ACHI 2011

The Fourth International Conference on Advances in Computer-Human Interactions

February 23-28, 2011 - Gosier

Guadeloupe, France

ACHI 2011 Editors

Les Miller, Iowa State University - Ames, USA

Silvana Roncagliolo, Pontificia Universidad Católica de Valparaíso, Chile

ACHI 2011

Foreword

The Fourth International Conference on Advances in Computer-Human Interactions [ACHI 2011], held between February 23-28, 2011 in Gosier, Guadeloupe, France, continued a series of events targeting traditional and advanced paradigms for computer-human interaction in multi-technology environments. The conference also covered fundamentals on interfaces and models, and highlighted new challenging industrial applications and research topics.

ACHI 2011 was proposed as 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 2011 also included a suite of specific domain applications, such as gaming, e-learning, social, medicine, teleconferencing and engineering.

We take here the opportunity to warmly thank all the members of the ACHI 2011 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 who dedicated much of their time and efforts to contribute to ACHI 2011. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ACHI 2011 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ACHI 2011 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of computer-human interaction.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the beautiful surroundings of Gosier, Guadeloupe, France.

ACHI 2011 Chairs

Les Miller, Iowa State University - Ames, USA
Alexandru Patriciu, McMaster University, Canada
Pierre Borne, Ecole Centrale de Lille, France
Anastasios A. Economides, University of Macedonia - Thessaloniki, Hellas

ACHI 2011

Committee

ACHI Advisory Chairs

Les Miller, Iowa State University - Ames, USA
Alexandru Patriciu, McMaster University, Canada
Pierre Borne, Ecole Centrale de Lille, France
Anastasios A. Economides, University of Macedonia - Thessaloniki, Hellas

ACHI 2011 Technical Program Committee

Eugenio Aguirre Molina, University of Granada, Spain
Cesar Alberto Collazos O., University of Cauca - Popayán, Colombia
Bernadetta Kwintiana Ane, Universität Stuttgart, Germany
Yuki Arase, Microsoft Research Asia, China
Peter Arato, Budapest University of Technology and Economics, Hungary
Karsten Berns, Technical University of Kaiserslautern, Germany
Pierre Borne, Ecole Centrale de Lille / IEEE Fellow, France
Luca Brayda, Istituto Italiano di Tecnologia (IIT)-Genova, Italy
Darwin G. Caldwell, Italian Institute of Technology - Genoa, Italy
Luis Carriço, University of Lisbon, Portugal
Ping Chen, University of Houston-Downtown, USA
Ryszard S. Choras, University of Technology & Life Sciences - Bydgoszcz, Poland
Timothy Coles, University of Wales Bangor & Italian Institute of Technology, UK
Chris Cornelis, Ghent University, Belgium
Anthony Cox, Centre for Psychology and Computing - Dartmouth, Canada
Lorcan Coyle, Lero-the Irish Software Engineering Research Centre, Ireland
Brian J. d'Auriol, Kyung Hee University, Korea
Dragos Datcu, Delft University of Technology - TUDelft, The Netherlands
Kerstin Dautenhahn, The University of Hertfordshire, UK
Berardina De Carolis, Università degli studi di Bari "Aldo Moro", Italy
Nancy Diniz, University of Hertfordshire - Hatfiel, UK
John M. Dolan, Carnegie Mellon University, USA
Mauro Dragone, University College Dublin (UCD), Ireland
Rolf Drechsler, University of Bremen, Germany
Carlos Duarte, University of Lisbon, Portugal
Anastasios A. Economides University of Macedonia - Thessaloniki, Hellas
Sandy El Helou, EPFL, Switzerland
Larbi Esmahi, Athabasca University, Canada
Elisabetta Farella, DEIS - University of Bologna, Italy
Maryanne Fisher, Saint Mary's University - Halifax, Canada
Somchart Fugkeaw, Thai Digital ID Co., Ltd. - Bangkok, Thailand
Mathias Funk, Eindhoven University of Technology, The Netherland
Matjaz Garns, Josef Stefan Institute- Ljubljana Slovenia
Denis Gillet, Swiss Federal Institute of Technology (EPFL), Switzerland

Jie Yin, CSIRO ICT Centre, Australia
Claudia Zapata del Rio, Pontifica Universidad Católica del Peru - Lima , Peru
Yu Zheng, Microsoft Research Asia - Beijing, P. R. China
Leyla Zhuhadar, Western Kentucky University/University of Louisville, USA

Table of Contents

Designing Healthcare Information System in Non-urban Area Using Neuroscientific Approach

Karim Fraoua, Christian Bourret
Equipe Dispositifs d'Information et de Communication à l'Ere Numérique (DICEN)
Université Paris-Est
77454 Marne-la-Vallée, France
fraoua@univ-mlv.fr, bourret@univ-mlv.fr

Abstract— **We will present at the beginning the situation of the Healthcare System in a non-urban area, and how to design a new conceptual framework in a Healthcare Information System. We will describe all the parameters of a significant good Healthcare System as viewed by a consumer. We will focus our works of what nowadays is known as human behavior or neuroscience. The analyses of information system must take into account much more the neuroscience approach, than limiting this analysis to the system components. We will focus our study to the non-urban people's interaction with an information system in a Healthcare area. We use a new approach especially neuroscience to represent the expected reactions in the human behavior and the impact expected in a medical healthcare information system. This consideration is due because of the wide range of motivators, and rewards that may induce irrational reactions that people show when they face to any new situation, especially those in non-urban area.**

Keywords- *Information System; HCI : Human Computer Interaction; neuroscience; non-urban area; healthcare*

I. INTRODUCTION

The non-urban areas know nowadays a major expansion that conducts cities to expand into rather rural areas for various reasons despite a lull observed during the 90's. Several types of non-urban areas have been identified, they have a lowest socio-economic indicators conjugated to a high demand of care. Geographically, these areas are located between the cities or urban area with highest infrastructure and rural areas devoid of any heavy hospital infrastructure, where the populations have changed their behavior to adapt to this lack of equipment and in which medicine is organized around the local doctor (primary care).

The population of non-urban area is displaced due to the enhancements of the property prices and the limited adaptation of the offer to a big family. The majority of these middle class populations consist of workers or employees. There are also a large community of immigrated from the first and second generation, the majority of these populations can be considered as currently disadvantaged groups, having low access to internet (50% for employees whereas it is more than 80% for urban population). The non-urban area represents only 18% of the Corsica area, but 75% of the Ile de France area [1]. Healthcare offer in these areas is relatively low as compared to urban area, with a density of medical offer below that of Paris or other city centers [1].

Nowadays, the medical situation in the non-urban Ile de France, which seems to be very similar to other cities, is considered as being in crisis [1,3]. The situation, according to the latest report of DREES (Direction de la recherche, des études, de l'évaluation et des statistiques) seems more disturbing especially from 2030 [4]. The same phenomena exist in United States for example, whereas disparities in access to medical care and utilization of services exist between urban and non-urban populations [2]. We observe in our preliminary results the same phenomena inside non-urban area. This result must be deeply investigated due to the fact that the population of non-urban area must be discriminated between well informed user and non educated user. This would be due to different reasons, as ethnicity, cultural effect, and non-connected population.

Our idea is to provide new theoretical approach of an Information System (IS) that will have a role in correcting social inequalities and to permit the access to a Healthcare Information System (HIS), in a context of cost reductions of healthcare system. This IS would be specifically developed for a non-urban people which suffer of a lack of access to existing IS or medical website. The main challenge remains the construction a system that matches user's capabilities. We will try in to open a new way for the construction of friendly and usable HIS.

II. DISCUSSION

It is admitted that the Healthcare Information Systems must be constructed in the perspective of the final users. Nevertheless, it is observed that a large number of HIS are designed without consideration of user-centered requirements. In consequence, when systems are created without further consideration such human behavior, users are dissatisfied and systems are not used, causing a lost of money and the degradation of the level of information of this population [5]. Numerous of health and medical information systems exist on the Internet in the form of aggregations of health data, but most of them are inaccessible to the population in non-urban area due principally to the symptoms described above, specially the low social level of this population. In a general observation, only 61% of Information System projects meet the requirements of the user needs [6]. A precise analysis of end-user permits to profile his characteristics such education, skill level, cultural environment, frequency of use, acceptance, expertise, knowledge, skills, cognitive capacities and limitations,

 ISBN: 978-1-61208-117-5

cultural background, times available for learning and training, and familiarity with IS.

Furthermore, it is admitted that the place and conditions in which the IS is located can play a great role in the interaction with the “patient” [7]. Experts agree that when the patients are more educated, they are more prompt to take their medication and to manage their own care [17].

This lack of these considerations make that many HIS are abandoned. This is not because misfit of technology, it is mainly due to the lack of systematic considerations of human being and behavior during the design and the implementation processes [8]. In designing of the most HIS, the importance of informing and actively engaging health care seekers in the planning and execution of their own treatment cannot be overstated. Efforts to achieve these goals must be based on integration of findings in studies from a wide range of disciplines bearing on motivation and decision making. To start this study, we must analyze in details the health’s user behavior trends which make them acting more as consumers than traditional patient, by describing the best medical ontology and whereas the IS can be located in a secure place for such user as a result of neuroscience trends, , in order to determine the most appropriate information system that can be deployed.

A. Theory of choice: From rational choice to neuroeconomic choice

In the rational model, the user is viewed as assigning utilitarian values to a range of presented options, both for information gathering and decision making. The preference based on utility function is the primary notion of economic rationality: the decision theory, game theory and the theory of general equilibrium model are based on this concept. This model of rational decisions making, based on VNM utility (von Neumann Morgenstern) [9], persists in contemporary decision theory and game theory. The rational agent is one who selects optimal decisions or strategies from his point of view or his self-interest. This last point seems important in this model, because it consider that agent have merely stable preferences and likelihood that all agents reason in the same manner. Nevertheless in neuroeconomic theory, the choice of human may be noted as puzzling in regard to rational model. The neuroeconomic theory gives a new sight of human choices.

Most of these choices are guided by psychological aspect, sometimes related to learning process. Indeed, in a recent report on the strategic development of health care system, Thaler *et al*. [10] show clearly these new trends of development of healthcare behavior. In this way the IS must be developed in preventive strategies that should not be intended only to inform but also encourage people to have new behavior in order to access to the IS. It was clearly identified that the difference between urban and non-urban people in use of services can be seen as differences in attitudes and behaviors toward seeking medical care [18].

B. The process of webdesign

Classical Human Computer Interaction (HCI) approaches remain indispensable to develop an information system, despite the fact that they are unable to answer to the complexity of healthcare information system. For this purpose, this study cannot be conducted without talking about HCI. Indeed, the HCI field is precisely positioned at the intersection of social sciences, behavioral sciences, computer sciences and information technology, in other words between psychology and informatics. Aiming in design a successful Information System depends on many elements as described above and they must be included during their implementation and evaluation, in order to satisfy the final users. Several models have been proposed in HCI, one of them that define the best information offer is described by The DeLone and McLean model [11,12].

Indeed this model contains six factor and some of them are interconnected: System quality, information quality, system use, user satisfaction, individual impacts and organizational impacts. In this way, an efficient system can be considered as an ineffective one if users don’t adopt it [19]. Moreover and until now, the HCI model gives a large place to the human characteristic, especially by the side of cognitive psychology and the social science. In support of this point of view, it has been shown that the social environment of the users affects also the information system process. From this analysis two parameters have been extracted [20]:

(1) Must share the same information by the users.
(2) Which resources are available to assist the users.

In this work, we choose to focus to the two main effects, structure effect and neuroeconomic approach, bridged to a classical approach used in HCI analysis as psychological and sociological approach, with implementing the healthcare information system in a secure place, and by building an ontology that correspond to level of the end-user and their aversion to the risk.

The main challenge in our innovative approach is to bring the patient to the information point access. Indeed the individual characteristics include attitude towards innovation and level of the user. In that way, the neuroscientific approach seems fundamental because it tends to study the factors which govern this acceptance. We recall briefly that the Technology Acceptance Model (TAM) has as a purpose to predict, and explain the user’s attitude toward an information system, by studying the factors which may influence him to accept or reject an IS [21].

Moreover, the style of the website is very important. Some websites use high or low medical jargon whilst some others are more or less formal. In both case, they seems inaccessible for a people with a low educational level, for this purpose the role of paramedical staff present in this place is to help them and to guide such user in their information research. They will play a role of assistance.

The adapted ontology must fit the behavioral attitude of these populations, otherwise it can put off someone who’s new to the information center disease and can be entirely appropriate by knowledgeable patients who belong generally to a high educated urban population. Describing purely formal ontologies or semiotic one, requires explicit and shared conceptualization, which are not be necessarily shared by these population having a low education level. In the configuration where the ontology is well adapted, the patient will be in a situation that may be considered as a very low risk for him and finally it will push him to act in favor to use the IS as a pre-diagnosis system or on self-medicament effect for example, which causes in France more than 10 000 death/year. From this point of view, the presence of forum in the website can plays a great role in avoiding this risky behavior specially addressed to these populations. Indeed, moderating forum embedded on the website could be critical to avoid potentially dramatic auto-medication.

Based on this approach describe above, the main contribution of this work, is to define a new way to create and validate an ontology in the field of medical IS, especially for the population situated in a low level area density of medical offer. For this reason, we try to extend the neuroscientific approach to the development of new IS and new ontology mainly related to a neuroscience using for example positive words that mimic reward, which is no longer a purely formal [13] or semiotic [14] which derives from Semantic Web or a "Socio-Semantic Web" and can be adopted only by a high educated people. Indeed, Zhang scheme gives a functional analysis that product an ontology of a given work domain. These considerations match with formal ontology. It includes (a) objects and their attributes, (b) resources and their types, (c) relations among entities and constraints on relations, (d) operations on single or multiple objects, transformations, relations, and constraints, and (e) workflow structures [8].

C. Information Point Access

The implementation of an information system must be made in a secure place like a city hall or a nursery office, or whatever paramedical office. It will create a confidence for the patient-user and this is the major opportunity to use a neuroeconomic approach in addition to a specific ontology derived from this theory. This can be considered as the reward by the end-user. Effectively, the reward in the case of healthcare system can be regarded as the result of the consultation in doctor's office and the main goal of this action: to get the information on therapy or to cure from disease. Indeed the presence of paramedical staff in a secure place, linked to a hospital network, with agreed information system, will make the patient more confident in such new medical approach. It has been proven that most frequently used quality criteria include accuracy, completeness and readability [15].

This can be obtained by a trustmarks website or when the IS is delivered by official organization. A recent experiment in an information systems shows, for those suffering from diabetes type II, one of the fastest growing health problems around the world, that such tools are more effective when it is designed to inform patients and integrate them as an actors into multiple-actor treatment teams including paramedical and social service personnel, along with nurses and doctors [16]. Within the Ile de France area, as in many similar communities elsewhere, this approach is particularly promising because it empowers patients to take an active role in their own treatment, and also because the distribution of nurses, social service and paramedical personnel is uniform throughout the non-urban areas.

III. CONCLUSION

This work is an exploratory work and still in progress; the observation made here about the difference between urban and non-urban populations is also observed elsewhere. The approach presented in this article indicates that efficiency dictates the patient choice and the user defines the quality of care as the primary utility function. The information tools for patients, is essential to improve the quality of care in non-urban areas. The IS must interact with local city hall or a paramedical staff, as it will involve the hospital network, to share data with the doctor offices and allow local hospital to act as a sentinel for access to these databases and to respond to outbreaks. The interface of Information System should be accessible, well designed and must correspond to the population's sociology, it is the main factor during the building of the ontologies. The importance of data will be crucial to regaining the trust of patients by ensuring the security of personal private data. The most important point of our approach focuses on the development of technologies needed to build and maintain public confidence in computer systems. This will help the e-health Information System to become ubiquitous for public whatever his education level or cultural belonging and especially for people socially and territorially isolated

These considerations invite us to plead for a new approach to build user-machine interfaces and to create neuroeconomic ontologies that user can use and which supports the reasoning of patients especially in decision making, rather than forcing them into a mode of thought which may be natural for machines or for well informed urban people, but not very useful for these populations, which are for the majority of them, socially excluded for a different reasons. In fact, the information system can't be considered as efficient if it is poorly adapted to their users, especially in their ways to use and to practice information.

REFERENCES

[1] Le Fur P. and Lucas-Gabrielli V., Rapport IRDES, 86, November 2004.

[2] Shi L., Macinko J., Starfield B., Politzer R., Wulu J. and Xu J., Public Health 119(8), 2005, pp. 699–710.

[3] La santé des franciliens, ORSIF, 2003.

[4] Numéro thématique, Les inégalités sociales de santé en France en 2006, Bulletin épidémiologique hebdomadaire, 2- 3, 23 janvier 2007.

[5] Johnson M. C., Johnson R.T. and Zhang J., Journal of Biomedical Informatics, 38, 2005, pp. 75–87.

[6] Williams D. and Kennedy M., In Proceedings of the 9th Annual International Symposium on Systems Engineering, International Council on Systems Engineering, Brighton, England, June 6–11, 1999.

[7] Bouwhuis D.G., Ergonomics, 43(7), 2000, pp. 908–919.

[8] Zhang J., Journal of Biomedical Informatics, 38(1), 2005, pp. 1-3.

[9] Neumann J. von, and Morgenstern O., Theory of Games and Economic Behavior, Princeton, NJ, Princeton University Press, 1944.

 ISBN: 978-1-61208-117-5

[10] (a) Oullier O., Aldini R., Thaler R. H. and Mullainathan S., In Nouvelles approches de la prévention en santé publique, 25, 2010, pp. 44-54.

(b) Oullier O. and Sauneron S., In Nouvelles approches de la prévention en santé publique, 25, 2010, pp. 55-65.

[11] DeLone W.H. and McLean E.R., In Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS_02). Big Island, Hawaii: Institute of Electrical and Electronics Engineers, Inc, 2002, pp. 238-248.

[12] DeLone W.H. and McLean E.R., Journal of Management Information Systems, 19(4), 2003, pp. 9-30.

[13] Guarino N., in Proceedings of the AIII Spring Symposium on Ontological Engineering, 1997, pp. 45-47.

[14] Zacklad M., In. Ingénierie des connaissances, IC'2005, Nice, Juin 2005.

[15] Eysenbach G and Köhler C., BMJ, 324, 2002, pp. 573-577.

[16] Mousquès J., Questions d'économie de la santé, Rapport IRDES, 89, Décembre 2004.

[17] Berg S., Asthma Magazine, 10(4), 2005, pp. 28-30.

[18] Harju B.L., Wuensch K.L., Kuhl E.A. and Cross N.J., Journal of Rural Health, 22, 2006, pp. 359–363.

[19] Ives B, Olson M. H. and Baroudi J.L., Commun ACM, 26(10), 1983, pp. 785–793.

[20] Hackos J. T., Redish J. C., User and task analysis for interface design, New York: Wiley; 1998

[21] Despont-Gros C., Mueller H. and Lovis C., Journal of Biomedical Informatics, 38, 2005, pp. 244–255.

Selecting the Right Task Model for Model-based User Interface Development

Gerrit Meixner
German Research Center for Artificial Intelligence
Innovative Factory Systems
Kaiserslautern, Germany
Gerrit.Meixner@dfki.de

Marc Seissler
University of Kaiserslautern
Center for Human-Machine-Interaction
Kaiserslautern, Germany
Marc.Seissler@mv.uni-kl.de

***Abstract* - This paper presents a taxonomy allowing for the evaluation of task models with a focus on their applicability in model-based user interface development processes. It further supports the verification and improvement of existing task models, and provides developers with a decision-making aid for the selection of the most suitable task model for their development process or project. Furthermore the taxonomy is applied on the Useware Markup Language 1.0. The results of the application are briefly described in this paper which led to the identification of substantial improvement potentials.**

Keywords - Task model, Taxonomy, Useware Markup Language, Model-based User Interface Development, MBUID.

I. INTRODUCTION

The improvement of human-machine-interaction is an important field of research reaching far back into the past [22]. Yet, for almost two decades, graphical user interfaces have dominated their interaction in most cases. In the future, a broader range of paradigms will emerge, allowing for multi-modal interaction incorporating e.g., visual, acoustic, and haptic input and output in parallel [41]. But also the growing number of heterogeneous platforms and devices utilized complementarily (e.g., PC's, smartphones, PDA) demand for the development of congeneric user interfaces for a plethora of target platforms; their consistency ensures their intuitive use and their users' satisfaction [16].

To meet the consistency requirement, factors such as reusability, flexibility, and platform-independence play an important role for the development of user interfaces [7]. Further, the perseverative development effort for every single platform, single platform or even single use context solution is way too high, so that a model-based approach to the abstract development of user interfaces appears to be favorable [31].

The purpose of a model-based approach is to identify high-level models, which, allow developers to specify and analyze interactive software applications from a more semantic oriented level rather than starting immediately to address the implementation level [18][36]. This allows them to concentrate on more important aspects without being immediately confused by many implementation details and then to have tools which, update the implementation in order to be consistent with high-level choices. Thus, by using models which capture semantically meaningful aspects, developers can more easily manage the increasing complexity of interactive applications and analyze them both during their development and when they have to be modified [28]. After having identified relevant abstractions for models, the next issue is specifying them through suitable languages that enable integration within development environments.

The pivotal model of a user-centric model-based development process is the task model [19]. Task models—developed during a user and use context analysis—are explicit representations of all user tasks [30]. Recently, several task modeling languages have been developed, which, differ, for example, in their degree of formalization, and their range of applications. To make the selection of a suitable task modeling language simpler, this paper introduces a task model taxonomy that enables all participants involved in an integrated MBUID (Model-based User Interface Development) process, to evaluate and compare task modeling languages.

The rest of this paper is structured as follows: Section II explains the proposed taxonomy for task models in detail. Section III gives a short introduction on the Useware Markup Language (useML) 1.0 followed by Section IV, which shows the application of the taxonomy on useML 1.0. The paper finishes with Section V, which gives a brief summary and an outlook on future activities.

II. THE TAXONOMY AND ITS CRITERIA

The proposed taxonomy focuses on the integration of task models into architectures for model-based development of user interfaces allowing for consistent and intuitive user interfaces for different modalities and platforms. For the evaluation of different task models, criteria describing relevant properties of these task models are needed. The criteria employed herein are based on initial work of [1] and [38], and are amended by additional criteria for task models with their application in MBUID. Following, the taxonomy and its criteria are described in detail.

A. *Criterion 1: Mightiness*

According to [26], a task model must help the developer to concentrate on tasks, activities, and actions. It must focus on the relevant aspects of task-oriented user interface specifications, without distracting by complexity. Yet, the granularity of the task definition is highly relevant. For the application of a task model in a MBUID process, the task model must comprise different levels of abstraction [15],

 ISBN: 978-1-61208-117-5

describing the whole bandwidth of interactions from abstract top-level tasks to concrete low-level actions. According to [34], it is commonly accepted that every person has her own mental representations (mental models) of task hierarchies. The hierarchical structure thereby constitutes the human's intuitive approach to the solution of complex tasks and problems. Consequently, complex tasks are divided into less complex sub-tasks [11] until a level is reached where sub-tasks can be performed easily. Normally, task models are divided into two levels of abstraction. With abstract tasks the user is able to model more complex tasks, e.g., "Edit a file." On the other hand a concrete task is an elemental or atomic task, e.g., "Enter a value." Tasks should not be modeled too detailed, e.g., like in GOMS [8] at least at development time [10].

Tasks can also be modeled from different perspectives. A task model should differentiate at least between interactive user tasks and pure system tasks [4]. Pure system tasks encapsulate only tasks which, are executed by the computer (e.g., database queries). This differentiation is preferable, because it allows for deducting when to create a user interface for an interactive system, and when to let the system perform a task automatically.

A further aspect determining the mightiness of a task model is its degree of formalization. Oftentimes, task modeling relies on informal descriptions, e.g., use cases [10] or instructional text [9]. According to [27], however, these informal descriptions do rarely sufficiently specify the semantics of single operators as well as the concatenation of multiple operators (i.e., to model complex expressions). These task models therefore lack a formal basis [33], which impedes their seamless integration into the model-based development of user interfaces [25]. On the one hand, developers need a clear syntax for specifying user interfaces, and on the other hand, they need an expressive semantic. Furthermore, the specification of a task model should be checked for correctness, e.g., with a compiler. For these reasons a task model should rather employ at least semi-formal semantics [24].

Using, for example, temporal operators (sometimes called qualitative temporal operators [14]) tasks can be put into clearly defined temporal orders [12]. The temporal order of sub-tasks is essential for task modeling [27] and opens up the road to a completely model-based development of user interfaces [15].

The attribution of optionality to tasks is another important feature of a task modeling language [1]. By itemizing a task as either optional or required, the automatic generation of appropriate user interfaces can be simplified. Similarly, the specification of cardinalities for tasks [26] allows for the automatic generation of loops and iterations. Several types of conditions can further specify when exactly tasks can, must, or should be performed. For example, logical [32] or temporal [14] conditions can be applied. Temporal conditions are also called quantitative temporal operators [14].

B. Criterion 2: Integratability

Due to the purpose of this taxonomy, the ease of a task model's integration into a consistent (or even already given) development process, tool-chain or software architecture [15], is an important basic criterion. Therefore it is necessary to have a complete model-based view, e.g., to integrate different other models (dialog model, presentation model, etc.) in the development process [37]. Among others, the unambiguity of tasks is essential, because every task must be identified unequivocally, in order to match tasks with interaction objects, and to perform automatic model transformations [40].

C. Criterion 3: Communicability

Although task modeling languages were not explicitly developed for communicating within certain projects, they are suitable means for improving the communication within a development team, and towards the users [29]. Task models can be employed to formalize [1], evaluate [32], simulate [27] and interactively validate [3] user requirements. A task model should therefore be easily, preferably intuitively understandable, and a task modeling language must be easy to learn and interpret. Semi-formal notations have shown to be optimally communicable [24] in heterogeneous development teams.

D. Criterion 4: Editability

This criterion defines how easy or difficult the creation and manipulation of a task model appears to the developer [6]. In general, we can distinguish between plain-text descriptions like e.g., GOMS [8] and graphical notations like e.g., CTT [26] or GTA [38]. For the creation of task models, graphical notations are better utilizable than textual notations [12]. For example, graphical notations depict hierarchical structures more intuitively understandable. Here, one can further distinguish between top-down approaches like CTT, and left-right orders such as in GTA.

Although this fourth criterion is correlated to the third one (communicability), they put different emphases. For every graphical notation, obviously, dedicated task model editors are essential [27].

E. Criterion 5: Adaptability

This criterion quantifies how easily a task model can be adapted to new situations and domains of applications. This applies especially to the development of user interfaces for different platforms and modalities of interaction. The adaptability criterion is correlated to the mightiness criterion. Especially while using task models in the development process of user interfaces for ubiquitous computing applications [39], run-time adaptability is an important criterion [5], which must be considered.

F. Criterion 6: Extensibility

The extensibility of a task modeling language is correlated to its mightiness and adaptability. This criterion reveals the ease or complicacy of extending the semantics and the graphical notation of the task modeling language.

This criterion is highly significant, because it is commonly agreed that there is no universal task modeling language which, can be applied to all domains and use cases [6]. In general, semi-formal notations are more easily extendable than fully formal ones. Formal notations are usually based on well-founded mathematical theories which, rarely allow for fast extensions.

G. *Criterion 7: Computability*

Computability quantifies the degree of automatable processing of task models. This criterion evaluates, among others, the data management, including the use of well-established and open standards like XML as data storage format. Proprietary formats should be avoided, because they significantly hinder the automatic processing of task models.

H. *Summary*

Some of the criteria are partly correlated, e.g., the Editability criterion is aiming in the same direction as the Communicability criterion, but their focus in terms of usability is quite different (see Figure 1). The Adaptability criterion is correlating with the Mightiness and the Extensibility criteria. Furthermore the Extensibility criterion is correlated to the Mightiness criterion.

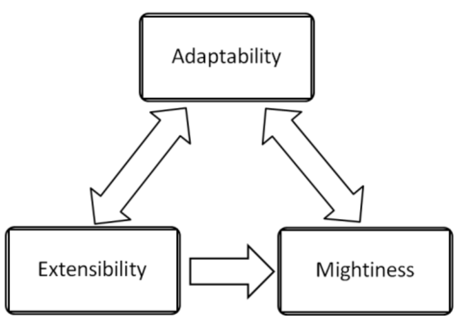


Figure 1: Correlating criteria

Table 1 shows all criteria and their possible values. All these possible values are more or less subjective. According to [6], the definition of more precise values is not possible, because there are no suitable metrics for value quantification.

TABLE I. CRITERIA AND VALUES

Criterion	Values
1. Mightiness	High, Medium, Low
a. Granularity	High, Medium, Low
b. Hierarchy	Yes, No
c. User- and system task	Yes, No
d. Degree of formalization	High, Medium, Low
e. Temporal operators	Yes, No
f. Optionality	Yes, No
g. Cardinality	Yes, No
h. Conditions	High, Medium, Low
2. Integratability	High, Medium, Low
3. Communicability	High, Medium, Low
4. Editability	High, Medium, Low
5. Adaptability	High, Medium, Low
6. Extensibility	High, Low
7. Computability	High, Low

III. USEWARE MARKUP LANGUAGE 1.0

The Useware Markup Language (useML) 1.0 had been developed by Reuther [32] to support the user- and task-oriented Useware Engineering Process [41] with a modeling language that could integrate, harmonize and represent the results of an initial analysis phase in one common, so-called use model in the domain of production automation. Accordingly, the use model abstracts platform-independent tasks, actions, activities, and operations into use objects that make up a hierarchically ordered structure. Each element of this structure can be annotated by attributes such as eligible user groups, access rights, importance. Use objects can be further structured into other use objects or elementary use objects. Elementary use objects represent the most basic, atomic activities of a user, such as entering a value or selecting an option. Currently, five types of elementary use objects exist [21]:

- Inform: the user gathers information from the user interface
- Trigger: starting, calling, or executing a certain function of the underlying technical device (e.g., a computer or field device)
- Select: choosing one or more items from a range of given ones
- Enter: entering an absolute value, overwriting previous values
- Change: making relative changes to an existing value or item

Figure 2 visualizes the structure of useML 1.0.

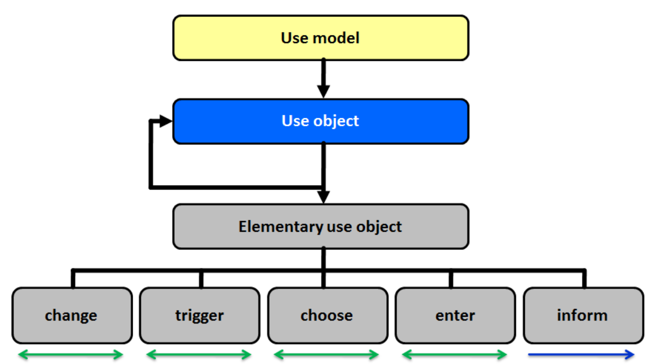


Figure 2: Schematic of useML 1.0

IV. APPLICATION OF THE TAXONOMY ON USEML 1.0

In the following subsections the application of the taxonomy on useML 1.0 is briefly described.

A. *Mightiness of useML 1.0*

useML 1.0's differentiation between use objects and five types of elementary use objects is sufficiently granular. With the classification of these elementary use objects types,

 ISBN: 978-1-61208-117-5

corresponding, abstract interaction objects can be determined [32]—which, the rougher differentiation of task types in the de facto standard CTT does not allow [2] [16] [35].
The use model or the useML 1.0 language, respectively, can be categorized as semi-formal. Though useML 1.0 is not based on formal mathematical fundamentals as e.g., Petri Nets [13], its structure is clearly defined by its XML schema. It allows, among others, for syntax and consistency checks which, ensure that only valid and correct use models can be created.

The use model by [32] focuses on the users' tasks, while those tasks which, are fulfilled solely by the (computer) system, can't be specified. Yet, for subsequently linking the use model to the application logic of a user interface, this task type is also required [2]. Querying a database might be such a pure system task which, however, might require that the query results are being presented to the user in an appropriate way. Pure system tasks can obviously be a part of a more complex, interactive action.

The hierarchical structure of the use model satisfies the Hierarchy sub-criterion of this taxonomy. Beside hierarchical structures, useML 1.0 also supports other structures, e.g., net structures. For the current useML 1.0 specification, however, no temporal operators were specified, which, constitutes a substantial limitation for the later integration of useML 1.0 into a fully model-based development process. In [32] Reuther himself admits that useML 1.0 does not possess temporal interdependencies between tasks. Task interdependencies must therefore be specified with other notations such as, e.g., activity diagrams. Such a semantic break, however, impedes developers in modeling the dynamics of a system, because they need to learn and use different notations and tools, whose results must then be consolidated manually. This further broadens the gap between Software- and Useware Engineering [41].

Although use models allow for specifying logical pre- and post-conditions, they don't support quantitative temporal conditions. Also, they lack means for specifying invariant conditions that must be fulfilled at any time during the accomplishment of the respective task. Finally, the current useML 1.0 version cannot indicate that certain use objects or elementary use objects are optional or required ones, respectively. Although there is a similar attribute which, can be set to a project-specific, relative value (between 1 and 10, for example), this is not an adequate mean for formally representing the optionality of a task. Accordingly, there are no language elements in useML 1.0 that specify the cardinality (repetitiveness) of a task's execution. The value of the Mightiness Criterion is based on the values of its sub criteria. Taking into account all the sub criteria, the value of the Mightiness criterion must be evaluated low.

B. *Integratability of useML 1.0*

Since no other models or modeling languages instead of use models or useML 1.0, respectively, have been applied and evaluated within projects pursuing the Useware Engineering Process, it is difficult to assess the applicability of use models into an integrated, MBUID architecture. Luyten mainly criticized the lack of dialog and presentation models complementing useML 1.0 [16]. Further, no unambiguous identifiers exist in useML 1.0 which, however, are required for linking (elementary) use objects to abstract or concrete interaction objects of a user interface—currently, use objects and elementary use objects can only be identified by their names that, of course, don't need to be unique. UseML 1.0 must therefore be extended to arrange for unique identifiers for (elementary) use objects, before it can be integrated into a complex architecture comprising multiple models representing relevant perspectives on the interaction between humans and machines. Until then, the integratability of useML 1.0 into such a model-based architecture must be rated low.

C. *Communicability of useML 1.0*

Since Useware Engineering demands for an interdisciplinary, cooperative approach [21], use models and useML 1.0 should be easily learnable and understandable. Being an XML dialect, in principal, useML 1.0 models can be viewed and edited with simple text or XML editors. Yet, these representations are difficult to read, understand, and validate. Readers with little knowledge in XML will have problems handling use models this way. Much better readability is achieved with the web-browser-like presentation of use models in the useML-Viewer by Reuther [32] (see Figure 3).

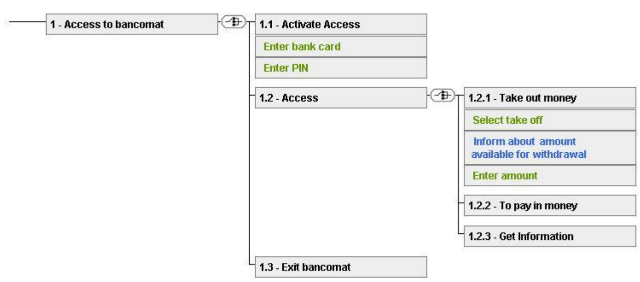


Figure 3: Excerpts of a use model as presented by the useML-Viewer

This HTML-based viewer allows for easily reading, understanding, and evaluating use models even without any knowledge in XML. It also prints use models using the web browsers' printer functions. However, the quality of the print is rather bad, among other reasons, because use models cannot be scaled to preferred paper sizes. Finally, the useML-Viewer can only display and print static use models, but does not provide means for interactive simulations or for the validation and evaluation of use models. Therefore, the communicability of useML 1.0 can only be rated medium.

D. *Editability of useML 1.0*

Though a simple editor may be sufficient for editing useML 1.0 models, XML editors are much more comfortable tools, especially those XML editors that run validity checks. Naturally, however, common versatile

 ISBN: 978-1-61208-117-5

XML editors from third party developers are not explicitly adapted to the specific needs of useML 1.0. Therefore, they cannot provide adequate means to simply and intuitively edit use models. The editability criterion of useML 1.0 must be rated low.

E. *Adaptability of useML 1.0*

useML 1.0 had been developed with the goal of supporting the systematic development of user interfaces for machines in the field of production automation. It focuses on the data acquisition and processing during the early phases of the Useware Engineering Process. Tasks, actions, and activities of a user are modeled in an abstract and platform-independent way. Thereby, the use model can be created already before the target platform has been specified. useML 1.0 provides for the incorporation of the final users and customers during the whole process, by allowing for the automatic generation of structure prototypes. The project-specific attributes (e.g., user groups, locations, device types) can be adjusted as needed, which, means that useML 1.0 can be employed for a huge variety of modalities, platforms, user groups, and projects. Among others, useML 1.0 has already been applied successfully, e.g., in the domain of clinical information system development [17]. In conclusion the adaptability criterion can be rated high.

F. *Extensibility of useML 1.0*

The fact that useML 1.0 is not strictly based upon well-grounded mathematical theories, actually simplifies its enhancement and semantic extension. This can simply be done by modifying the XML schema of useML 1.0. In most cases, however, not even this is necessary, because useML 1.0 comprises a separate XML schema containing project-specific attributes (e.g., user groups, locations, device types) which, can easily be adjusted without changing the useML 1.0's core schema. Since this allows for storing an unlimited number of use-case or domain-specific useML 1.0 schemes, the extensibility of useML 1.0 can be rated high.

G. *Computability of useML 1.0*

Since useML 1.0 is a XML dialect, use models can be further processed automatically. Employing dedicated transformations (e.g., XSLT style sheet transformations) prototypes can be generated directly from use models [21].

H. *Summary of the evaluation of useML 1.0*

The subsequently depicted table summarizes the evaluation of useML 1.0. Those criteria that were rated "No" or "Low", highlight severe deficits of the language. Figure 4 visualizes the results of the evaluation in a radar chart that reveals these deficits: They identify starting points for the upcoming, and for future improvements of the useML 1.0.

TABLE II. CRITERIA AND VALUES OF USEML 1.0

Criterion	Values
1. Mightiness	**Low**
a. Granularity	High
b. Hierarchy	Yes
c. User- and system task	**No**
d. Degree of formalization	Medium
e. Temporal operators	**No**
f. Optionality	**No**
g. Cardinality	**No**
h. Conditions	Medium
2. Integratability	**Low**
3. Communicability	Medium
4. Editability	**Low**
5. Adaptability	High
6. Extensibility	High
7. Computability	High

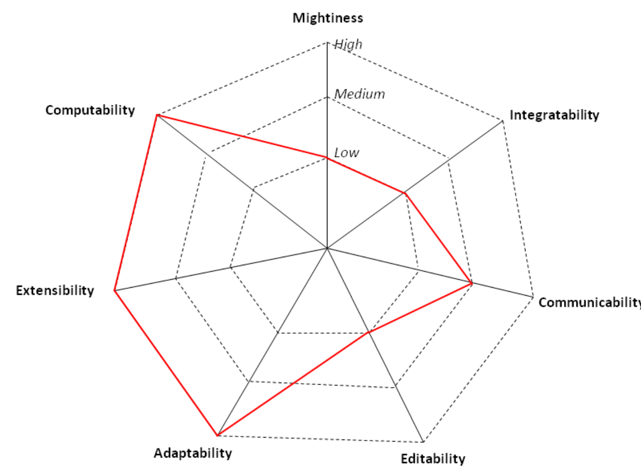


Figure 4: Results of the evaluation of useML 1.0

V. CONCLUSION AND OUTLOOK

In this paper, a taxonomy for task models has been proposed, to simplify the selection of the most suitable task model for projects employing model-based development processes for user interfaces. Furthermore to show the feasibility of the task model taxonomy, it has been applied on useML 1.0. Also the application of the taxonomy on useML 1.0 showed the need for enhancing useML 1.0 semantically.

Currently we're enhancing useML 1.0 in different aspects, according to the initial results of the application of the taxonomy. Additionally, we would like to improve the refinement of the criteria and apply this taxonomy to a selection of further task models, such as CTT [26] or AMBOSS [20] to proof the usefulness of this taxonomy.

REFERENCES

[1] S. Balbo, N. Ozkan, and C. Paris, "Choosing the right task modelling notation: A Taxonomy" in the Handbook of Task Analysis for

Human-Computer Interaction, D. Diaper and N. Stanton, Eds., Lawrence Erlbaum Associates, pp. 445–466, 2003.

[2] M. Baron and P. Girard, "SUIDT: A task model based GUI-Builder", Proc. of the 1st International Workshop on Task Models and Diagrams for User Interface Design, 2002.

[3] M. Biere, B. Bomsdorf, and G. Szwillus, „Specification and Simulation of Task Models with VTMB", Proc. of the 17th Annual CHI Conference on Human Factors in Computing Systems, ACM Press, New York, pp. 1–2, 1999.

[4] B. Bomsdorf and G. Szwillus, "From task to dialogue: Task based user interface design", SIGCHI Bulletin, vol. 30, nr. 4, pp. 40–42, 1998.

[5] K. Breiner, O. Maschino, D. Görlich, G. Meixner, and D. Zühlke, "Run-Time Adaptation of a Universal User Interface for Ambient Intelligent Production Environments", Proc. of the 13th International Conference on Human-Computer Interaction (HCII) 2009, LNCS 5613, pp. 663–672, 2009.

[6] P. Brun and M. Beaudouin-Lafon, "A taxonomy and evaluation of formalism for the specification of interactive systems", Proc. of the Conference on People and Computers, 1995.

[7] G. Calvary, J. Coutaz, J., and D. Thevenin, "A Unifying Reference Framework for the Development of Plastic User Interfaces", Proc. of the Eng. Human-Computer-Interaction Conference, pp. 173-191, 2001.

[8] S. K. Card, T. P. Moran, and A. Newell, "The psychology of human-computer interaction", Lawrence Erlbaum Associates, 1983.

[9] J. Carroll, "The Nurnberg Funnel: Designing Mini-malist Instruction for Practical Computer Skill", MIT Press, 1990.

[10] L. Constantine and L. Lockwood, "Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design". Addison-Wesley, 1999.

[11] A. Dittmar, "More precise descriptions of temporal relations within task models", Proc. of the 7th International Workshop on Interactive Systems: Design, Specification and Verification, pp. 151–168, 2000.

[12] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, "Human-Computer Interaction, 3rd ed., Prentice Hall, 2003.

[13] C. Girault and R. Valk, "Petri Nets for Systems Engineering", Springer, 2003.

[14] X. Lacaze and P. Palanque, "Comprehensive Handling of Temporal Issues in Task Models: What is needed and How to Support it?", Proc. of the 22th Annual CHI Conference on Human Factors in Computing Systems, 2004.

[15] Q. Limbourg, C. Pribeanu, and J. Vanderdonckt, "Towards Uniformed Task Models in a Model-Based Approach", Proc. of the 8th International Workshop on Interactive Systems: Design, Specification and Verification, pp. 164–182, 2001.

[16] K. Luyten, "Dynamic User Interface Generation for Mobile and Embedded Systems with Model-Based User Interface Development", PhD thesis, Transnationale Universiteit Limburg, 2004.

[17] G. Meixner, N. Thiels, and U. Klein, "SmartTransplantation – Allogeneic Stem Cell Transplantation as a Model for a Medical Expert System", Proc. of Usability & HCI for Medicine and Health Care, Graz, Austria, pp. 306–317, 2007.

[18] G. Meixner, D. Görlich, K. Breiner, H. Hußmann, A. Pleuß, S. Sauer, and J. Van den Bergh, "4th International Workshop on Model Driven Development of Advanced User Interfaces", CEUR Workshop Proceedings, Vol-439, 2009.

[19] G. Meixner, "Model-based Useware Engineering", W3C Workshop on Future Standards for Model-Based User Interfaces, Rome, Italy, 2010.

[20] M. Giese, T. Mistrzyk, A. Pfau, G. Szwillus, and M. Detten, „AMBOSS: A Task Modeling Approach for Safety-Critical Systems", Proc. of the 2nd Conference on Human-Centered Software Engineering and 7th international Workshop on Task Models and Diagrams, Pisa, Italy, pp. 98–109, 2008.

[21] K. S. Mukasa and A. Reuther, "The Useware Markup Language (useML) - Development of User-Centered Interface Using XML", Proc. Of the 9th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine-Systems, Atlanta, USA, 2004.

[22] B. Myers, "A brief history of human-computer interaction technology", interactions, vol. 5, nr. 2, pp. 44–54, 1998.

[23] H. Oberquelle, "Useware Design and Evolution: Bridging Social Thinking and Software Construction", in Social Thinking – Software Practice, Y. Dittrich, C. Floyd, and R. Klischewski Eds., MIT-Press, Cambridge, London, pp. 391–408, 2002.

[24] N. Ozkan, C. Paris, and S. Balbo, "Understanding a Task Model: An Experiment", Proc. of HCI on People and Computers, pp. 123–137, 1998.

[25] P. Palanque, R. Bastide, and V. Sengès, "Validating interactive system design through the verification of formal task and system models", Proc. of the IFIP Working Conference on Engineering for Human-Computer Interaction, pp. 189–212, 1995.

[26] F. Paternò, "Model-based design and evaluation of interactive applications", Springer, 1999.

[27] F. Paternò, "ConcurTaskTrees: An Engineered Notation for Task Models" in the Handbook of Task Analysis for Human-Computer Interaction, D. Diaper and N. Stanton, Eds., Lawrence Erlbaum Associates, pp. 483–501, 2003.

[28] F. Paternò, "*Model-based Tools for Pervasive Usability"*, Interacting with Computers, Elsevier, vol. 17, nr. 3, pp. 291–315, 2005.

[29] C. Paris, S. Balbo, and N. Ozkan, "Novel use of task models: Two case studies", in Cognitive task analysis, J. M. Schraagen, S. F. Chipmann and V. L. Shalin, Eds., Lawrence Erlbaum Associates, pp. 261–274, 2000.

[30] C. Paris, S. Lu, and K. Vander Linden, "Environments for the Construction and Use of Task Models" in the Handbook of Task Analysis for Human-Computer Interaction, D. Diaper and N. Stanton, Eds., Lawrence Erlbaum Associates, pp. 467–482, 2003.

[31] A. Puerta, "A Model-Based Interface Development Environment", IEEE Software, vol. 14, nr. 4, pp. 40–47, 1997.

[32] A. Reuther, "useML – systematische Entwicklung von Maschinenbediensystemen mit XML", Fortschritt-Berichte pak, nr. 8, Kaiserslautern, TU Kaiserslautern, PhD thesis, 2003.

[33] D. Scapin and C. Pierret-Golbreich, "Towards a method for task description: MAD", Proc. of the Conference on Work with DisplayUnits, pp. 27–34, 1989.

[34] S. Sebillotte, "Hierarchical planning as a method for task analysis: The example of office task analysis", Behavior and Information Technology, vol. 7, nr. 3, pp. 275–293, 1988.

[35] J. Tarby, "One Goal, Many Tasks, Many Devices: From Abstract User Task Specification to User Interfaces" in the Handbook of Task Analysis for Human-Computer Interaction, D. Diaper and N. Stanton, Eds., Lawrence Erlbaum Associates, pp. 531–550, 2003.

[36] J. Van den Bergh, G. Meixner, K. Breiner, A. Pleuß, S. Sauer, and H. Hußmann, "5th International Workshop on Model Driven Development of Advanced User Interfaces", CEUR Workshop Proceedings, Vol-617, 2010.

[37] J. Van den Bergh, G. Meixner, and S. Sauer, „MDDAUI 2010 workshop report", Proc. of the 5th International Workshop on Model Driven Development of Advanced User Interfaces, 2010.

[38] M. Van Welie, G. van der Veer, and A. Eliens, "An ontology for task world models", Proc. of the 5th International Workshop on Interactive Systems: Design, Specification and Verification, pp. 57–70, 1998.

[39] M. Weiser, "The computer for the 21st century", Scientific American, vol. 265, nr. 3, pp. 94–104, 1991.

[40] A. Wolff, P. Forbrig, A. Dittmar, and D. Reichart, „Linking GUI Elements to Tasks – Supporting an Evolutionary Design Process", Proc. of the 4th International Workshop on Task Models and Diagrams for User Interface Design, pp. 27–34, 2005.

[41] D. Zuehlke and N. Thiels, „Useware engineering: a methodology for the development of user-friendly interfaces", Library Hi Tech, vol. 26, nr. 1, pp. 126–140, 2008.

Discourse-based Interaction Models for Recommendation Processes

Dominik Ertl, Hermann Kaindl, Edin Arnautovic, Jürgen Falb, and Roman Popp
Vienna University of Technology,
Institute of Computer Technology
A–1040 Vienna, Austria
{*ertl, kaindl, arnautovic, falb, popp*}*@ict.tuwien.ac.at*

***Abstract*—Manually creating recommendation processes and their user interfaces usually requires a lot of effort. Therefore, we propose high-level interaction design and automated generation of user interfaces for building dialogue-based product advisors. For this purpose, this work introduces discourse models as interaction models used for modeling recommendation processes. Such discourse models refer to domain-of-discourse models, which represent, among other concepts, the products (and their related product categories) that are to be recommended. So, this paper presents how discourse-based interaction models and their corresponding domain-of-discourse models can be used for modeling recommendation processes.**

Keywords*—*Interaction design; discourse model; recommendation process.

I. Introduction

Dialogue-based product advisors have become very popular on the Web in the context of e-commerce. Web users try to inform themselves about certain products, and many companies use recommendation processes to offer information and stimulate demand for their products. The range of products to be recommended is nearly unlimited. There are several underlying principles to build such a dialogue-based product advisor [2]. The possible interactions between a user and the dialogue-based product advisor, however, are often programmed directly into a graphical user interface (GUI) without a high-level interaction design.

This paper introduces discourse models as interaction models of recommendation processes. We focus on processes in knowledge-based recommender systems that allow for *preference elicitation* and do not involve user modeling and/or profiling. We use discourses to model the possible interactions between a human and a computer-based recommendation process as dialogues between them.

Figure 1 shows an overview of our approach (concrete examples can be found below). It illustrates that such a discourse model refers to an underlying model of the domain-of-discourse, which represents the possible content of the dialogues (concrete values of answers). In principle, such domain-of-discourse models can be derived from (product) ontologies like *GoodRelations* [1]. Discourse models serve together with related domain-of-discourse-models as a basis for GUI generation (see, e.g., [6]). We also plan to generate

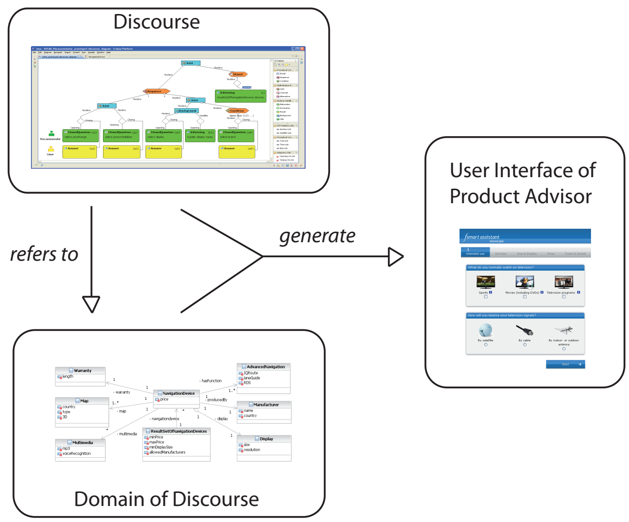


Figure 1. Overview of our approach.

GUIs for real recommendation processes to be used in commercial applications.

The remainder of this paper is organized in the following manner. First, we present background information about our discourse-based interaction models in general. Then we discuss related work. After that, we show how such a discourse model looks like for a recommendation process, using a class of recommendation dialogues for navigation devices as an example. Extending the same example, we finally show how a related domain-of-discourse model looks like.

II. Background

Recommendation processes for preference elicitation have a structure of (partially ordered) sequences of questions and answers. They are potentially clustered into groups of questions/answers that belong together according to their semantics.

Discourse-based interaction models fit well for defining such structures. Our approach to discourse modeling is based on several theories of human communication from several fields [5]. The key ingredients of our discourse models are

Communicative Acts as derived from speech acts [9], *Adjacency Pairs* adopted from Conversation Analysis [7], and RST relations inherited from Rhetorical Structure Theory (RST) [8].

Communicative Acts (semi-structured messages with intention) represent basic units of language communication. Thus, any communication can be seen as enacting Communicative Acts: acts such as making statements, giving commands, asking questions and so on. Communicative Acts carry the intention of the interaction (e.g., asking a question or issuing a request).

Adjacency Pairs are sequences of talk-turns that are specific to human (oral) communication, e.g., a question should have a related answer. RST relations specify relationships among text portions and associated constraints and effects. The relationships in a text are organized in a tree structure, where the rhetorical relations are associated with non-leaf nodes, and text portions with leaf nodes. In our work, we use RST for linking Adjacency Pairs of Communicative Acts and further structures made up of RST relations. We have also included procedural constructs, to provide means to express a particular order during discourse execution, to specify repetitions or conditional execution of different discourse parts. While we only use a few types of Communicative Acts and RST relations yet, our modeling tool-kit has been sufficient for modeling even real-world discourses.

III. Related Work

Chen et al. [3] present interaction design guidelines for a critiquing-based recommender system that acts like an artificial salesperson. It engages users in a dialogue where users can provide feedback in the form of critiques to the sample items that were shown to them. The authors point out that the feedback, in turn, enables the system to refine its understanding of the user's preferences and prediction of what the user "truly wants". Our work targets on interaction design as well, however, we do not provide user prediction behavior of the system but models in terms of discourses.

Doyle and Cunningham [4] analyze the problem of deciding on the set of questions to ask in a session of navigation-by-asking recommender systems, including the issue of optimal ordering of the questions. Their work includes the evaluation of different question-selection criteria. Our modeling approach is more comprehensive, however.

IV. Discourse Models for Recommendation Processes

Figure 2 shows a discourse model for our running example, that models high-level interactions for recommendations regarding a product, in our case a navigation device. The diagram shows *Communicative Acts* of two participating actors, a human user who is the potential customer (light, yellow boxes), and the computer (dark, green boxes). More precisely, the user interacts with the dialogue-based Product Advisor according to this process.

The recommendation process is modeled as a sequence of several questions and related answers (*Adjacency Pairs*, shown as diamonds in Figure 2), which are supposed to help the user of the dialogue-based Product Advisor finding a navigation device that fits his or her wishes and needs. This is a process of preference elicitation. The overall procedural construct used in this model is a *Sequence*. More precisely, this construct does not directly link these *Adjacency Pairs* but trees of discourse relations that contain them.

In the first branch of the *Sequence* relation (left in the figure), a *Joint* relation combines two of these *Adjacency Pairs*. Joints below the Sequence cluster questions that hold a semantic relation. The first question gathers information on the price range, defining the minimum and maximum price that the user is potentially willing to pay. The second question elicits all manufacturers of navigation devices the user is interested in. This is a *closed* question since it provides all available choices of manufacturers.

In the second branch of the *Sequence* relation (right in the figure), another *Joint* relation combines two more *closed questions* about the voice recognition and mp3 functions. This second Joint clusters *multimedia* questions. The question about choosing voice recognition is additionally related with another *Adjacency Pair* that has an *Informing* as opening Communicative Act (here, no closing Communicative Act is needed). *Background* is an RST relation that optionally informs the human user on additional details about the subject matter, e.g., more information on voice recognition. Moreover, with the procedural construct *Condition* we define a condition that has to be fulfilled at runtime to have this branch considered. In the running example, the question about the mp3 function is only asked if the user has selected voice recognition. So, in this context of recommendation processes, such a *Condition* implements a rule for determining whether a certain question will be asked by the Product Advisor, i.e., whether it will be displayed in its GUI.

Of course, several more questions and answers would be included here in a real recommendation process. For the purpose of this running example, however, let us assume that this is the whole sequence.

The *Joint* at the top of the sequence has a procedural construct *IfUntil* with an *Informing* about all products that match the preference elicitations, if any. Whenever an answer is selected in any question of this recommendation process, the updated list of matching products (the results) is presented to the user. However, the discourse model does not contain the rules for matching of products at runtime. They have to be contained in the application logic of the dialogue-based Product Advisor. Still, it should be clear that subsequently selecting a product to buy, and paying it can also be modeled according to our approach.

 ISBN: 978-1-61208-117-5

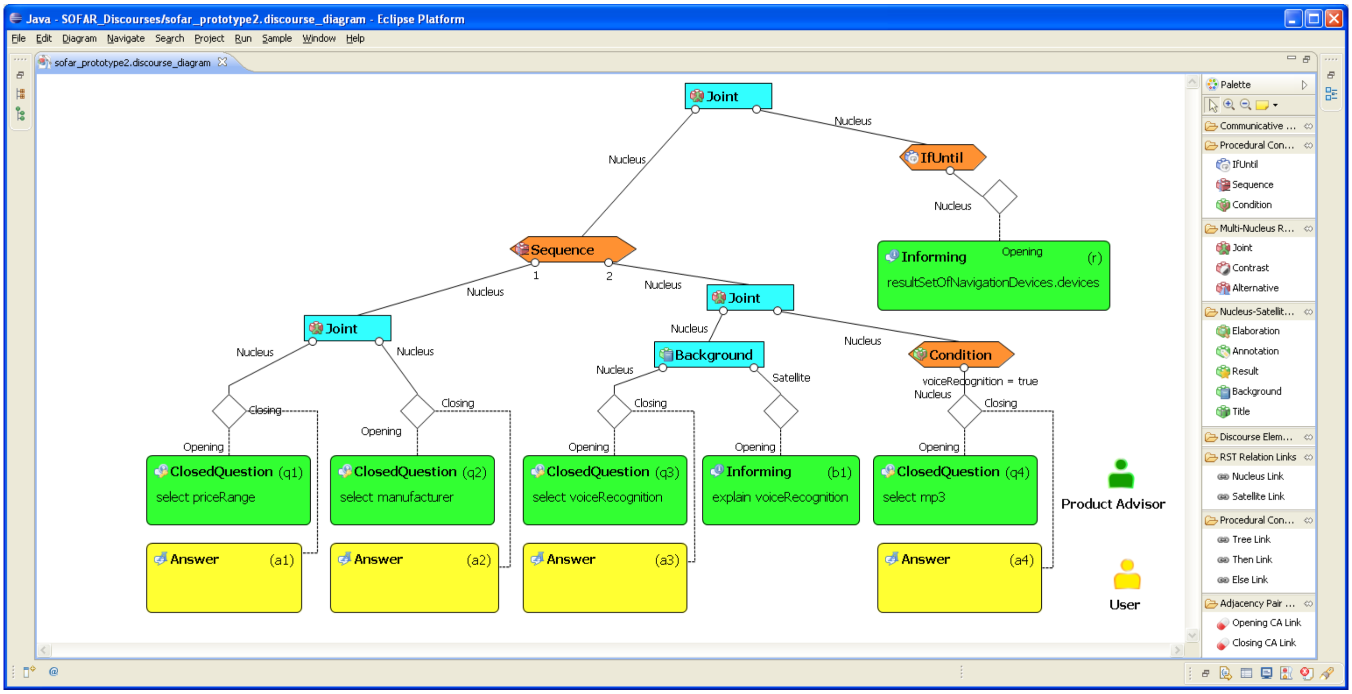


Figure 2. Discourse model representing a recommendation process for a navigation device.

In addition, *Communicative Acts* specify their propositional content, that refers to concepts in the domain of discourse. In Figure 2, these references are given in a short-hand notation, e.g., “select priceRange”. These references link to model elements in the domain-of-discourse model in our approach.

V. Domain-of-Discourse Models for Recommendation Processes

A domain-of-discourse model may have been informed from a related product ontology. However, for the dialogues in the course of the recommendation process, not all concepts from the ontology are relevant. So, the domain-of-discourse model may be seen as the subset of an ontology that is sufficient to specify the content of the Communicative Acts within the defined discourses.

Figure 3 shows a UML class diagram for an example of such a domain-of-discourse model from the domain of navigation devices. Such types are defined more generally, e.g., in a product ontology. A NavigationDevice in our example model has the attribute *price*, parts like a *Display*, as well as references to other properties and functions, e.g., *Multimedia* functions. These other properties and functions can have their own attributes.

Now let us explain more technically what the references from the discourse model to this model mean, through expanding the short-hands from Figure 2:

- select priceRange: *min and max price from Range(Min(all NavigationDevice.price), Max(all NavigationDevice.price))*. The Product Advisor requests the user to set the *minPrice* and *maxPrice* attributes of the instance named *resultSetOfNavigationDevices* of the class *ResultSetOfNavigationDevices*.
- select manufacturer: *select many manufacturer from all Manufacturer*. The dialogue-based Product Advisor presents a set of all available Manufacturers to the user, who selects one or more of them to specify the list of manufacturers of interest. The Product Advisor sets the attribute *manufacturersOfInterest* of *resultSetOfNavigationDevices* with this list.
- select voiceRecognition: *select voiceRecognition from MultiMedia*. The Product Advisor presents a check box that the user can select to get a navigation device with or without voice recognition. The Product Advisor sets the attribute *voiceRecognition* to true or false.
- explain voiceRecognition: The Product Advisor presents a background text explaining the feature to support the user in her decision if she wants to get a navigation device with or without voice recognition.
- select mp3: *select mp3 from MultiMedia*. The Product Advisor presents a check box to the user. Thus, the user can select to get a navigation device with or without mp3 capability. The Product Advisor then sets the attribute *mp3* to true or false.
- resultSetOfNavigationDevices.devices: *display ResultSetOfNavigationDevices::resultSetOfNavigation Devices.devices*. The GUI displays all Devices which are referenced by the resultSetOf

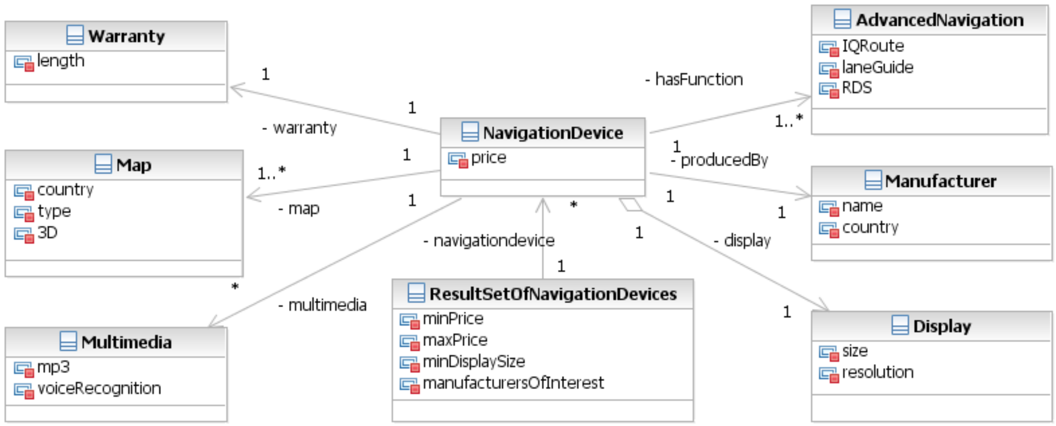


Figure 3. Domain-of-discourse model for a navigation device recommendation process.

NavigationDevices. This list contains only the devices fulfilling the defined constraints.

VI. Conclusion

In this paper, we show that and how a discourse model can represent a high-level interaction design of a recommendation process. Such discourse models are classes of dialogues that are possible between a human user and a Product Advisor that implements this recommendation process. Such a discourse model refers to a domain-of-discourse model in the sense that the latter specifies the content of the dialogues of the former.

A domain-of-discourse model may be part of a related ontology that is directly relevant for the dialogues. We currently work on support for extracting and deriving domain-of-discourse models from ontologies.

In our previous work, we have already generated multimodal user interfaces automatically from discourse models together with related domain-of-discourse models. We are currently working on feeding automatically generated structural GUI models into a real dialogue-based Product Advisor, that will generate its usual Web-based GUIs from them.

Acknowledgment

This research has been carried out in the SOFAR project (No. 825061), funded by the Austrian FIT-IT Program of the FFG and Smart Information Systems GmbH.

References

[1] Goodrelations: Last visited on December 10, 2010. http://www.heppnetz.de/projects/goodrelations/.

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[3] L. Chen and P. Pu. Interaction design guidelines on critiquing-based recommender systems. *User Modeling and User-Adapted Interaction*, 19(3):167–206, 2009.

[4] M. Doyle and P. Cunningham. A dynamic approach to reducing dialog in on-line decision guides. In *Proceedings of the 5th European Workshop on Advances in Case-Based Reasoning (EWCBR '00)*, pages 49–60. Springer-Verlag: London, UK, 2000.

[5] J. Falb, H. Kaindl, H. Horacek, C. Bogdan, R. Popp, and E. Arnautovic. A discourse model for interaction design based on theories of human communication. In *Extended Abstracts on Human Factors in Computing Systems (CHI '06)*, pages 754–759. ACM Press: New York, NY, 2006.

[6] J. Falb, S. Kavaldjian, R. Popp, D. Raneburger, E. Arnautovic, and H. Kaindl. Fully automatic user interface generation from discourse models. In *Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI '09)*, pages 475–476. ACM Press: New York, NY, 2009.

[7] P. Luff, D. Frohlich, and N. Gilbert. *Computers and Conversation*. Academic Press, London, UK, January 1990.

[8] W. C. Mann and S. Thompson. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.

[9] J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, England, 1969.

Combining agile methods and user-centered design to create a unique user experience: An empirical inquiry

Cynthia Y. Lester
Department of Computer Science
Tuskegee University
Tuskegee, Alabama, USA
cylester@tuskegee.edu

Abstract **- With the advent of the Internet and websites, many people believe that website development is as easy as dragging an icon here, placing a menu there, and adding a picture. However, there is more to website design than many people believe especially if you desire to develop a website that meets the needs of the user and follows software engineering principles. While there are many software process models and human-computer interaction activities that focus on the user, the integration of these activities is quite difficult, especially as it relates to website development. This paper presents the results of an empirical investigation that combined one activity of human-computer interaction, user-centered design, and one software engineering method, agile development into a small-scale development exercise that specifically focused on website development. The results from the study suggest that using the hybrid approach for small-scale projects is easy to implement, but is not without challenges.**

Keywords – agile development; human-computer interaction; software engineering; user-centered design

I. INTRODUCTION

The use of technology and the Internet is commonplace in today's society. In 1990, it was reported that there were less that 50 million users of the Internet in the U.S. However, by 2008 the U.S. reported approximately 230,630,000 Internet users [1]. Therefore, it stands to reason that with more users and more advanced systems, the user population of today's technology would be more technically savvy than those user groups of yesteryear. However, the average user is now less likely to understand the systems of today as compared to the users of a decade ago. Consequently, the designers and developers of these systems must ensure that the systems are designed with the three "use" words in mind so that the system is successful. Hence, the system must be useful, usable, and used [2]. The last of the "use" terms has not been a major factor until recently, thereby making the discipline of human-computer interaction increasingly more important.

Human-computer interaction (HCI) has been described 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 [3], [4]. 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 [5]. 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. Consequently, system developers need to further their understanding of the human, the user, and the interaction.

The aim of this paper is to present the results from an empirical inquiry that combined one activity of HCI, user-centered design, and one software engineering method, agile development, to develop a website for a small-sized business. The paper also touches on the theme of extreme programming as the implementation methodology for agile methods. While there are many different development strategies specifically for website design and development, a review by the author revealed that there was little consistency among the processes and some did not address user involvement or the user experience. Therefore, a hybrid approach using agile development and user-centered design was considered since both focus on the inclusion of the user throughout the development process.

The paper is divided into the following sections: the human, the system, and the interaction; traditional software methodologies; agile methods; user-centered design; a practical implementation combining the two methods; a discussion of the empirical investigation; and concluding thoughts. It is the desire of the author that the readers of the paper will see how closely related the two methodologies are and how they can be used together for small software development projects that yield high levels of user involvement while creating an enriched user and developer experience.

 ISBN: 978-1-61208-117-5

II. The Human, the System, and Interaction

A. The human user

The human user may be an individual or a group of users who employ the computer to accomplish a task. The human user may be a novice, intermediate, or expert who uses the technological system. Further, the human user may be a child using the system to complete a homework assignment or an adult performing a task at work. Additionally, the human user may be a person who has a physical or cognitive limitation which impacts his/her use with the computer-based system. No matter who the human user is, the goal when interacting with a computer system is to have a seamless interaction which accomplishes the task.

B. The computer

According to the *Random House Unabridged Dictionary*, a computer is defined as an electronic device designed to accept data, perform prescribed mathematical and logical operations at high speed, and display the results of these operations [6]. However, as computers become more complex, users expect more than just a display of the results of their operations. The term computer system is used to represent technology and technological systems. Consequently, technology or technological systems encompass many different aspects of computing. Users now require their systems to be able to provide answers to questions, to store various forms of information such as music, pictures, and videos, to create a virtual experience that physically may be unattainable, and to understand verbal, visual, audio, and tactile feedback, all with the click of a button. As the human user becomes to depend on these technological systems more, the interaction between the user and the system becomes more complex.

C. The interaction

Interaction is the communication between the user and the computer system. For computer systems to continue their wide spread popularity and to be used effectively, the computer system must be well designed. According to Sharp, Rogers, and Preece, a central concern of interaction design is to develop an interactive system that is usable [4]. More specifically, the computer system must be easy to use, easy to learn, thereby creating a user experience that is pleasing to the user. Consequently, when exploring the definition of interaction, four major components are present which include:

- The end user
- The person who has to perform a particular task
- The context in which the interaction takes place
- The technological systems that is being used

Each of these components has its own qualities and should be considered in the interaction between the computer system and the user. In his bestselling book, *The Design of Everyday Things*, Donald Norman writes about these components and how each must interact with the other, suggesting that the common design principles of visibility and affordance help to improve interaction [7]. The principle of visibility emphasizes the idea that the features of the system in which the user interacts should be clearly visible and accessible to human sense organs, which improves the interaction between the action and the actual operation [7]. The principle of affordance as suggested by Jef Raskin, should accommodate visibility such that the method of interacting with the system should be apparent, just by looking at it [8].

Therefore, in order to create an effective user experience, a designer of an interactive computer system must understand the user for which the system is being created, the technological system that is being developed and the interaction that will take place between the user and the computer system. However, traditional plan-driven software engineering methodologies often make integrating the user into the development process to achieve an effective user experience difficult.

III. Traditional Software Methodologies

Software engineering is defined as "being concerned with all aspects of the development and evolution of complex systems where software plays a major role. It is therefore concerned with hardware development, policy and process design and system deployment as well as software engineering [9]."

The term software engineering was first proposed at the 1968 NATO Software Engineering Conference held in Garmisch, Germany. The conference discussed the impending software crisis that was a result of the introduction of new computer hardware based on integrated circuits [9]. It was noted that with the introduction of this new hardware, computer systems were becoming more complex which dictated the need for more complex software systems. However, there was no formalized process to build these systems which put the computer industry at jeopardy because systems were often unreliable, difficult to maintain, costly, and inefficient [9]. Consequently, software engineering surfaced to combat the looming software crisis.

Since its inception, there have been many methodologies that have emerged that lead to the production of a software product. The most fundamental activities that are common among all software processes include [9]:

- *Software specification* – the functionality of the system and constraints imposed on system operations are identified and detailed
- *Software design and implementation* – the software is produced according to the specifications

- *Software validation* – the software is checked to ensure that it meets its specifications and provides the level of functionality as required by the user
- *Software evolution* – the software changes to meet the changing needs of the customer

The activities that formulate this view of software engineering came from a community that was responsible for developing large software systems that had a long life span. Moreover, the teams that used this methodology were typically large teams with members sometimes geographically separated and working on software projects for long periods of time [9]. Therefore, software development methodologies that resulted from this view of software engineering were often termed as "heavyweight" processes because they were plan-driven and involved overhead that dominated the software process [9]. However, great difficulty occurs when these methodologies are applied to smaller-sized businesses and their systems, because these methods lack the agility needed to meet the changing needs of the user. The next section presents an overview of an alternative to heavyweight processes, agile development.

IV. AGILE METHODS

In an effort to address the dissatisfaction that the heavyweight approaches to software engineering brought to small and medium-sized businesses and their system development, in the 1990s a new approach was introduced termed, "agile methods." Agile processes are stated to be a family of software development methodologies in which software is produced in short releases and iterations, allowing for greater change to occur during the design [10]. A typical iteration or sprint is anywhere from two to four weeks, but can vary. The agile methods allow for software development teams to focus on the software rather than the design and documentation [9]. The following list is stated to depict agile methods [9], [10]:

- *Short releases and iterations* - allow the work to be divided, thereby releasing the software to the customer as soon as possible and as often as possible
- *Incremental design* – the design is not completed initially, but is improved upon when more knowledge is acquired throughout the process
- *User involvement* – there is a high level of involvement with the user who provides continuous feedback
- *Minimal documentation* – source code is well documented and well-structured
- *Informal communication* – communication is maintained but not through formal documents
- *Change* – presume that the system will evolve and find a way to work with changing requirements and environments

More specifically, the agile manifesto states:

"We are uncovering better ways of developing software by doing it and helping others to do it.
Through this work we have come to value:
Individuals and interaction over processes and tools
Working software over comprehensive documentation
Customer collaboration over contract negotiation
Responding to change over following a plan
That is, while there is value in the items on the right, we value the items on the left more."

While agile methods are considered as lightweight processes as compared to their predecessors, it has been stated that it sometimes difficult especially after software delivery to keep the customer involved in the process [9]. Moreover, for extremely small software projects, the customer and the user may be one in the same, further complicating the development process. Therefore it is of interest to consider HCI, particularly user-centered design and the benefits it may have if combined with agile methods for software development. The next section introduces the concept of user-centered design.

V. THE USER-CENTERED DESIGN PROCESS

A central theme in HCI is to make the focus of design activity, 'user-centered'. According to human centered design processes for interactive systems, ISO 13407, "Human-centered design is an approach to interactive system development that focuses specifically on making systems usable. It is a "multi-disciplinary activity" [11]. User-centered design (UCD) tends to lead to fewer errors during development and lower maintenance costs over the lifetime of the computer software [12].

In contrast to the traditional methods of software development, user-centered design aims at understanding the user and designing the user interaction through an iterative process. At the center of user-centered design is the user with requirements emerging from user interaction with the system. Since the user-centered design process is an interactive one which allows users to interact with system designers to design a system, ultimately the needs of the user are met.

There are four basic components which help to define interaction [13]. Those components include:

- The end user
- The person who has to perform a particular task
- The context in which the interaction takes place
- The technological systems that is being used

Each of these components has its own qualities and should be considered in the design of the system. The UCD process allows for the exploration of each of these components. As with most methodologies, the UCD process can be broken down into four steps. These steps are analysis, design, implementation and deployment, and are shown in figure 1.

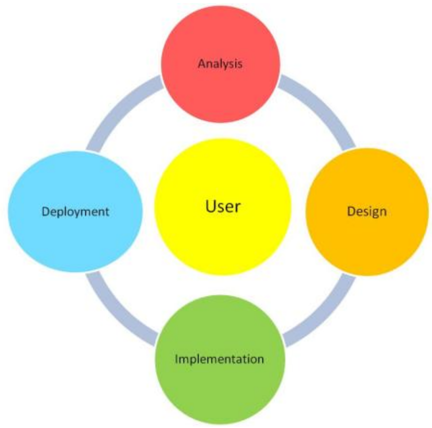


Figure 1. User-centered design model

VI. A PRACTICAL IMPLEMENTATION OF THE AGILE METHOD AND UCD

A. *The project*

The purpose of the project was to combine the principles found in user-centered design with the agile manifesto to develop a website for a customer who was also part of the user group.

B. *The stakeholders*

The stakeholders consisted of two groups: the customer who commissioned the project and the user group who consisted of selected parents and students. The customer was a program manager for a grant obtained to fund a Research Experience for Undergraduates (REU) in an integrative biosciences program at a mid-sized university. The customer has some technical expertise and expressed a desire to be involved in the entire development process. Therefore, to ensure that the customer who was also part of the user group was at the center of the process, bi-weekly meetings were established where updates were provided and prototypes were presented.

C. *The development team*

The development team consisted of two programmers who have expertise in website development and the principles of UCD.

D. *The implementation*

Extreme programming (XP) is probably one of the best known and most widely used agile methods [14], [15]. It was originally designed to address the needs of software development by small teams who faced changing requirements and system environments. XP was used in this empirical inquiry because it reflected the four following principles:

- Incremental development is supported through small, frequent releases
- Customer involvement is integral and supported throughout the process
- People are the main focus of the process not the development process
- Change is embraced as prototypes were constantly released to the user
- The design for the website was simple

XP was also used because it incorporates the concept of collaborative working. The most extensively investigated practice of XP is perhaps pair programming.

The basic premise of pair programming is that a pair of developers, work together during the development process. The developers sit as the same computer and develop the software. There have been several studies that have confirmed that pair programming is effective and can lead to better quality software [9]. However, some studies suggest that with more experienced programmers there is a loss of productivity [16]. Further in a study of nearly 500 students it was found that the stronger of the pair did most of the work, while the weaker of the pair did not improve in programming skill [17].

Yet, it was decided that pair programming would be used because it fosters communication between the team members working on the website and it supports the idea of collective ownership and responsibility. Moreover because the team consisted of only two members with similar programming backgrounds, pair programming proved to be a natural fit.

The first step in the project was to design user stories. User stories are requirements which can be implemented into a series of tasks [9]. User stories are often thought of as high-level requirement artifacts. There are several things to consider when developing user stories which include [18]:

- Stakeholder/customers write the user stories
- Simple tools like index cards to capture thoughts should be used
- The stories can be used to describe a variety of requirements
- Time for the pair programmers to implement the story should be considered
- Priority regarding implementation should be considered

In order to develop the user stories, the team met with the user group who supplied the information and the content

 ISBN: 978-1-61208-117-5

for the website. An example of a user story that was created for the website is found in figure 2.

Submitting the application
A student has decided to apply for the REU program. The application is an editable .pdf file that the student should be able to edit, complete, and submit online. The student may choose to print the application and mail the application to the program manager. The system should allow for online submissions as well as printing the hard copy for mailing.

Figure 2. Story card

After the user stories were developed, the story cards were broken into tasks and the user group was asked to organize the tasks according to priority of what should be implemented first. The objective of this step was to determine the resources needed for implementation. At the completion of this planning process, there were approximately twenty story cards with varying requirements which were organized according to priority.

In the next phase of the project, the development team began implementing the stories according to priority. It was imperative to the customer that the application for the program be the first item implemented. Once this was implemented, the prototype was delivered to the user group. The following is a timeline for the releases provided to the users. The project began September 2009.

TABLE 1. RELEASE TIMELINE

RELEASE	DELIVERY WEEK
Application	2
Homepage	3
Revised homepage	4
Sample project page	6
Revised sample project page	9
Pictorial from previous REU program	12
Resources/contact page	14
Delivery of completed website	16

VII. DISCUSSION

In this instance, the hybrid approach using the agile method and user-centered design for this small project was easy to implement. This section discusses the results from the study.

An exit interview with the user group revealed that they:

- Enjoyed being involved in the process
- Felt that their needs were being met
- Liked the idea of incremental releases

However, it was also noted that:

- The process was time consuming
- While there was some level of satisfaction with the progress of the project as the incremental releases were being delivered and the prototype was being used, after many weeks of meeting and seeing only a release, it was stated that it would be good "just to see the finished product"
- Confusion was also expressed with many technical aspects of the implementation

An exit interview with the development team revealed the following:

- Development was easier as they received immediate feedback from the user group
- Liked the interaction with the user group
- Appreciated the concept of pair programming

However, the team also stated:

- It was difficult to schedule meetings with the customer because of busy and conflicting schedules
- The users did not always communicate their ideas correctly which required rework of the prototype
- It was time consuming to meet for the pair programming experience due to busy and conflicting schedules as the website project was not the only project on which the individuals were working

VIII. CONCLUSION

The aim of this paper was to present the results from an empirical inquiry that focused on answering the question of how the concepts of agile methods and user-centered design could be combined to heighten user involvement in a small-scale software development project (i.e. website development). The author acknowledges that while there are many website development processes, there is inconsistency concerning the steps of the processes and many do not focus on a formalized method for actively involving the user. Consequently, the goal of the paper was to identify the broad steps involved in both agile methods, especially extreme programming, and in user-centered design and to explain how these steps could be used to create a valuable user and developer experience.

Results from the study revealed that using the agile method and user-centered design for small-scaled projects is easy to implement; however, there are certain challenges. While the user group enjoyed being a part of the process, they were overwhelmed by the involvement and certain technical aspects of development activities. Additionally,

 ISBN: 978-1-61208-117-5

the hybrid approach proved to be time consuming for both the user group and the development team.

Future work from this study includes adapting the hybrid approach to other small-scale software projects to ascertain if the type of software being developed determines the outcome of the project. Furthermore, the author intends to develop a case study specific to implementing XP and UCD.

The impact from this empirical inquiry is far reaching. It expands the dialogue that already exists among HCI researchers on how to effectively involve the user in development activities so that it is an enriched experienced. Furthermore, the study provides a foundation for future work on how light-weight software development methodologies and HCI activities can be combined for use in small-scale projects. In conclusion, as systems become more complex and user skill level decreases, it is important that designers of technology find more ways to create unique development experiences that meet both the needs of the user and the development team.

REFERENCES

[1] Internet users as percentage population. http://www.geohive.com/charts/ec_internet1.aspx (Accessed December 20, 2010).

[2] A. Dix A., J. Finlay G.B. Abowd and R. Beale. (2004). *Human-Computer Interaction.* Prentice Hall, 0130-461091, Boston, MA.

[3] Benyon, D; Davies, G; Keller, L.; Preece, J & Rogers, Y. (1998). *A Guide to Usability*, Addison Wesley, 0-201-6278-X, Reading, MA.

[4] Sharpe, H.; Rogers, Y. & Preece, J. (2007). *Interaction design: beyond human-computer interaction 2nd ed.* John Wiley & Sons Ltd, 978-0-470-01866-8, England.

[5] Preece, J.; Rogers, Y.; Sharp, H.; Benyon, D.; Holland, S. & Carey, T. (1994). *Human-Computer Interaction*. Addison Wesley, 0-201-62769-8, Reading, MA.

[6] “Computer.” Def. 1. (2005). *Random House Unabridged Dictionary*. 0-375-40383-3, New York, NY.

[7] Norman, D. (1998). The Design of Everyday Things. MIT Press, Cambridge, MA.

[8] Raskin, J. (2000). *The Humane Interface.* Addison Wesley, 0-2-1-37937-6, Boston, MA.

[9] I. Sommerville. (2011). *Software Engineering 9th Ed*. Addison Wesley, 13:978-0-13-703515-1, Boston, MA.

[10] Tsui, F. and O. Karam. (2011). *Essentials of Software Engineering 2nd Ed*. Jones and Bartlett Publishers, 13:978-0-7637-8634-5.

[11] *International Standard ISO 13407 (1999).* http://zonecours.hec.ca/documents/A2007-1-1395534.NormeISO13407.pdf. (Accessed on October 1, 2009).

[12] Schneiderman, B. 2005. Designing the User Interface 4th ed. Boston: Addison Wesley.

[13] H. Sharpe, Y. Rogers and J. Preece. (2007). *Interaction design: beyond human-computer interaction 2nd Ed.* John Wiley & Sons Ltd, 978-0-470-01866-8, England.

[14] Beck, K. (1999). Extreme programming explained: Embrace the change. Addison Wesley.

[15] Jefferies, R., A Anderson, C. Hendrickson. (2000). Extreme programming installed. In: The XP Series. Addison Wesley.

[16] Parrish, A. R. Smith, D. Hale, and J. Hale (2004). “A field study of developer pairs: Productivity impacts and implications.” *IEEE Software* 21 (5), 76-9.

[17] J. Schneider and L. Johnston. (2005). “eXtreme Programming – helpful or harmful in educating undergraduates?” The Journal of Systems and Software 74, 121-132.

[18] Agile Modeling (AM) Home Page: Effective Practices for Modeling and Documentation http://www.agilemodling.com (accessed February 15, 2010).

Personality and Mental Health Assessment: A Sensor-Based Behavior Analysis

Javier Eguez Guevara, Ryohei Onishi, Hiroyuki Umemuro
Department of Industrial Engineering and Management
Tokyo Institute of Technology
Tokyo, Japan
eguezguevara.j.aa@m.titech.ac.jp, ryh0024@yahoo.co.jp, umemuro.h.aa@m.titech.ac.jp

Kazuo Yano, Koji Ara
Central Research Laboratory
Hitachi, Ltd.
Tokyo, Japan
kazuo.yano.bb@hitachi.com, koji.ara.he@hitachi.com

***Abstract*—The purpose of this study was to estimate personality and mental health through behavior data measured by acceleration and voice intensity sensors. Results showed significant correlations between behavior and all personality and mental health traits studied except for openness. This methodology ascertained an effortless assessment of personality and mental health, which respects employee's privacy, and keeps up-to-date companies' workforce information.**

Keywords- human behavior; sensory technology; mental health; personality

I. INTRODUCTION

The wide study of personality theory and mental health has opened research directions to better understand personnel psychology. Companies are increasingly aiming to develop their human workforce by studying their employees' individual characteristics. Robbins [1] identified four individual-level variables, i.e. biographical characteristics, ability, personality, and learning, which have effects on employee performance and satisfaction.

However, since the use of questionnaire-based objective tests for both personality and mental health has been widely established [2], the time required for employers and employees to carry out such questionnaires is increasingly becoming wasteful and troublesome. On the other hand, recent technology enables to visualize office workers' interactions [3], identify human behavior within organizational situations and obtain associated tacit knowledge [4][5] without privacy intrusion or major burden.

The purpose of this study was to propose a sensor-based methodology to estimate behavior and to further demonstrate existing correlations between employee's behavior and, mental health and personality traits. Building on the aforementioned findings, this study provides conclusions involving personality traits and mental health in the working setting with a minimum required burden from both employers and employees.

This paper is organized as follows. Section II contains an overview of related personality and mental health studies in the workplace. In Section III, it is proposed a methodology to estimate behavior based on sensory data, which was used in Section IV to analyze the relationship with personality and mental health. Section V includes a summary of the paper and presents future work.

II. RESEARCH ON PERSONALITY AND MENTAL HEALTH

Personality has been defined as the characteristic manner in which one thinks, feels, behaves, and relates to others [6]. Robbins [1] claimed that all our behavior is at some extent explained by our personalities and experiences. Traits in personality psychology have been used to describe consistent inter-correlated behavior patterns [7]. The study of personality traits has increased the understanding of the differences between people's behavior in order to explain how certain personality traits better adapt for certain job types [1][8], how personality relates to the effective performance of teams [8][9][10], and how personality is a component of motivation [11].

With similar attention, mental health in the workplace has also been studied. A study has concluded that adverse psychosocial work conditions are predictors of depression worsening [12]. This result was independent from personality traits analyses, and demonstrated the importance of the study of mental health alone.

A. *Personality traits in this study*

The Five Factor Model (FFM) is a taxonomy, or descriptive model, of personality traits organized at the broadest level of abstraction in five factors or dimensions named: extraversion or surgency, agreeableness, conscientiousness, emotional stability versus neuroticism, and intellect or openness [7]. These traits became eventually known as the Big Five [13]. Extraversion describes traits relating energy, dominance, sociability, and positive emotions. Agreeableness includes traits such as altruism, tender-mindness, trust and modesty, defining a prosocial orientation towards others. Conscientiousness summarizes traits which facilitate goal-directed behavior. Neuroticism describes anxiety, sadness or irritability, contrasting emotional stability. Finally, openness describes the depth of an individual's mental and experiential life [14].

Locus of Causality traits are related to the motivation factor of an individual and it examines the source of the motivation when engaging on an activity. Locus of Causality's intrinsic motivation refers to doing something because it is inherently interesting, fun, or enjoyable. On the other hand, extrinsic motivation refers to doing something because it leads to a separable outcome, or because it responds to external demands. Each motivation trait has two secondary scales. Secondary scales for intrinsic motivation

 ISBN: 978-1-61208-117-5

are enjoyment and challenge. Challenge orientation is related to problem-solving, while enjoyment orientation is related to writing and art involvement. Secondary scales for extrinsic motivation are outward and compensation scales. Outward motivation entails personal endorsement and a feeling of choice. Compensation, on the other hand, involves mere compliance with an external control [15][16].

General Causality Orientation is referred as the individual differences that can be characterized in terms of people's understanding of the nature of causation of behavior [17]. In other words, these traits characterize the degree to which human behaviors are volitional or self-determined. There are three causality orientations, namely, autonomy, control, and impersonal orientation. Autonomy orientation trait involves a high degree of experienced choice related to the initiation and regulation of one's own behavior. Control orientation trait involves people's behavior following controls either in the environment or inside themselves. Impersonal orientation trait involves people's experiencing their behavior as being beyond their intentional control [17].

Self-monitoring people are described as showing considerable adaptability and behavior flexibility to external factors, being capable of behave differently in different situations [1].

Type A personality is the trait describing people which is aggressively involved to achieve more in less time. Highly rated Type A people are highly competitive, cannot cope with leisure time, and are continuously measuring their success [1].

B. *Mental health in this study*

The mental health statuses considered for this study were depression and happiness. The viewpoint from which these traits were analyzed was to relate depression, and stress, against job satisfaction characterized by happiness in the workplace.

III. Study 1: Estimation of Behavior Based on Sensory Data

This first study proposed a methodology to estimate human behavior at the workplace based on objective data measured by sensors. An experiment was done in order to investigate the possibility of identifying human behavior.

A. *Participants*

Two male participants volunteered for this experiment. They were aged 25 and 44, and were all in sound health condition.

B. *Apparatus*

Business Microscope (BM) [5] developed by Hitachi Corporation was used in this experiment. Users wore BM like a neck-hanging name-tag, and BM records data from its acceleration, face-to-face (IR), temperature, and voice intensity sensors.

C. *Procedure*

Each participant wore a BM device and acted out behavior categories for 2 hours as if they were engaging in daily office working activities. The characterized behavior categories were walking, talking, desk working, and not-working related behaviors like sleeping, eating/drinking, or simply being unoccupied. These last were grouped, and hereafter referred as idle category.

D. *Measurements*

For the purposes of this study, this experiment only used acceleration and voice intensity sensory raw data with a sampling frequency of 50 Hz. Due to privacy concerns and to a limitation of energy consumption, each datum was observed for 2s long and was acquired once every 10s. The data captured by the device was wirelessly transferred to a server where it was stored.

While wearing a BM device, participants acted out behaviors, switching them from one to another. Every time they switched behaviors, participants marked the time, the location, the posture, and the behavior being acted.

E. *Results*

Sensor data chosen from each behavior category was plotted for analysis to reveal distinctive characteristics representing each acted behavior. Such information was used to build a method to estimate behavior. Such behavior was compared with the actual behavior characterized by participants to find out *hit* and *false alarm* rates. A hit was defined by corresponding predicted and actual behaviors. False alarm on the contrary was defined by a mismatch between them.

A graphic method, the Receiver Operating Characteristic (ROC), was used to evaluate and compare the performances of signal-noise discrimination [18]. ROC was used to portray the optimal criteria to detect behaviors and to select the most effective prediction thresholds.

1) *Behavior detection criteria*

In the walking category plot the amplitude and the frequency of the oscillations of acceleration data were calculated. The amplitude of the curve was calculated with the subtraction of the curve's minimum data value from the maximum data value. As for the number of oscillations of the curve, it was used the zero crossing method. The zero cross line was determined as the data's average line. The number of times the curve crossed the zero-crossing line were added up to obtain the curve's frequency.

Sound intensity curve was represented by temporal changes of sound volume. The data's mean and the standard deviation was calculated and used to obtain thresholds for data characterized by sound representing a talking behavior.

For desk working behavior plot, back and forth acceleration data was investigated. It was found that the mean of acceleration data at time *t*, and the mean of acceleration data at time *t-10s*, tended to be comparable. As differences in mean values of these succeeding two time

 ISBN: 978-1-61208-117-5

points were limited in range, it was assumed that such behavior corresponded to small posture changes as those displayed by desk working behavior. Upper and lower limits of this range were calculated. Data found within this range was regarded as desk working behavior data.

As for the idle behavior category, the acceleration data's frequency of vibration was analyzed. The most suitable data for analysis was found along the acceleration's vertical direction; therefore the zero crossing number was used to calculate this behavior's data frequency.

2) Sequential detection method

The ROC curves for each behavior category are shown in Figure 1. The variance of dots in each graph represents the performance of criteria using various threshold combinations. The results showed that the best detection performance (represented by a red dot) was found in the following order: walking behavior category, talking, desk working, and idle behavior. For this reason it was adopted a sequential detection order which set the sensitivity [18] of each detection method as the detection order priority. Thus, walking behavior was the first category to be detected from the entire sensor data set. From the remaining data, talking behavior was detected, then desk working, and finally idle behavior category.

IV. STUDY 2: STUDY OF PERSONALITY AND MENTAL HEALTH BASED ON BEHAVIOR DATA

The purpose of this study was to analyze the relationships among personality and mental health, and behavior estimated by the method proposed in Section III. An experiment was conducted, and a correlation analysis was done in order to validate that sensory data can be used to assess personality and mental health.

A. Participants

Ninety two Japanese participants, 77 males and 15 females, ranging between 21 and 61 years old (M = 35.93, SD = 8.50), who worked as software developers at a certain company volunteered for the experiment. They were all capable of moving freely and perform routine office activities.

B. Apparatus

The apparatus for this study were the same as those used in study one. Refer to Section III.B.

C. Procedure

Participants wore individual BM devices every working day for 71 days, time in which they engaged in normal daily working activities. Participants' behavior was detected according to the procedure explained in Section III. Also participants conducted 8 sets of questionnaires, 5 relating personality, and 3 more relating mental health.

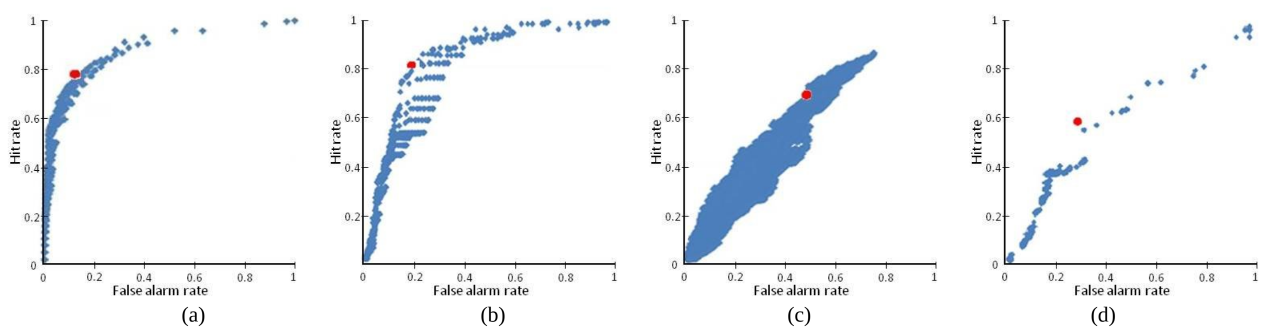
D. Measurements

To assess the big five personality, the Big Five Inventory [14], which consisted of 44 items, was used. The Work Preference Inventory (WPI) which consists of 30 items was used to assess Locus of Causality [16]. The 12-item General Causality Orientation Scale Questionnaire (GCOS) was used to assess General Causality Orientation [17]. It was also used the Self Monitoring trait questionnaire developed by Lennox and Wolfe [19], and the Type A questionnaire developed by Bortner [20].

As for mental health, two scales for depression and one for happiness were used. Although these states could be related to a general happiness scale, in this study the term mental health was used to describe them. The Center for Epidemiology Studies Depression Scale (CES-D) was developed by Radloff [21], and consisted of 20 items. Also the Beck Depression Inventory Second Edition (BDI-II) was used and consisted of 21 items. It is a self-administered questionnaire assessing the severity of depression in adults and adolescents [22]. The last mental health trait studied was satisfaction. The Oxford Happiness Questionnaire (OHQ) was used to assess this trait [23].

E. Results

This study considered unitary behavior samples and behavior events as measurement units. A behavior sample was defined as each datum in a set of data corresponding to an estimated behavior category (one sample per 10s). A behavior event was defined as the sequential group of 2 or more samples under the same behavior category. A chronological summary showing the time series of estimated behavior samples and events was prepared for each participant. This summary indicated what type of behavior category a participant engaged in and for how long.



(a) (b) (c) (d)

Figure 1. ROC curve for (a) walking detection, (b) talking detection, (c) desk working detection, (d) idle detection.

 ISBN: 978-1-61208-117-5

1) Detected behavior

After analyzing the number of behavior samples of the 92 participants, three outlier participants were excluded as they provided significantly less number of samples (< *M* – 2 x *SD*) due to their absence in the experimentation settings. The results hereafter report data from the remaining 89 participant, 74 males and 15 females (*M* = 36.07, *SD* = 8.56). From the total number of detected samples, 28% were detected as walking, 15% as talking, 30% as desk working, and 21% were detected as idle behavior. There were a 5% of samples which could not be detected as any of the proposed behaviors.

2) Behavior characteristic variables

Characteristic variables were obtained for the behavior samples and events of individual participants. The behavior characteristic variables (BCVs) were represented by letters triplets. The first letter of each triplet represented the behavior categories. This is W, T, D, and I represented walking, talking, desk working, and idle behaviors, respectively. The second letters in a triplet were A, T, E, and D, and represented time instances. A as a triplet's second letter represented the entire time span. T as a triplet's second letter represented the time ratio per day. This ratio was obtained by dividing a given behavior total time over the total time in a day. The letter E as a triplet's second letter represented the number of events per day. The letter D as the second letter in a triplet represented the behavior events duration. If the second letter of the triplet was A, the third letters of a triplet were T or E. In this case T represented the time ratio, and E represented an event. However if the second letter of the triplet was T, E, or D, then the third letter of a triplet were A, D, or M, which stood for average, standard deviation, and median.

Personalities like intrinsic or extrinsic motivation are estimated to be related to the variation of behavior. In order to assess this variation, the percentage of behavior-engaged time over all the experiment's time span and the number of events per day was calculated. It was also calculated the average, standard deviation and median of behavior-engaged time (time ratio T and number of events E) per day.

The concept of absorption is important for some personality traits and it is considered to be strongly related with uninterrupted behavior engagement. Therefore the average, standard deviation and median of the time continuance of each behavior were also calculated (triplets with "D" as the second letter).

3) Correlation among personality traits, mental health, and BCVs

Big Five personality scores and BCVs combinations whose correlations were significant are shown in Table I. Extraversion showed positive correlation with walking behavior variables. It was also found negatively correlated with talking events (TEA, TEM) and desk working related variables (DAE, DAT, DEA, DTA, DTM). These results suggested that people who often walk, and often spent their time away from their desks were likely to be extraverted.

TABLE I. PEARSON'S CORRELATION BETWEEN BIG FIVE PERSONALITY SCORES AND BCVS

	E	A	C	N	O
WAT	0.212 *	0.082	-0.142	0.006	0.076
WED	0.235 *	0.089	-0.109	0.118	0.147
WTA	0.216 *	0.091	-0.141	0.009	0.078
WTD	0.255 *	0.221*	0.015	-0.010	0.181
WTM	0.239 *	0.096	-0.130	0.010	0.081
WDA	0.238 *	0.120	-0.081	-0.102	0.059
WDM	0.226 *	0.128	-0.089	-0.039	0.104
TEA	-0.216 *	-0.164	0.032	0.129	0.018
TEM	-0.213 *	-0.140	0.043	0.110	0.027
TDD	-0.131	0.064	0.209 *	-0.041	-0.071
DAE	-0.298**	-0.029	-0.111	0.184	-0.002
DAT	-0.285**	0.000	-0.039	0.058	-0.142
DEA	-0.210 *	0.048	-0.132	0.216*	0.077
DEM	-0.204	0.099	-0.118	0.233*	0.078
DTA	-0.284**	-0.008	-0.036	0.059	-0.158
DTM	-0.273**	-0.007	-0.039	0.056	-0.154

n=89; **p<.01; *p<.05

E: extraversion, A: agreeableness, C: conscientiousness, N: neuroticism, O: openness

Intrinsic Locus of Causality personality scores and BCVs combinations whose correlations were significant are shown in Table II. Talking related variables were negatively correlated with intrinsic locus of causality, and both of its subscales, enjoyment and challenge. It can be argued that intrinsically motivated people have a strong preference for working individually without talking or interacting with people around. However, as it is shown in Table II, idle behavior variables were found positively correlated with challenge subscale alone. It might be argued that the nature of intrinsic locus of causality and challenge orientation, motivate these people to find time to think and reflect about their own initiatives.

Other results showed that extrinsic motivated people did not tend to stay in their desks or focus on their work for long periods as desk working related variable (DAT, DTA, DTM, DDM) were all negatively correlated with extrinsic locus of causality and compensation subscale.

TABLE II. PEARSON'S CORRELATION BETWEEN INTRINSIC LOCUS OF CAUSALITY AND SUBSCALES PERSONALITY SCORES AND BCVS

	Intrinsic	Enjoyment	Challenge
WTD	0.203	0.126	0.209*
TEA	-0.254*	-0.216*	-0.230*
TEM	-0.220*	-0.174	-0.214*
TTA	-0.211*	-0.210*	-0.173
TTM	-0.212*	-0.222*	-0.166
TDM	-0.226*	-0.231*	-0.174
DEA	-0.217*	-0.048	-0.283**
DEM	-0.168	0.005	-0.246*
IAT	0.280**	0.157	0.298**
ITA	0.285**	0.169	0.298**
ITD	0.209*	0.136	0.225*
ITM	0.264*	0.151	0.272*
IDA	0.239*	0.119	0.277**

n=89; **p<.01; *p<.05

 ISBN: 978-1-61208-117-5

General Causality Orientation personality scores and BCVs combinations whose correlations were significant are shown in Table III. Talking related variables correlated negatively with the autonomy trait. On the other hand, impersonal trait correlated positively with desk working and idle behavior related variables. It can be argued that people who do not actively engage or face external circumstances tend to spend their time at their desks.

These results implied that people with high impersonal score, whose behavior is marked by decisions beyond their control, tend to follow directions as they are told. In other words, these people might not leave their desks or stop working. These findings are comparable to idle behavior variables being positively correlated with impersonal trait. It can be argued that as these people did not leave their desks, they might have time to loosen up, even in front of their desks.

Self monitoring trait was positively correlated with walking related variables (WAT, WTA, WTM, WDA, WDM). This suggested that people with high sociability skills, or those rating high in self monitoring, engage for longer periods in walking behavior. High self-monitoring rated people are able to show striking contradictions between their public persona and their private self [1]. Thus, by the fact that self monitoring correlated negatively with talking related variables (TEA, TEM) it can be argued that even though these people regulate their behavior by walking or interacting with others, they might be reluctant to show their opinions by an apprehension of social disapproval.

Type A trait correlated positively with the number of walking events per day (WEM). This suggested that people who tended to walk more often are likely to be competitive or involved in achieving more in less time. It might be argued that these people are often walking around, looking for self-improving opportunities.

Mental health scores and BCVs combinations whose correlations were significant are shown in Table IV. Both depression scales utilized in this study presented similar results which highlighted positive correlation with talking, desk working, and idle behavior related variables (TEA, TEM, DAE, DEA, DEM, IEA). These results suggested that people who more often engaged in talking, desk working, and idle behaviors present higher depression or stress scores. By the fact that BDI-II depression scale positively correlated with walking idle events related variables (WEA, IEM) it can be argued that both, unoccupied behavior people or persistently walking people, might display high work depression or stress.

The OHQ results showed that high talking, and desk working behavior people often showed frustration or discontent (TAE, TAT, TEA, TEM, TTA, TTM, TDM, DAE, DEA, DEM). It can be argued that people who talked for longer periods, would be able to cope with dissatisfaction.

TABLE III. PEARSON'S CORRELATION BETWEEN GENERAL CAUSALITY ORIENTATION PERSONALITY SCORES AND BCVS

	Autonomy	Control	Impersonal
WED	0.163	0.233*	0.085
WTD	0.231*	0.133	-0.041
WDA	0.175	-0.120	-0.240*
WDD	0.148	-0.244*	-0.182
TAE	-0.238*	0.000	0.101
TAT	-0.273**	-0.016	0.065
TTA	-0.278**	-0.029	0.060
TTD	-0.305**	-0.034	0.054
TTM	-0.254*	-0.032	0.051
TDA	-0.216*	-0.029	0.009
TDM	-0.292**	-0.096	0.034
DAE	-0.016	0.059	0.304**
DEA	0.041	0.054	0.331**
DEM	0.083	0.087	0.323**
IEA	0.077	0.117	0.296**
IED	0.148	0.222*	0.165
IEM	0.100	0.147	0.301**

n=89; ***p*<.01; **p*<.05

V. CONCLUSION AND FUTURE WORK

The present study proposed a methodology to estimate personality from sensory data information. However studies pertaining personality with emphasis to the workplace are numerous, the established measuring method used by those studies were questionnaire tests. This study built up a clear methodology through which personality is estimated unobtrusively and without the need of questionnaires, through the use of acceleration and voice sensory information.

In Study One it was effectively detected walking, talking, desk working, and idle behaviors, with hit and false alarm rates of 0.78 and 0.12, 0.82 and 0.19, 0.69 and 0.48, and 0.59 and 0.28, respectively. In Study two, the correlation analysis showed significant correlations between behavior and all personality and mental health traits studied except for openness. Personality variables that showed significant correlations with greater extent of behavior variables were extraversion, intrinsic motivation, challenge, and happiness.

Also, the behavior category which showed significant correlation with the greater number of personality variables

TABLE IV. PEARSON'S CORRELATION BETWEEN MENTAL HEALTH SCORES AND BCVS

	CES-D	BDI-II	OHQ
WEA	0.103	0.221*	-0.083
WDA	-0.174	-0.139	0.237*
TAE	0.204	0.144	-0.215*
TAT	0.191	0.166	-0.235*
TEA	0.282**	0.267*	-0.287**
TEM	0.250*	0.271*	-0.267*
TTA	0.188	0.158	-0.228*
TTM	0.176	0.151	-0.211*
TDM	0.072	0.152	-0.270*
DAE	0.239*	0.254*	-0.255*
DEA	0.283**	0.371**	-0.283**
DEM	0.211*	0.355**	-0.255*
IEA	0.223*	0.264*	-0.156
IEM	0.173	0.246*	-0.119

n=89; ***p*<.01; **p*<.05

was desk working behavior category revealing 31 significant correlations; while the behavior category which showed significant correlation with the least number of personality variables was idle category with only 18 significant correlations.

The results in this study suggest that it is possible to effortlessly assess personality and mental health, respecting the privacy of employees, and without the need of questionnaires. What has been argued as a benefit of the use of questionnaires (greater choice) is a major weakness; the use of questionnaires allows for questionnaire items' omission or misrepresentation, thus affecting the overall effectiveness and goals of the assessment. This limitation affects the informant himself who is the ultimate beneficiary of the research efforts. Furthermore, personality is continuously shaped by experiences, and thus questionnaires are limited to cope with personality's changing nature. As the methodology presented in this study is set by continuously loading data, the personality and mental health information obtained will always provide up-to-date information. In addition, saving employers' and employees' time, is yet another benefit proposed by this study, which opens a new behavior estimation research direction, and thus its continuation is essential. Future studies should deepen this study's findings: it should consider additional working settings; the improvement of the behavior detection method including participants from both genders, and a larger set of behavior categories.

ACKNOWLEDGMENT

This research was based on the data provided by the World Signal Center, Hitachi, Ltd.

REFERENCES

[1] S. Robbins, "Foundation of Individual Behavior," in *Organizational Behavior*, 8th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1998, pp. 40-87.

[2] D. G. Winter and N. B. Barenbaum, "History of modern personality theory and research," in *Handbook of Personality*, L. A. Pervin and O. P. John, Eds. New York: The Guilford Press, 1999, pp. 3-27.

[3] J. Nishimura, N. Sato, and T. Kuroda, "Speaker siglet detection for Business Microscope," in *Proc. 7th International Conference on Machine Learning and Applications*, San Diego, 2008, pp. 376-381.

[4] K. Ara, N. Kanehira, D. Olguin Olguin, B. N. Waber, T. Kim, A. Mohan, P. Gloor, R. Laubacher, D. Oster, A. Pentland, and K. Yano, "Sensible organizations: Changing our business and work styles through sensor data," *J. Information Processing*, vol. 16, pp. 1-12, 2008.

[5] K. Yano and H. Kuriyama, "Human x sensor: How sensor information will change human, organization, and society," *Hitachi Hyouron*, vol. 89, No. 07, pp. 62-67, 2007.

[6] T. A. Widiger, R. Verheul, and W. van den Brink, "Personality and psychopathology," in *Handbook of Personality*, L. A. Pervin and O. P. John, Eds. New York: The Guilford Press, 1999, pp. 347-366.

[7] O. P. John and S. Srivastava, "The Big Five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of Personality*, L. A. Pervin and O. P. John, Eds. New York: The Guilford Press, 1999, pp. 102-138

[8] M. K. Mount and M. R. Barrick, "Five reasons why the 'Big Five' article has been frequently cited," *Personnel Psychology*, vol. 51, pp. 849-857, 1998.

[9] R. R. Reilly, G. S. Lynn, and Z. Aronson, "The role of personality in new product development team performance," *J. Engineering and Technology Management JET-M*, vol. 19, pp. 39-58. 2002.

[10] M. R. Barrick, G. L. Stewart, and M. Piotrowski, "Personality and job performance: Test of the mediating effects of motivation among sales representatives," *J. Applied Psychology*, vol. 87, No. 1, pp. 1-9, 2002.

[11] A. Furnham, L. Forde, and K. Ferrari, "Personality and work motivation," *Personality and Individual Differences,* vol. 26, pp. 1035-1043, 1999.

[12] S. Paterniti, T. Niedhammer, T. Lang, and S. M. Consoli, "Psychosocial factors at work, personality traits and depressive symptoms," *British Journal of Psychiatry,* vol. 181, pp. 111-117, 2002.

[13] L. R. Goldberg, "Language and individual differences: The search for universals in personality lexicons," in *Review of personality and social psychology*, vol. 2, L. Wheeler Ed. Beverly Hills, CA: Sage, 1981, pp. 141-165.

[14] V. Benet-Martinez and O. P. John, "Los cinco grandes across cultures and ethnic groups: Mutitrait multimethod analysis of the big five in Spanish and English," *J. Personality and Social Psychology*, vol. 75, No. 3, pp. 729-750, 1998.

[15] R. M. Ryan and E. L. Deci, " Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary Educational Psychology,* vol. 25, pp. 54-67, 2000.

[16] T. M. Amabile, K. G. Hill, B. A. Hennessey, and E. M. Tighe, "The work preference inventory: Assessing intrinsic and extrinsic motivational orientations," *J. Personality and Social Psychology*, vol. 66, pp. 950-967, 1994.

[17] E. L. Deci and R. M. Ryan, "The general causality orientations scale: Self-determination in personality," *J. Research in Personality*, vol. 19, pp. 109-134, 1985.

[18] C. D. Wickens and J. G. Hollands, "Signal detection, information theory, and absolute judgement," in *Engineering Psychology and Human Performance*, 3th ed. Upper Saddle River, New Jersey: Prentice-Hall. 2000, pp. 17-44.

[19] R. D. Lennox and R. N. Wolfe, "Revision of the self-monitoring scale," *J. Personality and Social Psychology*, vol. 46, pp. 1349-1364, 1984.

[20] R. W. Bortner, "A short rating scale as a potential measure of pattern A behavior," *J. Chronic Diseases,* vol. 22, No. 2, pp. 87-91, 1969.

[21] L. S. Radloff, "The CES-D scale: A self-report depression scale for research in the general population," *Applied Psychological Measurement*, vol. 1, pp. 385-401, 1977.

[22] A. T. Beck, R. A. Steer, and G. K. Brown, *Manual for the Beck Depression Inventory-II*, San Antonio, TX: The Psychological Corporation. 1996.

[23] P. Hills and M. Argyle, "The Oxford happiness questionnaire: a compact scale for the measurement of psychological well-being," *Personality and Individual Differences*, vol. 33, pp. 1073-1082, 2002.

Unpacking the Contents

A Conceptual Model for Understanding User Experience in User Psychology

Rebekah Rousi
Agora Center
University of Jyväskylä
rebekah.rousi@jyu.fi

Pertti Saariluoma
Agora Center
University of Jyväskylä
pertti.saariluoma@jyu.fi

Jaana Leikas
VTT Technical Research Centre of Finland
jaana.leikas@vtt.fi

Abstract—**Paradigm shifts can be noted to have taken place in several areas of user-technology research. The most obvious have been in terms of including users within the design process, either in terms of usability studies or user experience design. There have also been shifts towards viewing human-technology interaction as not only an optical experience, but also an embodied one. When exploring these factors it is easy to prioritize the physical over the psychological. User interactions with systems are more easily measured in terms of concrete outcomes rather than by subjective feelings and perceptions of interaction. Through the conference theme: user modeling and user focus, this study's purpose has been to uncover mental contents present during the moments of human-system interaction. The study has employed a range of design stimuli for users to encounter and evaluate, giving a holistic idea of the psychological components involved in the interactions. The article describes a conceptual model which has derived from a study of mobile phone icons in the context of their graphical user interfaces. This study shows that users draw on multiple dimensions of mental information contents when experiencing technology, these include: cognitive, practical, aesthetic and emotional. Although the dimensions somewhat overlap, shifts can be seen between the dominance of the dimensions when the experience is positive or negative.**

Keywords- user experience; mental contents; user psychology; conceptual model

I. INTRODUCTION

User interfaces (UIs) are the meeting point between technological products and users. Users, or humans, are complex organic systems in their own right. They come with varied physical, emotional and cognitive needs. Numerous factors impact the make-up of these users physically and mentally, and in order to get the connection between the user and technology 'just right', designers must consider and address these factors. As UIs become more advanced, user-centered matters are grow increasingly important. Thus, through 'user modeling and user focus', we aim to show that it is not just the physical and cognitive challenges of UI design that are important, but so too is a more encompassing view on psychological factors embedded within the user and their perception and interaction with the design.

This paper concentrates on the mental contents of user experience (UX). The term 'mental contents,' refers to information representations, which exist within an individual's mind. These representations, while constantly adapting and evolving, shape the way in which people experience phenomena [31][32][34][36]. These mental contents are shaped by a number of factors, none the least, by lived experience. Other factors include: cultural (national, gender, sexuality, religious, sub-cultural and people with disabilities), social, psychological, linguistic and geographical to name a few. The present study addresses two of these factors – national cultural and linguistic. Empirical material was collected in Australia and Finland in 2009 as a part of the Theseus [37] and ITEA2 Easy Interactions [38] projects examining the user psychology of user-system interactions.

We have been surprised to discover that astoundingly little, if anything, has really been done to understand what we consider to be the most important component of UX – mental contents. The study of mental contents is an important component of user psychology. In the next sub-section, user psychology and its application in the fields of human-technology interaction (HTI) is introduced in relation to UX. Section II details the method by explaining the picture technique and its rationale and detailing components such as the participants, measures and procedure. Section III illustrates the results and Section IV outlines a conceptual model of experiential contents. The paper is concluded in Section V, which summarizes and reflects on the findings of this study, posing further questions for future investigation.

A. User psychology in the exploration of user experience

In order to understand the mental processes that occur when users interact with technology, we must first understand the psychological pre-conditions of this interaction. This is what user psychology seeks to explain [1][2][3]. Scholars within the discipline of user psychology examine and construct psychologically justified explanatory models to influence design decisions [3]. With progress being made in the fields of UI development, it is hardly surprising that the area of UX design emerged. Emotional usability pioneer, Don Norman [4] articulates that designers and engineers know what works and users at least generally can learn how to use the products. But what lies at the heart of distinguishing two perfectly functioning products from one another, still remains somewhat of a mystery.

UX expands upon traditional usability studies, and is quite closely related to the Japanese Kansei engineering and North American emotional usability [4][5][6][7][8]. By now, the field of UX includes an abundance of conceptual models [7][9][10][11][12][13]. Inclusive in the models are dimensions such as user HTI perceptions, in addition to understandings of cultural and symbolic human-to-human interaction which impact HTI [14][15][16][17][18]. Important in the UX research paradigm is the quality improvement of HTI. Norman [4] stresses the affective and emotional aspects of interaction. Other scholars such as McCarthy and Wright [19] emphasize the role of culture as a meaning making tool in the process of UX. Battarbee [9] emphasizes the social nature of UX by demonstrating that it is not simply isolated within one individual, rather it can be and/or is a shared experience between multiple persons and communities. The major components of UX articulated by Väänänen-Vainio-Mattila et al. [20] encompass: interactional flow; pleasurable and hedonic aspects of product usage; and multisensory interaction.

The above examples represent studies undertaken in disciplines of cognitive science and task-related experimental analysis, phenomenology and qualitative analysis. They emphasize the role of feelings and emotions, perceptions and behavior in HTI [21]. Another key trait of the above mentioned models is that they broaden the perspective from human-computer-interaction (HCI), to other design products or interactive technologies. Thus, they concentrate on incorporating users' lived experience and natural interaction in the design process [19][22][23][24]. Despite this, attention is still lacking in regards to what we consider the most essential property of experience – information contents. Apart from phenomenology [25][26][27], and some aspects of cognitive models [28][29], information contents of mental representations have rarely been discussed in relation to UX [30]. Experience is the conscious part of human mental representations.

Previous work by Saariluoma [31][32][33][34][35][36] and Leikas and Saariluoma [24] is continued in this study, whereby a content-based approach has been taken towards the investigation of life based thought-related processes. Content-based thinking enables the formulation of new questions. Problems lie within explaining the mental contents of users. Contents are in a constant state of flux from one phase to the next, leading to a core question within our investigation: Given its ever-changing nature, how is conscious experience effectively operationalized, in order to provide a detailed understanding of UX from a psychological perspective?

In light of a study into user evaluations of mobile phone icons, the key aim of this paper is to lay the foundations of a conceptual model based on categorical dimensions of mental contents. The idea is not to provide a finished map of all the dimensions and categories, but to illustrate the formational stages of the framework, starting from a skeletal version of what has been achieved though the examination of user response to graphical user interface icons. The further these studies go into other design elements, bridging newer UI designs and prototypes, the more detailed and thorough the framework becomes.

II. METHOD

For this study, the picture sorts technique [46][47][48] 49] was employed to investigate the ways in which users prioritized and constructed explanations of why specific designs were preferred over others. In addition to carrying on the tradition of investigations which explore user emotional responses to designs, the picture sort technique focuses on collecting explanatory frameworks provided by users. As mentioned above, the challenge for us has been to gain an understanding of *how* participants see and mentally construct design products using the information content available to them.

The picture sorts method, developed during the 1950s [45], is an empirical technique used to explore Personal Construct Theory (PCT). Scholars of PCT advocate that people mentally register phenomena through constructs that they themselves create by means of mental information content. This information content includes and is shaped by social, environmental, cultural and psychological factors etc. In other words, via interaction with design, environments and other people we are constantly constructing and reformulating mental images of the phenomena we encounter. In fact, our sense of reality is based on these constructs. More and more within the field of HTI, the significance of methods such as the picture sort is being recognized [46][47][48][49][50].

A. *Participants*

In total, 35 subjects participated in this study. Fourteen people participated in Australia and 21 participated in Finland. Australian participants were aged from 26 to 61: 2 were 26-29; 3 were 32-35; 3 were 40-44; 4 were 50-54; 2 were 58-61. The mean age of the participants in Australia was 44.2 years old. The gender distribution of the participants in Australia was 8 females and 7 males. Ages of the Finnish participants ranged from 21 to 54: 9 were 21-29; 7 were 30-38; 3 were 43-46; and 2 were 51-54. The mean age of the Finnish participants was 33.1 years old. The gender distribution of the Finnish participants was 14 females and 7 males.

B. Measures

Twenty-two screen shots of icon menus were presented on a pack of picture cards. The icon menus were chosen from competing mobile phone brands, models and generations. The idea was to have a sample representing the most commonly used brands on the Australian and Finnish markets. Once these had been selected they were then printed on 200 gsm matt card. A USB recording device was used to record the experiments.

The participants were required to sort the cards into three piles: least attractive; attractive; most attractive. This was not a time based exercise, participants could undertake the experiment at their own pace to ensure that they carefully looked at the icons. We emphasized that it was *their* personal subjective preferences that we were interested in, and that the experiment was not an examination. Once the participant had finished sorting the cards into piles, they were asked to think of descriptive titles (words or phrases) and then reasons for these titles. The titles and the reasoning were first written down on an open answer questionnaire, and then the participants were asked to verbally elicit their responses. This was in case they were more likely to favor one of the explanatory forms over the other.

C. Procedure

Before conducting the experiments, a Statement of Ethics was applied for and obtained from the University of Jyväskylä, Finland, and the Edith Cowan University, Western Australia. The experiments were conducted in quiet, controlled environments in Finland and Australia. Generally, they took place one per time, but on several occasions there were two participants at different sides of the room. The researcher's role in the experiments was to distribute and explain the experiment components and answer questions. In Finland, a native speaking Finnish research assistant was used to conduct and explain the experiments.

When entering the experiment setting, participants were given an information handout to read about the procedure. Participants were asked to sign a 'Notice of Consent' agreeing to the use of the material obtained during the experiments. Participants completed a personal details form asking: age, gender, cultural ethnicity, highest education level, profession, mobile phone user skills (expert; advanced; intermediate; beginner) and model of mobile phone. The steps of the experiment were explained to the participants. In return for participating in the experiments participants were awarded with one free movie ticket.

III. RESULTS

In an attempt to understand the relationship between particular qualities and positive and negative user experiences, the titles and explanations given by the participants were divided into categories. Through content analysis eight categories arose in the Australian data and nine from the Finnish data. The categories were: aesthetic appeal; clarity; icons, colors and layout; intuitiveness; amounts of icons on screen; understandability; labels; size; and the category of shapes emerged from the Finnish data. The positive and negative descriptions were treated separately when counting and grouping adjectives and phrases.

In the positive descriptions, clarity was the most common category (31%) used by the Australians. Intuitiveness was the next most frequent category (19%). Descriptions relating to intuitiveness included "the need to think" and "informative". Later on in the Finnish results this was mostly linked to the characteristic of familiarity. However, considering this study was supposed to focus on aesthetic attraction, it only featured in 18% of the Australian explanations. Examples of this can be seen in statements such as "good looking", "aesthetically interesting", "aesthetically appealing", "funky" etc. Comments referring directly to the icons themselves, their colors and layout, featured 11% of the time in the Australian results, which was the same as understandability (e.g., easy to understand, more understandable). On the practical side, even though the focus was on the icons, labels were mentioned in 5% of the Australians' positive comments. Interestingly, the practical dimensions (labels, size and amount-arrangement) were mentioned by participants of 50 years of age and over. Their descriptions related more to usability and physical limitations than to aesthetic appeal.

In the Finnish participants' positive evaluations, aesthetic appeal was mentioned 27% of the time. Clarity was the next most frequent category, featuring in 19% of the descriptions. The icons, their colors and layout were the next most frequently used (18%). Intuitiveness, as well as amounts of icons on screen and arrangement were both present in 10% of the comments.

There are some differences between the negative descriptions of icons, colors and layout (Australians 4% and Finns 18%) and negative descriptions regarding size (Australians 12% and Finns 3%). However, one substantial finding is connected to the way that practical usability (clarity) took preference over aesthetic appeal when Australians were positively evaluating icons. In order to see these relationships more clearly, figs. 1 and 2 below illustrate the comparisons between the distribution of adjectives used for positive and negative descriptions, in relation to each national group. This is achieved by showing the percentage of comments (descriptions) which were allocated to each of the adjective categories.

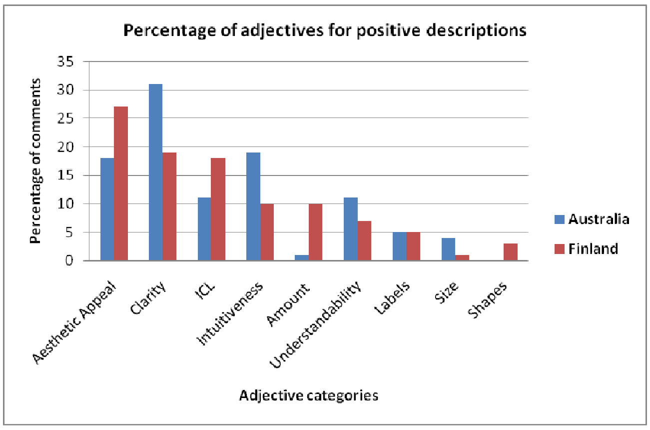


Figure 1. Percentage of adjectives for positive descriptions

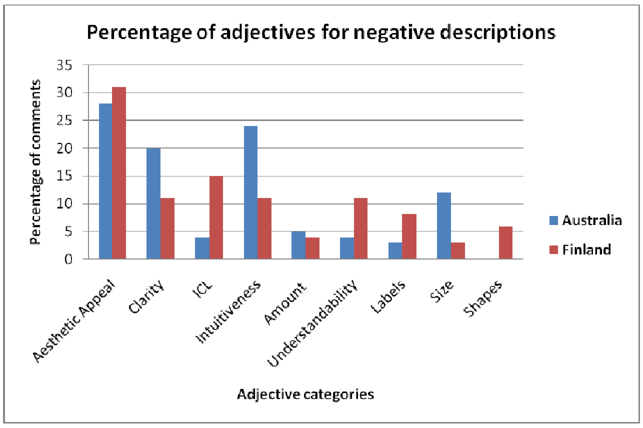


Figure 2. Percentage of adjectives for positive descriptions

The distributions show that content recognized as clarity was present particularly when Australian participants positively experienced icon designs. When negatively experiencing the icons, content associated with aesthetic appeal played an extremely active role amongst participants in both national groups. When considering the dominance of the dimension in positive descriptions, clarity (perceived simplicity and ease-of-use) is a main desired quality for positive UX, even more so than an abstract more subjective conception of aesthetic appeal. This affirms models such as TAM (technology acceptance model) [51][52][53] and their advocacy of the importance of perceived ease-of-use.

IV. Conceptual Model of Experiential Contents

A means by which we may express the mental contents involved in UX is to examine the categorical groups which people generate to justify evaluations. Individual remarks differ from one participant to the other. However, on a higher categorical level, clear dimensions of contents may be extracted from this qualitative data. The information content available enables insight into the formation and design of UX on the basis of it is described by the users themselves.

A. A User Psychology Model of the Categorical Dimensions of UX

In order to make sense of these results, and to create a framework by which future user psychology investigations of this nature may be guided, a model of categorical dimensions of UX has been plotted. This is not a polished product, but the basis upon which a larger framework for psychologically understanding UX may be developed.

From the results, slight tendencies may be observed within the categories which highlight the elements participants focused on during positive or negative interaction. In order to approach the results from a deeper perspective, the categories were divided into adjective dimensions which describe the products' physical attributes, and categories which describe the participants' internal/emotional attributes allocated to the icon designs. The division can be seen in the diagram below.

Fig. 3 below shows the basis of this user psychology model of the categorical dimensions of UX. It is a simplified mind map of the categories extracted from the data, and organized into 4 main dimensions: cognitive, emotional, practical, and aesthetic. Furthermore, the diagram has been divided vertically into the dimensions which can be seen from the point-of-view of the user (top) and the point-of-view of the product or design (bottom).

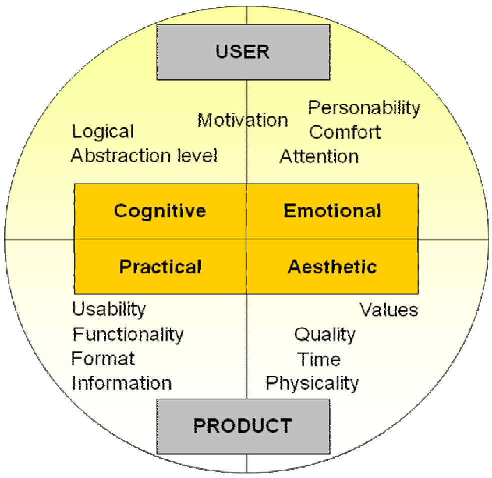


Figure 3. Categorical dimensions of UX

 ISBN: 978-1-61208-117-5

The featured categories noted in the results section were used as a basis to form themes featured in the diagram above. The resulting themes informed our understanding of the categorical dimensions. These dimensions summarize the users' explanatory constructs in terms of designs' practical and aesthetic properties, and the user's cognitive and emotional properties. Theme allocation within these specific dimensions is not a straight forward task. Themes such as motivation and format-physicality, may be applied to multiple dimensions – i.e., motivation, through adjectives such as "invigorating", "stimulating" and "interesting", was allocated to the emotional dimension, but can be attributed the theme of learnability.

Through analysis of positive, semi-positive and negative comments it was observed that positive comments mostly concentrated on the technical and aesthetic dimensions. Participants focused more on how the icon designs worked within themselves, than on how they as users were emotionally affected by the designs. The positive technical comments focused on formatting – such as labels, size, amount of icons. The positive aesthetic comments focused on aesthetic values-qualities and physicality. Fig. 4 below demonstrates this relationship.

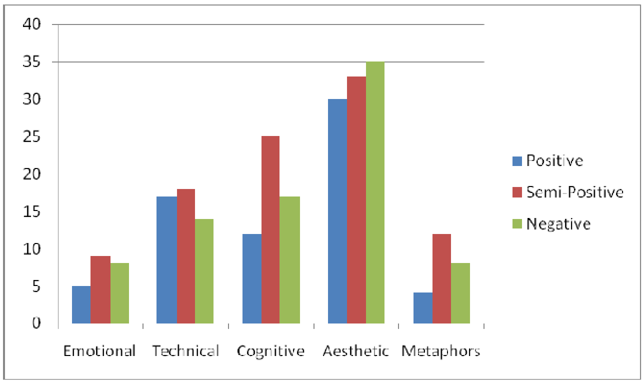


Figure 4. Adjective distribution – positive, semi-positive and negative

The semi-positive comments used to describe the attractive pile seem to emphasize the technical, cognitive and aesthetic dimensions. Thirty-three different semi-positive terms were used within the aesthetic dimension, in comparison to 30 positive and 35 negative comments. The themes emphasized in the semi-positive comments included: emotional – attention, motivation, identification and comfort; technical – usage and format; cognitive – abstraction, motivation and logic; aesthetic – values/quality, time and physicality. Notably, more metaphors were used to describe semi-positive traits than were used in the cases of positive or negative traits. These metaphors included: "basic mass" and "Linux GDM window", describing banality in the icon designs.

As with the positive comments, the negative comments concentrated on the technical dimension. Yet, the emotional dimension played a greater role in the negative (and semi-positive) comments than in the positive comments – i.e., attentional and motivational ("dull", "boring", "annoying"). Moreover, the negative comments emphasized the aesthetic dimension (values and physicality themes).

V. CONCLUSION

Based on the results of this experiment some interesting observations may be made regarding emotional usability. The findings of this experiment show that less emphasis was placed on the user's own emotional responses when positively evaluating icon designs. Emotions came to the fore when designs were experienced as negative. Interpreting design via aesthetic categories was common throughout the process of evaluation. However, the technical dimension was more important in the positive evaluations than the negative evaluations. This may be due to the fact that participants found ease in articulating technical qualities when explaining why they positively experienced certain designs.

A principal theoretical issue has been opened here. This pertains to the way that we can investigate how users encode technical devices; and how we may examine this encoding process. When allowing participants to generate reasoning behind their choices, we are able to glimpse the mental contents involved in the way they emotionally appraise products. These findings articulate: critical UX design characteristics; why they are critical; the complexity and multidimensionality of mental contents; and possibilities for benefiting from the development of design based on our knowledge of mental contents.

We believe that these empirical findings should be considered in terms of content-based psychological thinking [31][33][44]. Speech output reflects conscious experience, behind which are systems of subconscious mental contents. Thus, according to this UX means the conscious experience of encoding objects. People are guided by their mental representations. If information content is understood, then cognitive and emotional responses can be used to explain behavior. Thus, we pose that knowledge of mental contents will improve the analysis of human technology interaction.

We are no longer primarily interested in computations (as seen in Newell and Simon [28] or Fodor [41]). Instead, we investigate mental contents in the late Wittgenstein [54] sense as systems of languages, meanings and thoughts [33][55]. Percepts, concepts, beliefs, mental models, schemas and other forms of representations such as images or emotions have their contents. Therefore, theoretical and explanatory concepts are often the contents and not the format in which contents have been constructed [55][36].

This was a small-scale study designed to develop methods and theoretical frames to measure attractiveness in terms of user psychology. However, as a result of this small sample, questions have emerged regarding the nature of emotional usability in everyday design – should more emphasis be placed on inducing positive emotions, or on reducing negative ones? We also question the relationship between thought and language in such a study, as language never precisely produces the sensations experienced by the user. The results show the complexity of conscious and verbal experience, and moreover, that four major categories give a broad idea about how people classify icons. The key purpose of the underlying conceptual model is to shed light on what people actually experience. This form of information allows for mental representations and life to be connected [23][24]. UX is not an abstract model, but rather, refers to multiple factors in the make-up of a person. This means that mental content always plays a role.

ACKNOWLEDGMENTS

We would like to thank the funders of the ITEA2 Easy Interactions and Theseus projects, Tekes, for enabling us to undertake this research. In regards to experiment development and implementation in Australia, we would also like to thank our collaborators at: the Edith Cowan University; Curtin University of Technology; and Adelaide University.

REFERENCES

[1] A. Oulasvirta, and P. Saariluoma. “Long-term working memory and interrupting messages in human computer interaction,” Behaviour & Information Technology.Taylor and Francis, 2004.

[2] A. Oulasvirta, and P. Saariluoma. Surviving task interruptions: Investigating the implications of long term working memory theory. Amsterdam: Elsevier, 2006.

[3] P. Saariluoma. Explanatory frameworks for interaction design. London: Springer-Verlag, 2004.

[4] D. Norman. Emotional design: why we love or hate everyday things. New York: Basic Books, 2004.

[5] M. Hassenzahl, and N. Tractinsky. “User experience – a research agenda,” Behaviour and Information Technology, vol. 25 issue 2, 2006, pp. 91–97.

[6] ISO DIS 9241-210:2010. Ergonomics of human system interaction – Part 210: Human-centred design for interactive systems. Geneva: International Standardization Organization, 2010.

[7] M. Kuniavsky. Observing the User Experience. San Francisco: Morgan Kaufman Publishers, 2003.

[8] M. Nagamachi. “Perspectives and new trend of Kansei/Affective Engineering,” The TQM Journal, vol. 20 issue 4, 2008, pp. 290-298.

[9] K. Battarbee. Co-experience – understanding user experiences in social interaction. Helsinki: University of Art and Design Helsinki, 2004.

[10] A. Cooper, R. Reimann, and D. Cronin. About Face 3. The essentials of interaction design. Indianapolis: Wiley, 2007.

[11] J. Forlizzi, and K. Battarbee. “Understanding Experience in Interactive Systems,” In Proceedings Designing Interactive Systems: processes, practices, methods, and techniques. Cambridge MA: ACM, 2004, pp. 261–268.

[12] M. Hassenzahl. “The interplay of beauty, goodness and usability in interactive products,” Human-Computer Interaction, vol. 19, 2004, pp. 319–349.

[13] D. Russell. “User experience research group: Understanding the complete user interaction,” In SIGCHI Bulletin vol. 30 issue 2, 1998, pp. 90-94.

[14] L. Arhippainen, and M. Tähti. “Empirical evaluation of user experience in two adaptive mobile application prototypes,” In Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia (MUM 2003), Norrköping, Sweden, 2003, pp. 27-34.

[15] M. Buchenau, and J. Fulton-Suri. “Experience Prototyping,” In Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques. New York City: ACM, 2000, pp. 424-433.

[16] J. Forlizzi, and S. Ford. “The Building Blocks of Experience: An Early Framework for Interaction Designers,” In Proceedings of Designing interactive systems: processes, practices, methods, and techniques. New York: ACM, 2000, pp. 419–423.

[17] A. Kankainen. Thinking model and tools for understanding user experience related to information appliance product concepts. Espoo: Helsinki University of Technology, 2002.

[18] T. Mattelmäki. *Design Probes*. Helsinki: University of Art and Design in Helsinki, 2006.

[19] J. McCarthy, and P. Wright. *Technology as Experience*. MIT Press: Cambridge, 2004.

[20] K. Väänänen-Vainio-Mattila, H. Väätäjä, and T. Vainio. “Opportunities and challenges of designing the service user experience (SUX) in Web 2.0,” In Saariluoma, P. & Isomäki, H. (Eds.), Future interaction design II, London: Springer, 2009, pp. 117-139.

[21] D. Swallow, M. Blythe, and P. Wright. “Grounding experience: Relating theory and method to evaluate the user experience of smartphones,” In Proceedings of the 2005 Conference on European Association of Cognitive Ergonomics. New York: ACM Press, 2006, pp. 91-98.

[22] Hammel, M. “The aesthetics of use,” In Bertelsen, O.W., Petersen, M.G. & Pold, S. (Eds.), Aesthetic approaches to human-computer interaction. Proceedings of the NordiCHI 2004 Workshop, Tampere, Finland. Aarhus: Department of Computer Science, University of Aarhus, 2004, pp. 17-19.

[23] J. Leikas. Life-Based Design - A holistic approach to designing human-technology interaction. VTT Publications 726. Helsinki: Edita Prima Oy, 2009.

[24] J. Leikas, and P. Saariluoma. “'Worth' and mental contents in designing for ageing citizens' form of life,” Gerontechnology, vol. 7, 2008, pp. 305-318.

[25] P. Dourish. Seeking a Foundation for Context-Aware Computing, 2001, doi: 10.1207/S15327051HCI16234_07.

[26] P. Dourish. Where the Action Is: The Foundations of Embodied Interaction. Cambridge, Mass: MIT Press, 2001.

[27] J. Rod. “Post Human-Centered Design Approach for Ubiquity,” In Proceedings of Digital Art and Culture Conference 2009, <http://www.escholarship.org/uc/item/7nx6199f> 13.12.2010.

[28] A. Newell, and H. Simon. Human problem solving. Englewood Cliffs: Prentice Hall, 1972.

[29] S. Payne. “Mental models in Human Computer Interaction,” In A. Sears & J. Jacko (Eds), The human-computer interaction

handbook: fundamentals, evolving technologies. New York: Taylor and Francis Group LLC, 2008, pp. 63-76.

[30] A. Markman Knowledge Representation, 2nd Edition. Mahwah: Lawrence Erlbaum Associates, 2000.

[31] P. Saariluoma. "Apperception and restructuring in chess players' problem solving," In Gilhooly, K., Keane, M., Logie, R., & Erdos, G. (Eds.), Lines of thought: reflections on the psychology of thinking. London: Wiley, 1990, pp. 41-57.

[32] P. Saariluoma. Chess and content oriented psychology of thinking. Psihologica, vol. 22, 2001, pp. 143-164.

[33] P. Saariluoma. Thinking in work life: from errors to opportunities (in Finnish). Porvoo: WSOY, 2003.

[34] P. Saariluoma, K. Nevala, and M. Karvinen. Content-based analysis of modes in design engineering. Berlin: Springer, 2006.

[35] P, Saariluoma, and I. Maarttola. Stumbling blocks in novice building design, 2003, <https://www.mit.jyu.fi/agora-center/inbct/InBCT23/Stumbling_blocks.pdf> 13.12.2010

[36] P. Saariluoma, and K. Nevala. The focus of content-based approach to design engineering a reply to Eder. Pilsen: University of Western Bohemia, 2006.

[37] University of Jyväskylä. Theseus Project. 2008-2010, <https://www.jyu.fi/erillis/agoracenter/tutkimus/acprojektit/katsy/engtheseus> 30.12.2010.

[38] University of Jyväskylä. ITEA 2 Easy Interactions Project. 2007-2010, <https://www.jyu.fi/erillis/agoracenter/en/research/projects/katsy/itea2> 30.12.2010.

[39] A. Collins, and M. Quillian. "Retrieval Time from Semantic Memory," Journal of Verbal Learning and Verbal Behavior, vol. 8, 1969, pp. 240-247.

[40] A. Cooper, R. Reimann, and D. Cronin. About Face 3. The essentials of interaction design. Indianapolis: Wiley, 2007.

[41] J. Fodor. Concepts: A Potboiler. Cognition, vol. 50, 1995, pp. 133-151.

[42] J. R. Anderson, M. Matessa, and C. Lebiere. "ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention." Human-Computer Interaction, vol. 12, 1997, pp. 439-462.

[43] S. Card, T. Moran, and A. Newell. The Psychology of Human-Computer Interaction. Hillsdale: Lawrence Erlbaum Associates, 1983.

[44] P. Saariluoma. "Error in chess: Apperception restructuring view," Psychological Research, vol. 54, 1992, pp. 17-26.

[45] G. A. Kelly. The psychology of personal constructs. New York: W.W. Norton, 1955.

[46] G. Rugg, and P. McGeorge. "The sorting techniques: A tutorial paper on card sorts, picture sorts and item sorts," Expert Systems, vol. 14, issue 2, 1997, pp. 80-93.

[47] G. Rugg, and P. McGeorge. "Concept sorting: The sorting techniques," In A. Kent & C. Hall (Eds.) Encyclopedia of Library and Information Science. New York: Marcel Dekker, 1999, pp. 43-70.

[48] M-D. Shieh, W. Yan, and C-H. Chen. "Soliciting customer requirements for product redesign based on picture sorts and ART2 neural network," Expert Systems with Applications, vol. 34, 2008, pp. 194-204.

[49] G. Rugg, and P. McGeorge, P. "Concept sorting: The sorting techniques," In A. Kent & C. Hall (Eds.) Encyclopedia of Library and Information Science. New York: Marcel Dekker, 1999, pp. 43-70.

[50] W. Yan, C-H. Chen, and L.P. Khoo. "Artificial Intelligence for Engineering Design, Analysis and Manufacturing. An integrated approach to the elicitation of customer requirements for engineering design using picture sorts and fuzzy evaluation," Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **vol.** 16* 2, 2002 April, pp. 59-71.

[51] F.D. Davis. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," MIS Quarterly, vol. 13, 1989, pp. 319-340.

[52] F.D. Davis, R. Bagozzi, and P. Warshaw. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, vol. 35, 1989, pp. 982-1003.

[53] P. Dourish. Seeking a Foundation for Context-Aware Computing, 2001, <http://www.dourish.com/embodied/essay.pdf> 13.12.2010

[54] L. Wittgenstein. Philosophical Investigations. Oxford: Blackwell Publishers Ltd, 1958.

[55] P. Saariluoma. *Foundational Analysis*. London: Routledge, 1997.

Query Cluster: A Method for Web Search Behavior

Jinyoung Kim
LG Electronics
Seoul, Korea
e-mail: amjinyoung.kim@lge.com

Moonsung Kim
Graduate School of Convergence Science and Technology
Seoul National University
Suwon, Korea
e-mail: amenra@snu.ac.kr

Joongseek Lee
Graduate School of Convergence Science and Technology
Seoul National University
Suwon, Korea
e-mail: joonlee8@snu.ac.kr

***Abstract*— There have been intensive research on user web search behavior since the late 1990s. Previous researchers collected data from search engines and analyzed explicit data (queries) to understand the characteristics of the user's search process, while other researchers analyzed the data of recruited subjects under experimental settings to understand the behavioral patterns in web usage. Although these researches provided an understanding of what users are searching for and how they are searching, both approaches did not provide rich user contexts that capture the reason why users are motivated to search, how long users' tasks (session) last, and other factors affecting user's search behavior. In this paper, we propose 'Clustered Query' as the unit of analysis in web search behavior studies. We found that users make their own Clustered-queries that yield better overview on their web search pattern, yet detailed individual web traces intact. The methodology consists of three phases and *Log Catcher*, *Query Cluster*, *Monitoring tool*, and Retrospective Interview technique are used in each phase. At the end of this paper, we also illustrate the process of the pilot and main study where the methodology is modified and validated.**

Keywords-Web search behavior; Methodology; User Intent; User Context

I. INTRODUCTION

As stated in the 2005 Pew Internet Report, 'Web has become the new normal' in the way of modern life [1]. Web and information retrieval has become the dominant issue in the field of information studies ever since. With the development of web and mobile, there also have been changes of human information behavior. The strategic redesigning of web search services such as Google [2] and NAVER [3] has brought major changes from the way we recognize the needs of information to the way we engage information seeking behavior. Search assistance features such as 'real-time issues' and 'related keywords recommendation' have opened up new ways of searching by generating user's needs or by providing shortcuts to reach the information a user wants. The Web serves users' daily information behavior, and the mobile platform is accelerating this phenomenon. Users no longer seek for information just for their jobs, tasks, or expertise but also for everyday curiosity and fun. Even more, they do not need to seek for information as the information comes to the users.

However, previous researchers have focused on the framework that illustrates user's information behavior and the task-related information needs and process. Although these studies have contributed to the information behavior studies, several constraints are also perceived such as a lack of empirical studies supporting the framework and a lack of user data in natural settings. The studies mostly relied on the qualitative research methodology such as in-depth-interviews to acquire user data. Other researches concentrated on the analysis of users' daily web usage and search patterns using quantitative data collected through search engine logs or customized tools. These researches are restricted to understand user's context as they collected and analyzed a 'series of queries' that random users typed.

In this paper, we describe a methodology to capture the user's usual web search behavior and the context of the web search behavior. The methodology allows researchers to collect data from users' web activity logs in natural settings and accumulations of context to affect the web user's search behavior. The rest of the paper is organized as follows: In Section 2, we reviewed the previous studies of web search log. The method, Query Cluster, and the refinement of method are introduced in Section 3 and Section 4, respectively. We concluded Section 5 with the discussions and future steps of the study.

II. RELATED WORKS

Web log data allowed researchers to track back user's information behavior rather than to assume with user's recollection or diary data. According to Jansen and Spink [4], web-searching studies can be categorized into three methodologies: (1) transaction-log analysis, (2) experimental setting analysis, and (3) issues related to web searching. In this paper, we focus studies on the transaction-log analysis and on the experimental setting analysis.

Transaction-log analysis web-searching studies are one of the major streams that analyze data acquired from search engines to understand the characteristics of web searching behavior. These researches are meaningful as most web users gather to search the engine/portal looking for new information. Researchers extracted the characteristics of web searching by investigating the frequency of query occurrence, the average length of query, the typical query session, or the relevance among queries to improve current web search engine [4][5].

 ISBN: 978-1-61208-117-5

One approach of transaction-log studies is to investigate tactics or strategies in user's web searching. Silverstein et al. and Jansen examined query characteristics and correlation of query logs from the Alta-Vista search transaction- logs [5][6]. Different approaches have been applied to cluster and classify search queries. Ross and Wolfram analyzed the co-occurrence of query terms among the Excite search engine queries [7]. They presented a hierarchical cluster analysis comprising the topics. Shi and Yang also developed a method to identify related queries by extracting and segmenting query sessions and mining association rules from a Taiwanese search engine [8]. While these studies mostly focused on extracting topics of query terms, Rose and Levinson were concerned with understanding the users' intrinsic goals of user searches [9]. They characterized the user search goals – Navigational, Informational, and Resource – that are derived from Broder's 'Taxonomy of Web Search', and manually classified the searching queries of the three goals [10].

Transaction-log studies have strength as they deal with a large number of data of random users, and less likely to be affected by trends [5]. However, it is hard to observe the behavioral pattern of a user and to understand the user context with anonymously collected data. Researchers have to rely on the log data that shows when users search and what they search for, and cannot report in a user-centered manner because of the lack of contextual information [11]. It also has limitation that the analysis may reflect the characteristics of the search engine.

Experimental setting studies, on the other hand, are to analyze data acquired from the customized tools installed on the participant's computer or using the web browser. Participants are recruited for the experiment and their web search pattern is analyzed while the transaction-log studies mainly focus on the analysis of obtained data. These researches cover topics from the characteristics of interaction during information seeking to the context of information seeking.

Choo et al. observed the web seeking behavior of 34 knowledge workers to find out their information needs and information seeking preferences [12]. They extracted the significant episodes during web usage through in-depth-interviews. A customized tool, WebTracker, collected participants' web log data of URL calls/requests, browser menu selections (i.e., reload, back, and forward) and the collected data was used as the background information for the interview. They identified 61 significant episodes of information seeking and categorized them into 4 complementary modes of information seeking.

Sellen et al. studied how and why knowledge workers use the web with a methodology that combined diaries and interviews. They interviewed 24 workers about their web search history with web history references written on the worker's personal computer. Participants were asked to tell a story of their searching activities and to rate their web activities with respect to the success/failure, significance of the activity, and time spent on the activity [13].

Kelly proposed a method for collecting the user data about information seeking contexts and behaviors in natural environments [14]. Seven PhD students used laptops equipped with a client-side logger. The students reported 5 variables – endurance, frequency, stage, persistence, and familiarity- related to the tasks and topics of their web seeking behavior, and usefulness ratings and confidence of the document. More details were obtained through the exit interview at the end of the research.

Kellar et al. also examined how users interact with their web browsers during information-seeking tasks [15]. 21 students installed a custom-built web browser that collected visited websites and browser menu logs. Students reported their own browsing histories in task types -fact finding, information gathering, just browsing, transactions, and others- and task descriptions through electronic diaries or real-time reports. Experimental setting analysis usually uses a combined methodology to obtain qualitative data and quantitative log. A small group of participants is recruited for the research and the customized tool collects user's web log and interview follows. Although the experimental setting analysis provides qualitative data of user context, it still has restrictions of small data sets. Also participants sometimes forgot about the past research behavior as interviews or clustering assignment are delayed [14] and researchers missed the details of user context as they focused on the browser controlling behavior [15].

The purpose of this study is to develop a method to overcome limitations of prior studies. We used 'Clustered Query' as the unit of analysis that is grouped by users, instead of 'session' that are mainly used in the transaction-log studies. Clustered Query is a meaningful unit that shows the duration of attention toward a topic and the steps of search. Log Catcher installed in the participants' personal computers and collect logs of the natural web searching behavior. The log data provides quantitative information to the researchers such as duration, a number of 'Clustered Query' and a number of queries in each 'Clustered Query' on a day. It also helps participants answer the questionnaire and helps researchers obtain qualitative information.

III. METHODOLOGY: QUERY CLUSTER

In this section, we describe the methodology, query cluster, to collect user contexts in order to understand the user's intentions in web searching. The first part of this section presents the three phases of the research model and the terms frequently used in this paper. Each phase contains a description of tools that we have developed for the research. For the next part, the pilot and main study that we carried out to validate the methodology is introduced.

Our research model consists of three phases: (1) the Set-up Phase, (2) Experiment Phase, and (3) Revision Phase. For each phase, we developed tools to acquire user contexts and intents of web searching behavior (see Figure 1). We borrowed the theories of Marchionini and Jones and Brown

 ISBN: 978-1-61208-117-5

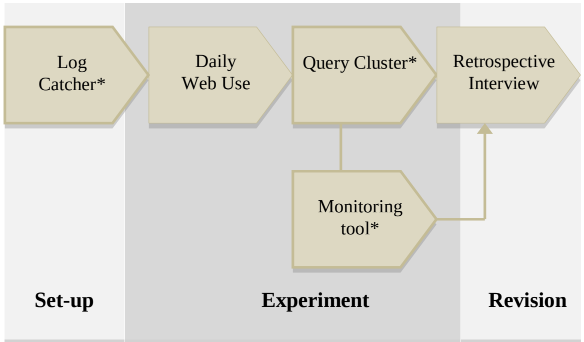


Figure 1. Research Model: 3 Phases

and categorized user contexts into 6 factors, which are information seeker, setting/physical context, task, search system, domain knowledge, and outcomes [16][17]. User data of 6 factors are mainly collected in the experiment phase, and other phases support to fill in the missing information.

For the Set-up phase, participants are given a package consisting of a questionnaire, a log catcher software, and research guidance video. After 1-2 days of stabilizing and learning period, the experiment phase starts. Participants are expected to use their computer daily as usual for two weeks and to access the research website for the assignments, which are clustering the queries of the previous day and answering questions. Researchers monitor participants' daily assignments and note the clustered queries to ask the context. In the revision phase, researchers carry out in-depth-interviews of participants about the clustered-queries and obtain rich user context.

Frequently used terms are defined as follows:

Queries are the keywords that were typed by the user during the process of web portal search activities. In this paper, we collected queries from 7 popular search engines (Google, Yahoo, Naver, Daum, Nate, Paran, and YouTube) in Korea. We focused on the queries rather than the general web activity logs because a series of queries represent user's information problems and reflect user's wide-ranging information needs [18].

General web activity logs are the website logs that users visited except for the search engine queries.

Cluster is the action of grouping a number of queries that are made under the same objective or intent.

Clustered-queries are the cluster that participants made with their queries collected on the previous day. Each cluster has one or more queries and is titled with subject-representative words. Kelly and Kellar et al. applied similar methods to make participants categorize their own web logs one by one [14][15].

Assignment consists of *Cluster*-ing and answering to the questionnaire about the *Clustered-queries*. A participant is supposed to complete a daily assignment about previous day's web activities log.

TABLE I. FACTORS AFFECTING INFORMATION SEEKING

Marchionini (1997) /Jones & Brown (2004)	Research Model		
	Set-up	*Experiment*	*Revision*
Information Seeker	-	Motivation	
Setting/Physical Context	Search assistance	Trigger (Physical-context)	Physical-context
Task	-	Clustered-queries	
Search System	Information Source	Search engine Search assistance	
Domain knowledge	-	-	Interview
Outcomes	-	Satisfaction	

A. Set-up Phase

During the set-up phase, researchers understand the basic information about users' web behavior, and participants install the provided tool to their computing environments and learn the tasks for the experiment. An experiment package is provided to the participants, which consists of a guidance video on the tasks and the flows of the experiment, the log catcher tool, and the entry questionnaire. Researchers communicate with participants to inform whether the log data are collected well and to train them how to cluster.

1) Entry Questionnaire

Participants are expected to fill out entry questionnaires at the start of the research. The questionnaire contains 10 items relating to demographic, Internet usage, information ground on the web, and web searching behavior. The questionnaire result provides an understanding of the web search behavior of participants, and some items such as computer type, the period of web usage, the objectives of web search can be used to screen participants.

2) Log Catcher Installation

A client-side log-collecting tool, *Log Catcher* (Log Catcher was written in C# and used windows process hooking mechanism), is published via the research website in a packaged wizard format. As the objective of this research is to understand users' web activities in natural settings, participants are encouraged to install Log Catcher on their personal computer. Log Catcher is designed to collect web activity logs and the condition of collecting web activity logs is informed during the installation process. The following information is collected through Log Catcher:

- Access Time: the time that the participant visited the web site
- Page Title: head title of web page

 ISBN: 978-1-61208-117-5

TABLE II. INFORMATION COLLECTED BY LOG CATCHER

Access Time	Page Title	URL	Query	Search Engine	Search service	Browser ID	Web Browser
2010-07-22 12:03:27	Sports Today	http://stoo.asiae.co.kr				1114682	Internet Explorer
2010-07-22 12:04:55	Web Hard™	http://www.webhard.co.kr				66218	Internet Explorer
2010-07-22 12:06:09	Naver Search	http://search.naver.com	Sports News	Naver	Web Search	131428	Internet Explorer
2010-07-22 12:06:44	Naver	http://www.naver.com				131428	Internet Explorer
2010-07-22 12:06:45	Naver Search	http://search.naver.com	Free Hi-Pass	Naver	Web Search	131428	Internet Explorer

- URL: URL of visited web page
- Query: typed keywords on the search engine
- Search Engine: search engine that the participants accessed to query
- Search Service: specific search service provided by the search engine (e.g., web search, image search, news search)
- IP address: participant's IP address
- Web Browser: Web browser's process ID to distinguish the different web browser windows and tabs
- Web Browser Name: Web browser name that the participant is using (e.g., Microsoft Internet Explorer, Mozilla Firefox, Google Chrome)

Participants should input their activation code provided with the installation package to finish the installation process. The activation code enables researchers to track the participant's status such as whether the participant was successful in installing the Log Catcher and whether the log data was sent to the server. After installing, Log Catcher is launched automatically when the participant turns on his/her computer and the data collected are sent to the research database every 5 minutes.

Log Catcher stores the participant's log data until the data is sent to the database to prevent data loss in cases of network failures or unexpected system shutdowns.

B. Experiment Phase

When Log Catcher is stabilized on the participants' computer, the Experiment Phase follows. Participants are guided to explore the web daily in the same way as they did previously. The daily web activities are collected in the database, and the participant visits the research website to do their assignment. The assignment is consisted of two parts: Cluster and Questionnaires for the Clustered-queries. Participants view their own web activity logs on the research website and cluster the queries in groups according to the rules that they learned. After the clustering finishes, the participant answers to the questionnaire corresponded to the clustered-queries.

1) Definition of 'day'

We defined that a 'day in web usage' is between 5:00 a.m. and 5:00 a.m. the next day. Due to the number of participants surfing the web until dawn just before they go to work or school, 12:00 am to 12:00 am collection may bring problems for participants in clustering their queries. r their own queries.

2) Query Cluster

Participants are instructed to access the research website and to cluster their own queries that were collected on the previous day. For this, we built a specialized web-log clustering and reporting tool, *Query Cluster*, to help participants to cluster and answer the questionnaire easily. As the design of the Query Cluster is similar to the web card-sorting tool, participants reported that they have no difficulties in using the tool and the tool helped them understand how to cluster their own queries.

When logging in to the Query Cluster with his/her activation code, a participant can view the experiment date page and click the previous date for the assignment. The date is activated if a participant missed the assignment and deactivated if he/she completed the assignment or no logs have yet been collected.

(a) My web activity logs: The web activity logs of the selected date are shown on the left column. It contains the logs of all websites visited and shows the page title, query texts that user typed, search service, search engines and domains. The web activity logs are listed chronically, and we intentionally left time blanks for any web activity logs that are not collected for an hour. Time blanks and general web activity logs other than the search engine queries are provided for the participants to help them to recollect the reasons why they typed the queries and to cluster the queries. Query logs are lightly shadowed to distinguish from general web activity logs, and the search engine logo is displayed with query logs.

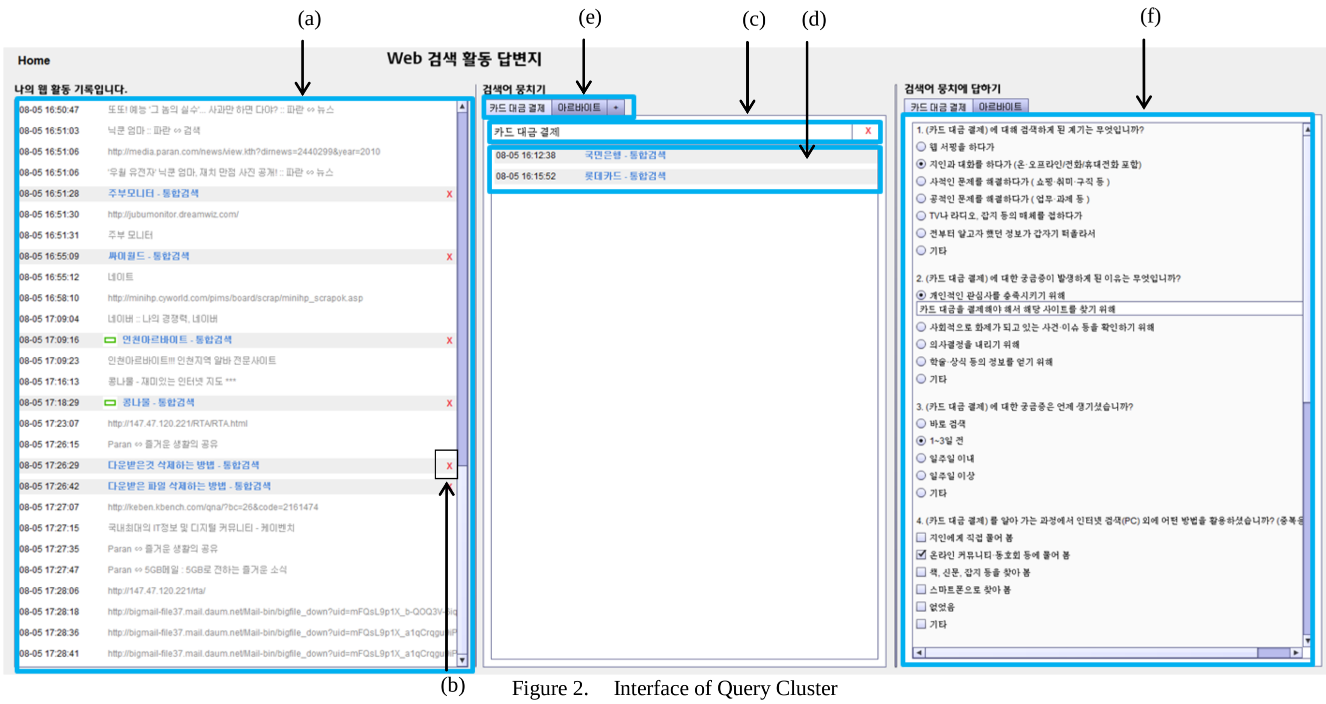


Figure 2. Interface of Query Cluster

(b) Query Delete button: Participants can remove their own queries from the list in case they do not want to cluster the queries. We added this function after the pilot study as it was observed that several queries were too private to share with researchers or sometimes misspelled. For every deleted query, the reason why it needs to be deleted was specified in order to prevent abusive usage.

(c) Title of Clustered-queries: The Clustered-queries is titled with subject-representative words that participants typed. (e) represents the titles of Clustered-queries that a participant made.

(d) Clustered-queries: Participants drag and drop the queries from the left column to the right column to make a cluster. Queries that are moved to the right column are disabled on the left column.

(e) Titles of Clustered-queries: Participants can make a new cluster by clicking '+' button. A blank title bar and box for clustering appears.

(f) Questionnaire: Each cluster has a questionnaire and participants should fill out the questionnaires to finish the daily assignment. The details are explained in the 'contextual questionnaire' section.

3) *Contextual Questionnaire*

When participants click the Next button after finishing the *Cluster*-ing, the contextual questionnaire is displayed. We used five variables to obtain the user data of factors affecting the information seeking:

- **Motivation** is the goal that participants want to reach by resolving the recognized information gap. The multiple choices consists of 5 items that are derived from the past information behavior studies, which are sense making/reducing uncertainty, decision making, problem solving, fact knowing for personal reason and others. A text area appears when a participant selects an answer for the details of motivation.
- **Trigger** is the internal or external cue for a participant to conduct the type of query. 6 multiple choices are provided for selection, which are during web surfing, communicating with others, working, media consuming, personal affairs, and others. We also asked participants when their motivation arised for the first time.
- **Physical Context** is the user's environmental context which includes where, with whom, doing what. Although we asked participants to answer the question about physical contexts, we dropped the question as the other questions or interviews could provide the physical contexts.
- **Search assistance features** are the additional functions that a participant uses during the process of search. Choices include auto-completion, related keywords, real-time issues, spell-check, and nothing.
- **Satisfaction** consists of three components: how much the participant gets satisfied, whether a participant reaches the goal through search behavior, and whether he/she would try additional searches later.

C. *Revision Phase*

In the revision phase, researchers review the collected data and conduct in-depth-interviews to fill the missing clues for the user context.

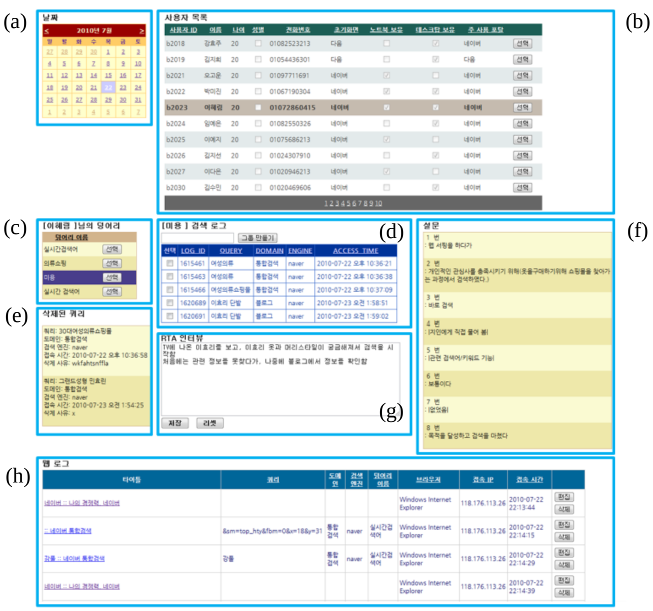


Figure 3. Interface of Monitoring Tool

1) Monitoring Tool

We developed a monitoring tool for researchers. The monitoring tool enables researchers to check the status of research, to extract unusual queries/answers at a glance and to interview with well-formatted data. The interface of this tool is shown in Figure 3.

(a) Calendar: Researchers can select the date to check.

(b) List of participants: Participants whose data were collected on the selected date are listed. The list contains id, demographic information, and background information of participants. (e.g., name, age, gender, phone number, search engine preference).

(c) Titles of Clustered-queries

(d) Clustered-queries: Clustered-queries are listed when a researcher selects a title of clustered-queries. The information contains query text, search engine, searched domain, and access date and time. Researchers can create or modify clusters if the participant asks to change.

(e) Deleted queries: The deleted queries and reasons appear if the participant removes the query from the logs.

(f) Questionnaire answers about the cluster

(g) Memo box: Researchers can leave notes about the findings during the interview.

(h) Web activities log: Contains both queries and general web activity logs.

2) Exit Interview

A retrospective interview technique is applied for the exit interview. Although the technique is not considered to be appropriate to collect accurate and objective data, it is useful to build a history of event or exploratory experience by making participants recall their aspects of past experiences [19].

Researchers carry out the exit interview via telephone once a week and twice during the 2-week research period. Before the interview, researchers reviewed the Clustered-queries and Contextual Questionnaire of users. Participants are asked to tell a story about the situation and motivation of Clustered-queries and the relations among the queries in a cluster.

Participants' domain knowledge about the clustered-queries also can be asked through the exit interview: whether a participant is accustomed to the topic, which information grounds he/she relies on for seeking information about the topic, or how frequently a participant searches for the related topic.

IV. Research Process

The entire research process is designed to modify problems and to prove the validity of the methodology. Several changes have been made through the pilot study, and tools and survey questions were reviewed. After refining the methodology, we carried out the main study with a large number of participants and found concerning points when applying the methodology.

A. *Pilot Study*

Among 8 participants recruited for the pilot study, 4 were females and 4 were males and all participants were between the ages of 20-40. We considered the participants' job, where 4 were undergraduates and 4 were paid workers.

During the 6-day pilot study, most participants set up the log catcher tool in their personal computer because of the security issues in collecting logs at their workplace computers. The participants had selectively sent their web search logs by turning on and off the log catcher tool.

Participants clustered their log histories on the web, and answered survey questions about each log clusters on a daily basis. At the end of the pilot study, the researcher interviewed the participants via mobile phone for 30 minutes or more to acquire user contexts in web searching activities.

B. *Problems*

We found several problems through the pilot study and modified the research process and details to enhance the participants' engagement and to acquire valid user data from the study.

1) Selective Report of Log Histories

In the pilot study, participants selectively reported their log histories by clicking the on-off button provided. We asked participants to turn on the log catcher tool at least two hours a day to acquire equivalent amount of data from all subjects and to respect their privacy. However, some participants got confused when turning the log catcher on and failed to report their search histories. Other participants intentionally hid their search histories and invented data for the report. The log catcher tool was modified to collect the whole log histories during web use. In the main study, we recruited participants who accepted this condition and added a function to delete queries on the tool.

2) Erros in Clustering

Throughout the pilot study, we noticed that participants made mistakes when clustering their own queries at times. Some participants, for example, classified queries into different clusters although the queries were made with the same motivation. Other examples are inappropriate titles, for example, a participant titled clustered-queries as ‘day 1’. We trained those participants to cluster queries based on a motivation, and to title clustered-queries to represent the subject of the cluster.

Most errors were found in the early stages of the experiment, as participants are not accustomed to the experiment and the clustering rule at first. Therefore, the data of first 1~2 days should be reviewed carefully and researchers should communicate with participants to understand how to cluster and title. However several cases were reported during the interviews and it was also required to provide a cluster-modification function on the administration tool. In the main study, researchers combine or separate queries during the interview.

3) Surveys about Clustered-queries

The initial version of the contextual questionnaire contains several open-ended questions to ask participants to answer the physical contexts or search reasons. However, we found that the participants were not willingly answering the open-ended question of each clustered queries. We decided to drop the burdensome questions as the exit interview study and other questions can cover them. We also found that our participants usually search at home, the physical contexts are not considered as crucial feature in this study.

C. Main Study

After revising the methodology, we carried out the main study with a large number of participants. 100 participants were recruited for the main study and the demographic ratio of participants were similar to the Korean demographic data except for the geographic; 25 male undergraduates and 25 female undergraduates in their 20’s, and 25 male paid workers and 15 female paid workers-10 housewives in their 30’s.

For 14 days of the main study, most participants installed the log catcher tool on their personal computer. Entire log histories of the participants were collected on the server and participants were asked to cluster queries and to answer the surveys to each clustered queries of the previous day. During the main study, participants were interviewed twice about the context of the clustered queries or correlation between the clustered queries found on another day.

V. CONCLUSION

This paper presents a methodology to understand the user context in web search behavior with Clustered Query as a research unit. The methodology consists of three phases - set-up, experiment, and revision phase-, and the methodology is refined and validated through the pilot and main study. User context is defined as 6 factors and tools are introduced that are developed to obtain user data in each phase.

The contributions of the proposed methodology are usability and user-oriented approach. Participants highly engage in the experiment phase by clustering their own queries, and provide meaningful clusters that cannot be captured through previous log analysis studies. Our methodology improved the previous quantitative and qualitative approaches by collecting quantitative data of users’ web activities logs and qualitative data of questionnaires and interviews. The self clustered-queries deliver valuable data to understand the user intents and the task session.

For the next step, we will analyze the data obtained through the pilot and main study focusing on the categorization of user intent and its effect on search behavior.

REFERENCES

[1] The Mainstreaming of Online Life. http://www.pewinternet.org/~/media//Files/Reports/2005/Internet_Status_2005.pdf.pdf, 01.07.2011

[2] http://www.google.com, 01.07.2011

[3] http://www.naver.com, 01.07.2011

[4] B. Jansen and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," Information Processing & Management, vol. 42, Jan. 2006, pp. 248-263, doi:10.1016/j.ipm.2004.10.007.

[5] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, “Analysis of a very large web search engine query log," ACM SIGIR Forum, vol. 33, Fall 1999, pp. 6-12, doi:10.1145/331403.331405.

[6] B. Jansen, A. Spink, and J. Pedersen, “A temporal comparison of AltaVista Web searching,” Journal of the American Society for Information Science and Technology, vol. 56, Feb. 2005, pp. 559–570, doi:10.1002/asi.20145.

[7] N. Ross and D. Wolfram, “End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine,” Journal of the American Society for Information Science, vol. 51, Jun. 2000, pp. 949–958, doi: 10.1002/1097-4571(2000)51:10<949::AID-ASI70>3.0.CO;2-5.

[8] X. Shi and C. Yang, “Mining related queries from Web search engine query logs using an improved association rule mining model,” Journal of the American Society for Information Science and Technology, vol. 58, Aug. 2007, pp. 1871–1883, doi: 10.1002/asi.20632.

[9] D. Rose and D. Levinson, “Understanding user goals in web search,” Proceedings of the 13th international conference on World Wide Web (WWW 04), ACM, 2004, pp. 13-19, doi:10.1145/988672.988675.

[10] A. Broder, “A taxonomy of web search,” ACM SIGIR Forum, vol. 36, Fall 2002, pp. 3-10, doi:10.1145/792550.792552.

[11] B. Jansen, A. Spink, and T. Saracevic, “Real life, real users, and real needs: a study and analysis of user queries on the web,” Information Processing & Management, vol. 36, Mar. 2000, pp. 207–227, doi:10.1016/S0306-4573(99)00056-4.

[12] C. Choo, B. Detlor, and D. Turnbull, “Information Seeking on the Web: An Integrated Model of Browsing and Searching,” Proceedings of the ASIS Annual Meeting, Information Today, vol. 36, Oct. 1999, pp. 3–16.

[13] A. Sellen, R. Murphy, and K. Shaw, “How knowledge workers use the web,” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(CHI 02), ACM, 2002, pp. 227-234, doi:10.1145/503376.503418.

[14] D. Kelly, “Measuring online information seeking context, Part 1: Background and method,” Journal of the American Society for Information Science and Technology, vol. 57, Sep. 2006, pp. 1729–1739, doi:10.1002/asi.20483.

[15] M. Kellar, C. Watters, and M. Shepherd, “A field study characterizing Web-based information-seeking tasks,” Journal of the American Society for Information Science and Technology, vol. 58, Mar. 2007, pp. 999–1018, doi:10.1002/asi.20590.

[16] G. Marchionini, “Information Seeking in Electronic Environments,” Cambridge University Press, 1995, pp. 33-38.

[17] G. Jones and P. Brown, “The role of context in information retrieval,” Proceedings of the ACM SIGIR 2004 Workshop on Information Retrieval in Context, ACM, 2004, pp. 20-22.

[18] S. Rieh and H. Xie, “Analysis of multiple query reformulations on the web: The interactive information retrieval context,” Information Processing & Management, vol. 42, May 2006, pp. 751–768, doi:10.1016/j.ipm.2005.05.005.

[19] B. Montgomery and S. Duck, eds. “Studying Interpersonal Interaction. Guilford Publishing”, The Guilford Press, 1991, pp. 162-178.

Temporal Aspects of Human-machine Interaction in the Perception of Visual Information

Igor Petukhov
Dept. of Radio Engineering
Mari State Technical University, MarSTU
Yoshkar-Ola, Russia
e-mail: laboratory502@rambler.ru

***Abstract*—The vision system is one of the major sensory systems in the process of human-machine interaction. To improve the quality of interaction it is necessary to evaluate optimal parameters for the speed of perception and volume of visual information. New methods of evaluating the time of visual perception as well as the time of recovery and lability of the vision system are elaborated determining the inertia of the vision system and its ability to perceive visual information. New data on the temporal parameters of visual information processing are obtained through experimental research. The inertia of visual perception is ascertained to depend on the operator's fatigue in the working process.**

Keywords-human-machine interaction; visual perception; inertia

I. INTRODUCTION

The problem of providing human-machine interaction is considered to be relevant and important. The human factor is responsible for more than 70% of all industrial and traffic accidents [1]. The reason for this is the fact that a human's professional activity is becoming more and more complex. As a result, the load on the perceiving, identifying and decision making systems of a person is becoming increasingly heavy.

A mismatch between the information load and the capacities of human perception results in early fatigue of a human operator and errors in his/her work.

Human errors are caused by many factors, poor human-computer interface (HCI), amount of operator's workload, his/her experience, shift-fatigue, etc. being among them [2, 3].

The most badly influenced is vision system providing perception of 80% of all the information actively involved in professional activity.

Temporal parameters of the vision system stipulate the requirements for HCI organization: the speed of information presentation, color, shape and size of objects and their complexity [4, 5].

The fatigue of an operator, the number of errors and labor productivity depend on the parameters of an HCI [6].

The development of HCIs adaptively configured to user, depending on the degree of his/her fatigue and difficulty of the tasks performed is of current interest.

All this requires knowledge on the mechanisms of visual information processing, the temporal dynamics of visual perception indicators.

Thus, the study of information interaction between a human operator and a technical system, aiming at determining the optimal modes of their operation, is of paramount importance.

The objective of the current work is to obtain new data on the processes of perceiving and processing visual information.

In this paper, a new method of recovery time determining, method of visual perception time evaluation and method of the vision system lability evaluation are developed. The results of modeling and new data on the temporal parameters of visual information processing are obtained.

II. APPROACHES TO RELIABILITY CONTROL

According to the analysis carried out, enhancing reliability of man-machine systems requires taking two basic approaches:

- Adapting the structure of information flow and interface parameters to the needs of an individual human operator [7].
- Adapting the user to a computerized system [8].

There is a progress in the achievement of certain quality indicators in both approaches.

Adaptive interfaces are aimed at changing and optimizing themselves automatically, depending on the workload of the operator and dynamically changing characteristics of the medium [9].

It is known that using intellectual adaptive interfaces allows considerably reducing the time expenses of the operator and the volume of his/her work, helping to make correct decisions [10].

Adaptation of the interface parameters is carried out through the automatic correction of the user errors [11], changes in the level of complexity of the interface [12], adaptation to the intensity of information exchange between the user and the system, customization of the technical system to the aims and intentions of the user [13-15], etc.

Simplification of information blocks is one of perspective methods of the organization of modern user interfaces [16].

The methods and means of adapting a user to the system are mainly based on the user study and training. To organize

 ISBN: 978-1-61208-117-5

adaptation of the user to the system it is necessary to take into account his/her physiological characteristics, behavior and physical state, for which testing (diagnosis) of the user to form his/her psycho-dynamic portrait is implemented [17].

It should be noted that both methods are based on the data about the processes of information perception and processing by a human.

There is a large number of methods for studying the processes of information perception such as the critical flicker frequency (CFF) [18], the time of visual identification [19, 20], the visual sensation [21] and etc. obtained by numerous experimental data on the temporal aspects of processing.

At the same time, the data obtained by different researchers are not systematic by nature and sometimes contradictory.

This brings about the development of new highly accurate and reliable methods and systematization of methodological base.

III. RESEARCH METHODS

The information used by the operator strongly differs by complexity, the form of representation and information content.

To determine potential ability of perception it is necessary to use the information being the simplest for perception. This provides the maximum speed of information processing.

Such condition is satisfied by light pulses of rectangular shape of the "yes-no" type.

Besides, this method of evaluation of the critical flicker frequency is well known in the studying of the processes of visual perception [18].

Critical flicker frequency is the frequency of light flashings per second at which the subjective flicker fusion takes place.

The disadvantage of this method is its low accuracy, low reliability of measurement results.

The research results of the CFF are achieved at successive masking caused by a rhythmic sequence of light pulses, which hinders visual information processing.

A number of new research methods into the visual perception time aspects based on the CFF method has been worked out.

A. The method of recovery time determining

Recovery time (RT) of the visual system is the time during which the vision system is not able to accept a new flow of information. It is the time interval between the first and the second light pulses during which the impulses are perceived separately.

To determine the recovery time a human is supposed to be exposed to a sequence of pairs of light pulses of fixed duration τ_{imp} (Fig. 1).

The pause between light pulses t_{PI} is reduced to critical value t_{cr} at which the fusion of light pulses in a pair is observed (Fig. 2).

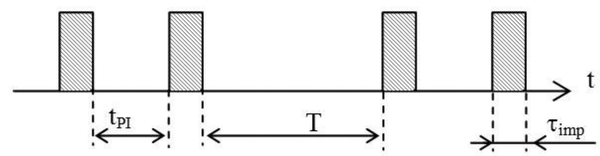


Figure 1. The time diagram of sequence of pair light impulses

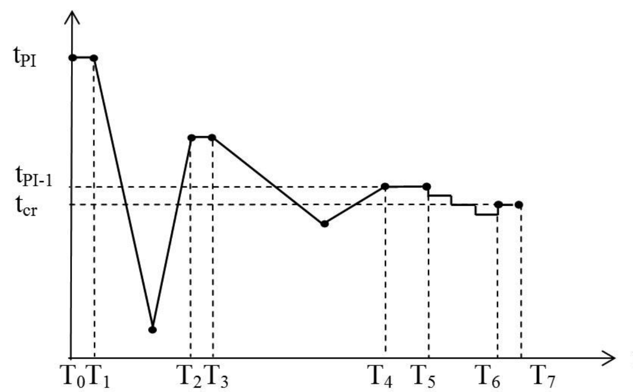


Figure 2. Changing the length of the pause diagram.

The pause at this point is taken equal to the recovery time of the vision system.

To shorten the measurement time at intervals $T_0 – T_4$ the change of the pause length occurs continuously at different speeds. To enhance the accuracy at the final stage (interval $T_5 – T_7$) the change of the pause length occurs discretely with a step of 0.1 ms.

It is known that at light pulse duration of 40-50 ms the processes of neurons receptive fields restructuring in the vision system terminate. Therefore, the duration of light pulses τ_{imp} is chosen to be 50 ms not to disrupt the process of visual perception.

It is also known that a consistent visual masking is observed at time intervals less than 800 ms. Therefore, the time interval between pairs of light pulses T is equal to 1 sec.

As the operator works with visual information of various colors, it is meaningful to carry out research of vision system restoration time on several colors.

The measurement technique will be modified as follows (Fig. 3).

After determining the recovery time on one of the colors $t_{cr\ 1}$ (T_7), the pause between the paired light pulses is increased by a random value. Then the color of the light pulses is changed, and the process of evaluating the recovery time is repeated. The duration of the pause varies discretely with a step of 0.1 ms.

 ISBN: 978-1-61208-117-5

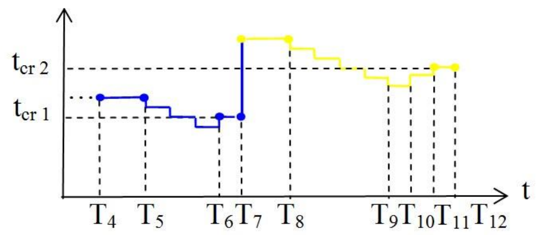


Figure 3. Changing the length of the pause with different colors diagram.

When using light pulses of different colors the process of measuring on the interval $T_7 - T_{12}$ is repeated as many times as necessary.

B. The method of visual perception time evaluation

Time of visual perception (TVP) is an integral parameter of visual perception inertia. TVP is the period from the start of a short exposure test stimulus to the inclusion of a masking stimulus, when the latter cannot interfere with the test stimulus comprehension.

The time for visual information processing is known to depend on the exposure time of test stimulus.

The advantage of the method for determining TVP consists in the fact that it takes into account the duration of the light stimulus exposure.

To define TVP it is offered to use the method of visual system restoration time evaluation. Similarly, a human is exposed to a sequence of paired light pulses, and the duration of a pause at which light pulses in a pair fuse is evaluated.

TVP is accepted equal to the sum of the pulse τ_{imp} and the pause t_{cr} duration at the time of light pulses fusion.

The method of determining TVP technique is similar to the recovery time evaluation method described above.

C. The method of the vision system lability evaluation

The lability of the vision system is an integral value of the lability of the central nervous system and changes in functional status of a human as a whole.

Lability characterizes the ability of the nervous system to react to excitation in exact conformity with the rhythm of stimulation.

The determination of lability is also suggested to be carried out, basing on the method of visual perception time evaluation.

To define the lability of a vision system the examinee is exposed to paired light pulses of fixed duration. In each series of experiments the duration of light pulses is different and equals 110, 90, 70, 50, 30, 10, 5 and 1 ms.

At each duration of light pulses a pause between light pulses in a pair at which impulses fuse is evaluated.

The method of a pause evaluation is shown in Fig. 4.

For each duration of light pulses TVP is calculated and a graph of TVP is built based on the duration of the light pulse, evaluating the minimum functions.

The lability of a human vision system is taken equal to the value of the repetition frequency of light pulses at the minimum point of Hz function:

$$F=1/(\tau_{imp}+t_{cr})$$

τ_{imp} – duration of the light pulse in seconds; t_{cr} – pause length between the light pulses during which the light pulses fuse.

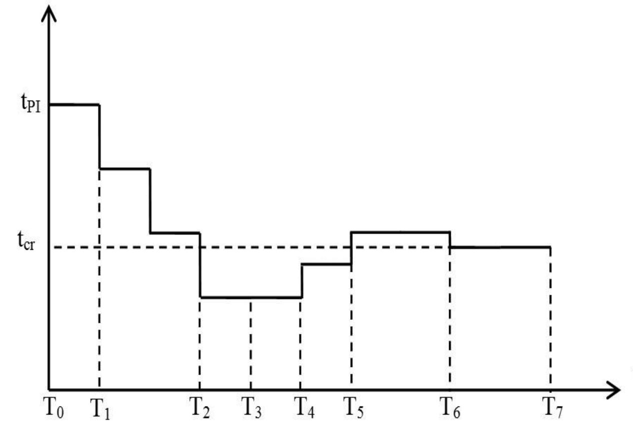


Figure 4. Changing the length of the pause diagram.

IV. MODEL OF THE VISION SYSTEM

As the analysis of literature shows, there is a great number of vision system and its subsystems models.

Our task was to develop a model that can be used for the methods elaborated and can explain the differences in temporal dynamics at the perception of light stimuli of different colors. The structural model based on the data about the vision system structure is presented in Fig. 5.

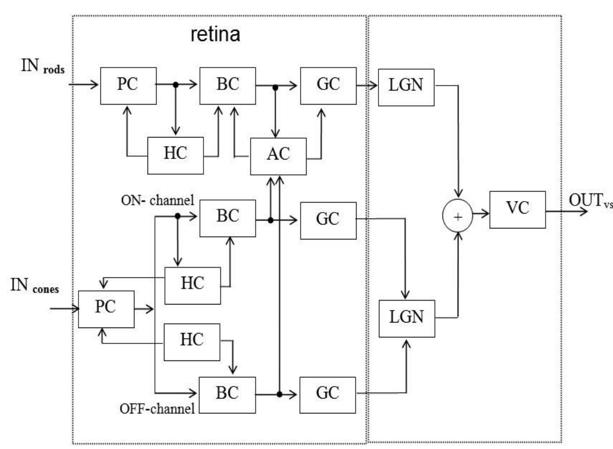


Figure 5. Scheme of the visual system, PC - photoreceptor cell, BC - bipolar cells, HC - horizontal cells, AC - amacrine cells, GC - ganglion cells, LGN - lateral geniculate nucleus, VC - visual cortex, VS – visual system.

 ISBN: 978-1-61208-117-5

This model considers division on rods and cones paths of perception and on-, off-channels.

This agrees with the data obtained by T. Gollisch & M. Meister, 2008 [22], that for adequate simulation of the vision system it is necessary to consider the parallel ON- OFF channels.

For mathematical modeling the well-known transfer functions of neurons and retinal neurons were used.

The results of modeling the perception of paired light pulses of 10 ms are shown in Fig. 6.

The dependence of visual perception on the pulse duration is shown in Fig. 7.

Here the point on the graph with duration of 10 ms corresponds to the lability of the visual system.

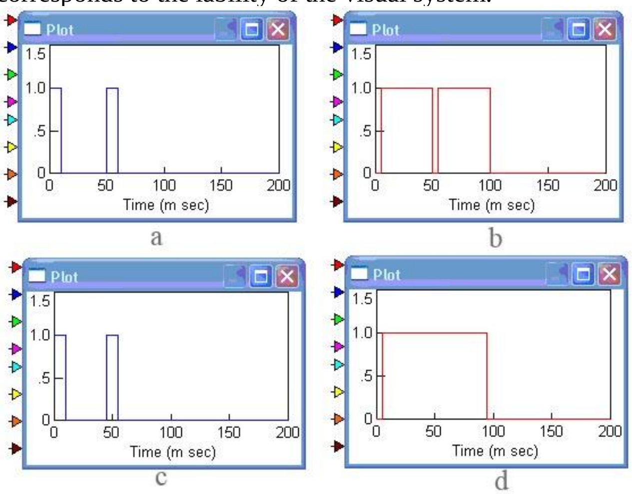


Figure 6. Results of modeling the perception of paired light pulses of 10 ms; a) timing diagram of the two input pulses of 10 ms, separated by a pause of 40 ms; b) timing diagram of the output pulses model presented in Fig. 6 a; c) timing diagram of the two input pulses of 10 ms, separated by a pause of 35 ms; d) timing diagram of the output model to pulses presented in Fig. 6 c.

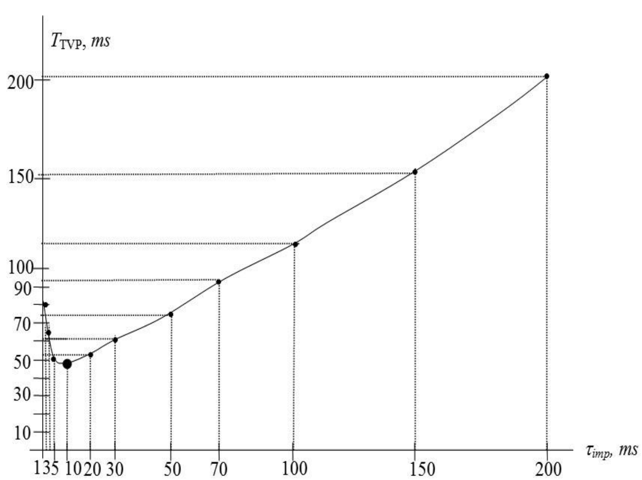


Figure 7. Dependence of visual perception on the pulse duration diagram.

V. EXPERIMENTAL RESULTS

Experimental research was conducted in a group of 30 operators. The time of vision system restoration, the time of visual perception and lability were evaluated, basing on the methods mentioned above. The source of light pulses was LED of yellow color.

The experimental results are presented in the table below.

TABLE I.

№	Estimates of the temporal parameters		
	Value of time parameter	***Individual values***	***Median for the group***
1	Recovery time	12,7 … 37,9 ms	22,9 мс
2	Time of visual perception	62,7 … 87,9 ms.	72,9 мс
3	Lability of the visual system	11,1 … 24,6 Hz.	19,4 Гц

The data concerning the time of restoration and the time of visual perception as a whole correlate with the data from other researchers. The accuracy of the methods developed above is 18-25% higher in comparison with the methods known.

The data on lability of visual system coincide with the data received by means of electrophysiological methods and do not coincide with the data received by means of the CFF method.

In the method described above we eliminate masking effect obtained during the perception of paired light pulses. This explains the difference in the data obtained from those obtained by the CFF method.

The accuracy of the method described above is 17-38 % higher than of the CFF method.

The main advantage of the methods developed is their simplicity; they provide an opportunity to evaluate visual exhaustion.

VI. CONCLUSION

Thus, we have developed new methods of studying temporal aspects of human-computer interaction in the perception of visual information.

All the methods developed are protected by Russian patents for inventions.

We developed the model of visual perception; established the dependence of time of visual perception on the duration of impulses.

Separation on rod and cone paths allows simulating the change of light and the spectrum of light pulses. This is determined by the fact that the rods are most sensitive to blue color and the cones are most sensitive to yellow one. In this RGB triplet and theme map of the ganglion cell were ignored.

We obtained new data about the time of restoration, the time of visual perception and lability of a vision system.

The data are consistent with the simulation data, which confirms the adequacy of the model.

The results obtained can be used to develop human-focused communication devices, displays as well. They will allow to choose characteristics of video terminal devices so that the maximum consistency of a video display device and a vision system be achieved.

It will also be possible to raise the exchange intensity in the "display-person" system, to increase the reliability of data communication, and to reduce the negative influence of a video terminal on a vision system.

The results of the work can be used to develop human centered HCIs, as well as adaptive HCIs taking into account the peculiarities of visual perception and their change at shift-fatigue of an operator.

ACKNOWLEDGMENT

The research results reflected in the paper are obtained with grant support under the analytical departmental target program «Development of scientific potential of the higher school (2009-2010)» № 2.1.2/4841"Methodical, algorithmic and soft-technical maintenance of research of time aspects of touch perception of the person-operator» and Russian Foundation for Basic Research № 10-08-97019.

Special thanks to Dr. V. Rozhentsov for collaborative research.

REFERENCES

[1] J. Zhang, V.L. Patel, T. Shortliffe, M. Freed, and R. Remington, "The nature of human error: An emerging interdisciplinary perspective", Proceedings of the 22-th Annual Conference of the Cognitive Science Soceity (CogSci2000), Institute for Research in Cognitive Science University of Pennsylvani, USA. 2000, date accessed 23/12/10, http://acad88.sahs.uth.tmc.edu/research/publications/human%20error.doc.

[2] C.P. Shelton, "Human Interface/Human Error", Report, Carnegie Mellon University, Feb. 1999, date accessed 23/12/10, http://www.ece.cmu.edu/~koopman/des_s99/human.

[3] A. Kumar and P.K. Sinha, "Human Error Control in Railways", Jordan Journal of Mechanical and Industrial Engineering, vol. 2, No 4, December, 2008, ISSN 1995-6665, pp. 183 - 190.

[4] G.D. Abowd, "Formal Aspects of Human-Computer Interaction", Doctoral Dissertation, Univ. of Oxford, UK, June 1991, 232 p.

[5] L.C. Begosso and L.V. Filgueiras, "Human error simulation as an aid to HCI design for critical systems", Proceedings of the VII Brazilian symposium on Human factors in computing systems (IHC '06), pp. 120-127, doi:10.1145/1298023.1298040.

[6] F.A. Schreiber, "A Model for the Human Computer Interface Evaluation in Safety Critical Computer Applications", Proceedings of the IEEE International Conference and Workshop: Engineering of Computer-Based Systems, March 1998, Jerusalem, Israel, pp. 178-182.

[7] M. Jipp, E. Badreddin, C.Abkai, and J.Hesser, "Individual Ability-based System Configuration. Cognitive Profiling with Bayesian Networks", Proceedings of the IEEE International Conference: Systems, Man and Cybernetics (SMC 2008), Oct. 2008, pp. 3359 – 3364, doi:10.1109/ICSMC.2008.4811816.

[8] A.Yu. Burov, "Psychophysiological features of adaptation to stress and risk assessment at operator's work", Proceedings of the 18th Digital Avionics Systems Conference, 1999, vol. 2, pp. 6.D.6-1 – 6.D.6-4, doi:10.1109/DASC.1999.821983.

[9] J.D.Cress, L.J.Hettinger, W.T. Nelson, and M.W. Haas, "An approach to developing adaptive interface technology for advanced airborne crew stations", Dayton Section Symposium, 1997., The 14th Annual AESS/IEEE. 1997. - Fairborn, OH , USA, pp. 5-10, doi:10.1109/DAYTON.1997.595089.

[10] M.Hou and R.D. Kobierski, "Intelligent adaptive interfaces: Summary report on design, development, and evaluation of intelligent adaptive interfaces for the control of multiple UAVs from an airborne platform", DEFENCE RESEARCH AND DEVELOPMENT TORONTO (CANADA). 2006. – 88 p.

[11] D. Leadbetter, A. Hussey, P. Lindsay, A. Neal, and M. Humphreys, "Towards Model Based Prediction of Human Error Rates in Interactive", J. Australian Computer Science Communication, vol. 23, January-February 2001, pp. 3-12.

[12] N. Sarter, "Coping with complexity through adaptive interface design", Proceedings of the 12th international conference on Human-computer interaction: intelligent multimodal interaction environments, Beijing, China, 2007, pp. 493-498.

[13] ATCA, "Next Generation Air Transportation System Transformation [Special Edition]", Journal of Air Traffic Control 48(1), 2006, pp. 1-64.

[14] J. Wreathall, D.D. Woods, A.J. Bing, and K. Christoffersen, "Relative Risk of Workload Transitions in Positive Train Control", Technical Reports (DOT/FRA/ORD-07/12), Federal Railroad Administration, Washington, D.C., March 2007, 55 p.

[15] U. Ahlstrom, "An Eye for the Air Traffic Controller Workload", Journal of the Transportation Research Forum, 2007, vol. 46, No 3, pp. 103-108.

[16] G. Song, "Analysis of Web Page Complexity Through Visual Segmentation", Human-Computer Interaction. HCI Applications and Services: Lecture Notes in Computer Science, 2007, vol. 4553, 2007, pp. 114-123, doi: 10.1007/978-3-540-73111-5_14

[17] F. Joseph Dyro, "Clinical engineering handbook", Elsevier, Academic press series in biomedical engineering, Medical, 2004, 674 p.

[18] E.F. Wells, G.M. Bernstein, B.W. Scott, P.J. Bennett, and J.R. Mendelson, "Critical flicker frequency responses in visual cortex" Experimental Brain Research, vol. 139, No.1, pp. 106-110, doi: 10.1007/s002210100721.

[19] J.M. Henderson, "Two representational systems in dynamic visual identification", J Exp Psychol Gen., 1994, Dec., vol.123, No. 4, pp. 410-436.

[20] D. Harris, "Human-Computer Interaction: Theory and Practice (part 2)", Volume 2 (Human Factors and Ergonomics), 2003, 1480 p.

[21] D.D. Dilks, C.I. Baker, Y. Liu, and N. Kanwisher, "Referred Visual Sensations": Rapid Perceptual Elongation after Visual Cortical Deprivation", The Journal of Neuroscience, 2009, July, vol. 29(28), pp. 8960–8964.

[22] T. Gollisch and M. Meister, "Modeling convergent ON and OFF pathways in the early visual system", Biol Cybern, 2008, November, 1999, vol.(4-5), pp. 263–278, doi: 10.1007/s00422-008-0252-y.

Application of User Involvement and Quality Function Deployment to Design Intelligent Service Systems

Víctor Acinas Garzón
Communications Division
Instalaciones Inabensa S.A.
Madrid, Spain
victor.acinas@inabensa.abengoa.com

Abstract—**This paper presents a new method for capturing user needs in product design. The method links user needs to product design by combining user involvement techniques together with Quality Function Deployment (QFD). The methodology is applicable for product development as well as the design of novel products that cannot base its design on incremental improvements of existing products. The methodology is illustrated through application to the design of a novel intelligent service system that aims to create channels for communication and interaction between people with some kind of special need and their environment.**

Keywords-intelligent service system design; user involvement; quality function deployment; accessibility; inclusion.

I. INTRODUCTION

The information and knowledge society provides scant attention to users with special needs, despite the increase in number and needs [1], [2]. The information and knowledge society should not only cover mainstream user's demand, but also individuals needs through software services. It is of great importance to develop new methodologies that take users with special needs into consideration from the early steps of product design.

A widely adopted product design methodology is Quality Function Deployment (QFD). QFD is spread across different fields, including software development [3][4]. This methodology translates the voice of the customer into design requirements and product specifications through the House of Quality Matrix, see Fig. 1. However, QFD tool can find difficulties to capture [5], understand and organize user needs [6] as well as connecting technical requirements and setting targets according to quality of service [7]. Moreover, QFD can not be used in multiple product design [8] and innovative interactive systems [9].

To improve QFD and exploit its flexibility [10], QFD is combined with other methodologies such as Kano Model, FMEA, KC and TRIZ [10][11][12][13]. However, modifications to QFD lack from an active user involvement that is of special relevance for social applications [14].

For these reasons, INREDIS (RElationship INterfaces between the environment and DISabled people) project [15] developed a design methodology that on one side researches user needs, including user with special needs, in connection with the information and knowledge society and, on the other, current and emerging technological solutions that can meet user needs through tailored software services.

The design methodology was used to design an intelligent service system that creates channels for communication and interaction accessible to all users. INREDIS on-going work is to develop the intelligent service system that will enable the creation of channels for communication and interaction among people with special needs and the information and knowledge society according to the design methodology results. The intelligent service system goal is to deliver accessible software services that meet a wider range of needs than current technological solutions and is capable of evolving in relation to the needs of the users and the environment. As a result, in this paper INREDIS design methodology is illustrated in connection with the design of INREDIS intelligent service system design. Thus, in connection with the design of INREDIS intelligent service, Section II describes the design methodology, Section III presents the design methodology results and Section IV the conclusions and future work.

II. DESIGN METHODOLOGY

INREDIS design methodology consists of four major steps as outlined in Fig. 1. These steps represent a comprehensive analysis of technological solutions and users, including those with special needs. This methodology overcomes traditional QFD difficulties in capturing [5] understanding and organizing user needs [6] by complementing traditional methods used in QFD [5] with user involvement. Furthermore, by relying on user involvement, it is possible to develop novel products [9] [16].

The first step of the intelligent system design is to research the relationship between users, including those with special needs, and technology, in particular with technologies that are characteristic of the information and

 ISBN: 978-1-61208-117-5

knowledge society. This analysis seeks to identify the main barriers that users with and without functional diversity encounter in using technology.

For this, user tests [17][18], heuristic analysis [19], surveys [20], interviews [14][21] and the people led innovation methodology [16] were carried out in order to understand the real-world problems in interaction between users, devices and technologies in the environment and pinpoint interoperability issues among devices and systems in different contexts. This step yields a list of user requirements as well as importance values of each requirement.

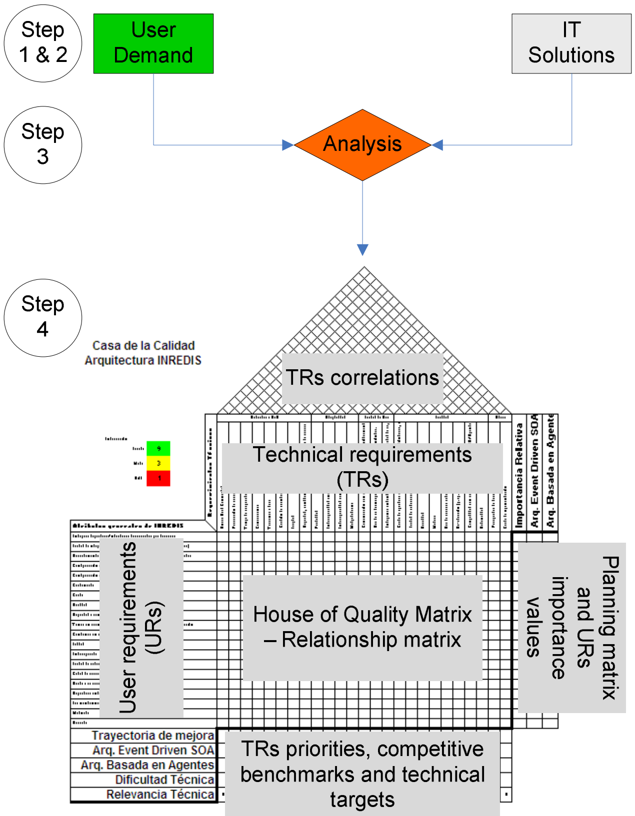


Figure 1. INREDIS intelligent service system design methodology outline.

The contexts of study are: telecare, domestic, urban, banking, buy/sell, educational and work. Furthermore, six types of user profiles with specific special needs are evaluated, visual, hearing, mobility, handling, cognitive and elderly. These contexts and profiles are considered across all the design steps.

The second step is to analyze the current technological environment, taking into account different emerging technologies that may contribute to the development of the technological environment and to further the state of the art of these technologies. Thus, this step investigates the phenomenon of converging technologies in the information and knowledge society and relevant to users with special needs, while it takes into account other usage social considerations. This step yields a list of technical requirements as well as their correlations.

Hence, the results from step one and two are analyzed in step three to select the most relevant issues for the design and development of an intelligent service system for users with special needs. This analysis yields the values to include in the QFD, which are, on one side, the technical requirements, priorities, targets and competitive benchmarks with other solutions, and, on the other, the planning matrix. The selection of these is done through different group sessions that include experts from INREDIS project with user and technical backgrounds. The resulting issues form the starting point for the next and final step.

Drawing from the information gathered to this point, the fourth step ensures relating user and technical perspectives through a QFD [3][22]. The House of Quality function deployment matrix provides a cornerstone to engage technical and user experts in a discussion on the relative importance and relationship among technical characteristics and customer attributes when designing and developing an intelligent system for users with special needs. Furthermore, this will enable to prioritize technical solutions requirements relevant to the information and knowledge society and in line with users with special needs. This is done by weighting the importance of the user input together with the technical requirements and their relation.

III. DESIGN METHODOLOGY RESULTS

The described design methodology surfaced accessibility and usability barriers as well as preferences with interconnected with technical characteristics. These shows the way to develop novel solutions or implementations that break down barriers with the information and knowledge society to provide more efficiently and effectively software services. The design yielded that the most relevant technical characteristics when designing an intelligent service system to create accessible channels for communication and interaction are:

1. Openness: must be based on standards and/or free and open source technologies to ease adoption.
2. Interoperability: the intelligent service system has to interoperate with the widest range of software services and technological solutions.
3. Portability: the intelligent system must be able to operate on different devices with different platform solutions.
4. Implementation cost: the intelligent service system must be easy to access/install and use as well as to update.
5. IT architecture compatibility: the intelligent service system must be compatible with other IT architectures and paradigms such as service oriented, event driven and intelligent agent architectures.

Interestingly, these technical characteristics can be summarised, leaving implementation costs aside, as a surge

 ISBN: 978-1-61208-117-5

towards an intelligent service system that interoperates with current or upcoming solutions whether open, private or standardised. Furthermore, as to integrate its usage and functionality in the daily life of the user and the contexts under study, it shall also be available everywhere. For these reasons, the main focus shall be placed on creating an intelligent service system that is accessible, interoperable and ubiquitous to users with special needs, and, thus, to everyone, see Figure 2.

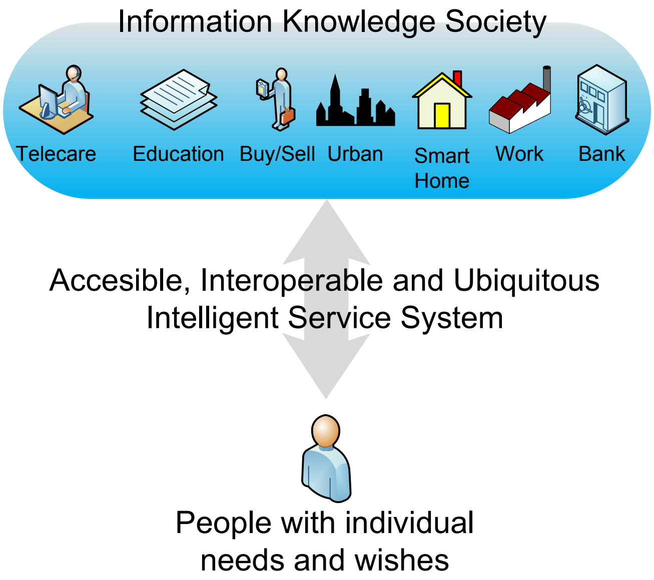


Figure 2. Outline of intelligent service system design.

Only by implementing an intelligent service system focused on accessible, interoperable and ubiquitous technical characteristics users will adopt it to impact daily life. Ultimately, users with special needs can communicate and interact with the information and knowledge society in virtually all contexts.

A. Current software services market

Despite the growing needs and number of users with special needs, according to the design studies in step 1 and step 2, the current market scenario does not offer accessible software services that fulfil all needs. So far, the trend has been to incorporate technological advances as adaptations to existing software services that make them accessible and usable by e.g. elderly and people with disabilities. However, the studies show that, often, adapted software services do not even entirely cover targeted user needs. Since these services were planned from inception to mainstream users, adaptations to cover other needs are at best, troublesome and costly and at worse, unattainable. Fig. 3 shows the current market situation, software services depicted in a fading colour indicate needs not fully covered.

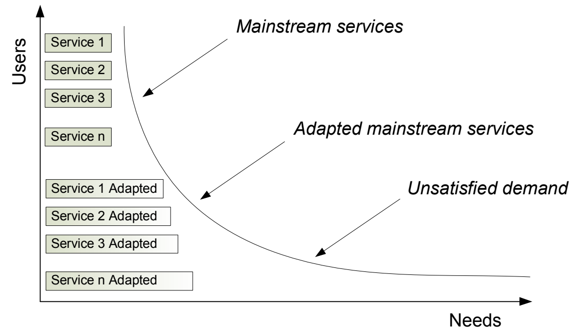


Figure 3. Market software services situation according to the number of users and needs.

For instance, despite banking software services may increase the fonts to enable its use for people with visual impairments; the user may require plain language to fully understand what is shown. Even more, adaptations often involve heterogeneous modifications that hinder the development of standardized technologies that can operate with any type of device [23]. This creates a fragmented and non-structured market jeopardising user involvement in the information and knowledge society.

B. Future software market

For these reasons, the future market of software services should address fully the user needs and overcome software service adaptation deficiencies [24]. Hence, according to the design methodology an intelligent service system has to create channels for communication and interaction which provide tailored services to fulfil user's special needs comprehensively through accessible, interoperable and ubiquitous software services. In doing so, the services may need to be, not only aggregated with others but complemented with assistive technologies. Fig. 4, illustrates the composition of tailored software services.

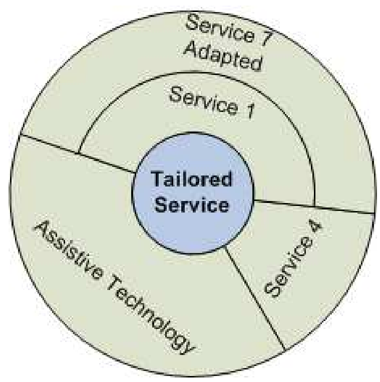


Figure 4. Example of tailored software service composition.

Interestingly, tailored software services are not only of use for people with special needs. A common example is the difficulties everyone experiences when looking rich-colour

 ISBN: 978-1-61208-117-5

interfaces under direct sunlight. It would be of help to all users if the interface changed automatically to a high-contrast. Fig. 5 shows how tailored software services generated through an accessible, interoperable and ubiquitous intelligent service system can fullfil users' needs demand through service aggregation.

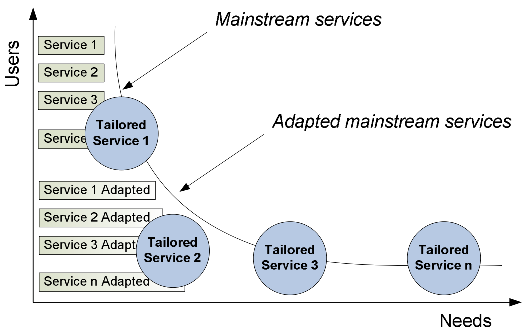


Figure 5. Market scenario according to the number of users and needs.

Thus, rather than developing or modifying software services an accessible, interoperable and ubiquitous intelligent service system builds on existing software services and creates, in line with user needs, channels for communication and interaction with the information knowledge society [25]. Furthermore, this design leaves the door open to all software service providers to interoperate with the intelligent service system and, thus, improves user satisfaction, while co-existing with current mainstream (adapted) software services.

IV. Conclusion and Future Work

The application of user involvement together with the Quality of House function deployment presented in this paper enables building an intelligent service system to meet comprehensively user needs for software services. This approach points out the importance to create tailored software services that meet not only mainstream users but also users with special needs. These tailored software services can be created through the design of an intelligent service system that implements the characteristics of accessibility, interoperability and ubiquity and takes into account the user preferences. In this way, current software services can be accessed and use anywhere together with others to provide tailored services that match with the user needs and wishes.

In line with the presented design, the next step is to develop INREDIS intelligent service system to build tailored software services that meet both mainstream users and users with special needs. Later on, INREDIS intelligent service system will be validated in real life tests within the considered project scenarios; mobility, electronic media, telecare, domestic, urban, banking, shopping, educational and work.

Acknowledgment

The research described in this paper arises from a Spanish research project called INREDIS (INterfaces for RElations between Environment and people with DISabilities) [http://www.inredis.es]. INREDIS is led by Technosite and funded by CDTI (Industrial Technology Development Centre), under the CENIT (National Strategic Technical Research Consortia) Programme, in the framework of the Spanish government's INGENIO 2010 initiative.

The opinions expressed in this paper are those of the authors and are not necessarily those of the INREDIS project's partners or of the CDTI.

References

[1] MeAC. Measuring Progress of eAcessibility in Europe. Assessment of the Status of eAccessibility in Europe. Main Report. October (2007).

[2] Ageing characterises the demographic perspectives of the European societies. Eurostat report 72/2008.

[3] L. Chan and M. Wu, "QFD: a literature review", European Journal of Operational Research, Vol. 143, pp.463-497, 2002.

[4] J. R. Sharma, S.A. Tabarno, and A. M. Rawani, "Integrating QFD with software development engineering for higher customer satisfaction", Indian Journal of Information Science & Technology, May, Vol. 1, No. 2, pp.35-43, 2006.

[5] R. E. Zultner, "Objectives: The missing piece of the requirements puzzle," Proc. Structured Development Froum XI, pp. 1-11, 1990.

[6] J. J. Cristiano, J.K. Liker, and C. C. White, "Key factors in the succesful application of quality function deployment," IEEE Trans. Eng. Manag., vol.48, no. 1, pp. 81-95, 2001.

[7] L. Xiaoqing and Z. Lianzhang, "Design of SOA Based Web Service Systems Using QFD for Satisfaction of Quality of Service Requirements," Proc. IEEE International Conference on Web Services, pp. 567-574, 2009.

[8] Y. Sireli, P. Kauffmann, and E. Ozan, "Integration of Kano's Model into QFD for Multiple Product Design," IEEE Transactions on Engineering Management, 54(2), pp.380-390, 2007.

[9] A. Dearden, S. Howard, "Capturing user requirements and priorities for innovative interactive systems," Computer Human Interaction Conference, pp. 160–167, 1998.

[10] A. Hassan, A. Siadat, J. Y. Dantan, and P. Martin, "Interoperability of QFD, FMEA, and KCs methods in the product development process," IEEE International Conference on Industrial Engineering and Engineering Management, pp. 403-407, 2009.

[11] S. Li, Y. Ma, G. Yang, and Y. Li, "An Integrated Mode Research of QFD and TRIZ and Its Applications," Second International Workshop on Computer Science and Engineering, vol. 1, pp.548-552, 2009.

[12] C. Ankur, R. Jain, A. Singh, and P. Mishra, "Integration of Kano's Model into quality function deployment (QFD)," International Journal of Advanced Manufacturing Technology, pp. 268-278, 2010.

[13] K. S. Chin, L.S. Zheng, and L. Wei, "A hybrid roughcut process planning for quality," International Journal of Advanced Manufacturing Technology, vol. 22, pp. 733–743, 2003.

[14] M. Valles, Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional. Síntesis. Madrid, 1997.

[15] INREDIS: INREDIS project, www.inredis.es last visited on December 2010.

[16] Institute of Innovation for Human Well-being (I2BC), http://www.i2bc.es/inicio/pli last visited on December 2010.

[17] J. Lores, T. Granollers, and S. Lana, Introducción a la interacción persona-ordenador. AIPO. Madrid, 2002.

[18] K. Pernice and J. Nielsen, How to Conduct usability Evaluations for Accesibility: Methodology Guidelines for Yesting Websites and Intranets With Users Who Use Assistive Technology. Freemont: Nielsen Norman Group, 2001.

[19] J. Nielsen, Usability engineering. London, UK: Academia Press Profesional, Boston, MA., 1993.

[20] F. Alvira, La encuesta: una perspectiva general metodológica. CIS. Madrid, 2004.

[21] L. Alonso, La mirada cualitativa en sociología. Fundamentos. Madrid, 1998.

[22] Y. Akao, Quality Function Deployment: Integrating Customer Requirements into Product Design. Productivity Press. Cambridge, 1990.

[23] A. Sloane, Home-Oriented Informatics and Telematics, Springer-IFIP, pp. 323-335. Springer Boston, 2005.

[24] C. A. Velasco, Y. Mohamad, and J. Pullmann, “Dynamic Adaptation of Web 2.0 Applications by Combining Extended Device Profiles. Universal Access in Human-Computer Interaction. Applications and Services,” Lecture Notes in Computer Science, Volume 5616, pp. 797-802, Springer, Berlin / Heidelberg, 2009.

[25] M. Pous and L. Ceccaroni, “Multimodal Interaction in Distributed and Ubiquitous Computing,” Fifth International Conference on Internetand Web Applications and Services (ICIW), pp. 457-462, Barcelona, 2010.

Usability Heuristics for Grid Computing Applications

Cristian Rusu[1], Silvana Roncagliolo[1], Gonzalo Tapia[1], Danae Hayvar[1], Virginica Rusu[2], Dorian Gorgan[3]

[1]Pontificia Universidad Catolica de Valparaiso, Valparaiso, Chile
e-mail: cristian.rusu@ucv.cl, silvana@ucv.cl, g.tapia.aldea@gmail.com, dhayvar@gmail.com

[2]Universidad de Playa Ancha, Valparaiso, Chile
e-mail: virginica.rusu@upla.cl

[3]Technical University of Cluj-Napoca, Cluj-Napoca, Romania
e-mail: dorian.gorgan@cs.utcluj.ro

Abstract — **Usability evaluation for applications based on emerging information technology brings new challenges. Grid Computing is a relatively new, distributed computing technology, based on sharing different types of computational resources, located in various geographic locations. Technical knowledge of grid users is expected to decrease in the future; that is why the usability of Grid Computing applications will become a main issue. There is a need for new usability evaluation methods or at least for the use of traditional evaluations in novel ways. A set of heuristics is proposed and validated, in order to help the heuristic evaluations of Grid Computing applications.**

Keywords – usability; usability heuristics; grid computing applications

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, which are not necessarily at the same geographical location. The process is transparent for users, allowing the use of resources as a single supercomputer.

There are many projects worldwide making use of grid infrastructure, most of them for scientific purposes. Current research usually focuses on Grid Computing based application development from a technical point of view, rather than a user–centered approach. There is a necessity to establish methodologies that could lead to applications with a high level of usability. Such methodologies have to include accurate usability evaluations.

The usability evaluation of a software system is one of the most important stages in the user centered design approach. It allows obtaining the usability characteristics of a software system and the extent to which the usability attributes, usability paradigms and usability principles are being implemented [1].

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 paper focuses on usability evaluation of Grid Computing applications by heuristic evaluations. A set of 12 specific usability heuristics is proposed and validated. Section 2 highlights the basic features of Grid Computing applications and the challenge of their usability evaluation. Section 3 presents a proposal of usability heuristics, which validation is described in section 4. Conclusions are presented in section 5.

II. USABILITY IN GRID COMPUTING APPLICATIONS

Grid Computing applications aim to solve problems that usually require a large number of processing cycles, storage and access of large amounts of data, sometimes distantly located or administered by various organizations, access to specialized equipment, and inter-organizational collaboration of users.

Grid Computing is defined by a set of basic features: abstraction, resource sharing, flexibility, decentralized management and control, scalability, high performance, security, generalization, personalization, heterogeneity. The use of Grid Computing technology has significant advantages: allows independent administrative domains, offers a good cost/performance ratio, enables the sharing of multiple types of resources, allows the integration of heterogeneous systems and resources, offers great fault adaptability and the capacity of easily adding new resources or replacing old ones, to provide new features [3] [4].

Grid Computing cover a wide range of application fields, and it is particularly useful in science, where experiments, simulations, or other research need a power that cannot be offered by standalone supercomputers or clusters of isolated organizations. Some Grid Computing applications are processor-intensive; others may require massive storage.

Depending on the type of resources that are mainly used, the main types of grid are:

- *Computing Grid*: designed to provide as much computing power as possible.

 ISBN: 978-1-61208-117-5

- *Data Grid*: allows the management and sharing of huge amounts of distributed data.
- *Service Grid*: provides services that cannot be provided by a single computer.
- *Equipment Grid*: provides access to special type of equipment, not easily available, either because of their high cost, geographic localization, or other difficulties.

The use of Grid Computing applications includes:

- *Job Submission*: Users specify the definition of tasks to execute and data to use.
- *Monitoring*: A monitoring interface allows users to check the status of the processing.
- *Visualization*: When jobs processing has finished, a visualization interface shows the results.
- *Web Portals*: Nowadays many Grid Computing based projects offer access to their services through Web applications, by Web portals.

Grid Computing users, their knowledge and specific tasks may be categorized as follows [5]:

- *Service end-user*: low technical knowledge; data input and grid services user.
- *Service end-user execute*: some technical knowledge; job submission.
- *Power user agnostic of grid resource nodes*: high technical knowledge; application development.
- *Power user requiring specific grid resource nodes*: high technical knowledge; application development, aware of specific grid resources nodes.
- *Power user developing a service*: high technical knowledge; services development.
- *Service provider*: high technical knowledge; identity and authorization management.
- *Infrastructure system administrator*: high security and infrastructure knowledge; grid nodes system administration.

The current use of Grid Computing is at the hand of experts and researchers with extensive (specific and technical) knowledge. Most of the Grid Computing applications users have nowadays similar knowledge and similar background, but we may infer that such similarities will be no longer the rule.

It is expected that in the future the technical knowledge of grid users will decrease. The number of users belonging to the first and the second of the above mentioned categories is growing fast. That is why we think the usability of Grid Computing applications will become a main issue.

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 [6].

Usability is not a one-dimensional property of the interface; it is a combination of factors. Effectiveness refers to the accuracy level that the user achieves goals. Efficiency refers to the resources employed by the user to accomplish these goals. Finally, the satisfaction is related to the comfort of the user during the interaction with the software system.

Usability evaluation methods are commonly divided into inspection and testing methods. Inspection methods find usability problems based on the expertise of usability professionals. Testing methods find usability problems through the observation of the users while they use (and comment on) a system interface [7].

Heuristic evaluation is a widely used inspection method. A group of evaluators inspect the interface design based on the usability principles (heuristics). 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 significant.

Grid Computing has evolved from scripts to portals and Web interfaces, therefore usability heuristics for Grid Computing should be developed from this new perspective [8] [9] [10].

III. Defining Grid Computing Usability Heuristics

In order to develop specific usability heuristics for Grid Computing applications the following steps were followed [11]:

- An *exploratory* stage, to collect bibliography related with the main topics of the research: Grid Computing applications, usability evaluation, and usability heuristics.
- A *descriptive* stage, to highlight the most important characteristics of the previously collected information, in order to formalize the main concepts associated with the research.
- A *correlational* stage, to identify the characteristics that the usability heuristics for Grid Computing applications should have, based on traditional heuristics and case studies analysis.
- An *explicative* stage, to formally specify the set of the proposed heuristics, using a standard template.
- A *validation* (experimental) stage, to check the new heuristic against traditional heuristics by experiments, through heuristic evaluations performed on selected case studies, complemented by user tests.
- A *refinement* stage, based on the feedback from the validation stage.

Based on the well known and widely used Nielsen's 10 heuristics and extensively analyzing several Grid Computing applications, especially *GreenView* [12], a set of 12 new usability heuristics was developed for heuristic evaluations of Grid Computing applications.

Grid Computing heuristics were specified using the following template:

- *ID, Name and Definition*: Heuristic's identifier, name and definition.
- *Explanation*: Heuristic's detailed explanation, including references to usability principles, typical

usability problems, and related usability heuristics proposed by other authors.

- *Examples*: Examples of heuristic's violation and compliance.
- *Benefits*: Expected usability benefits, when the heuristic is accomplished.
- *Problems*: Anticipated problems of heuristic misunderstanding, when performing heuristic evaluations.

The 12 proposed heuristics were grouped in three categories: (1) *Design and Aesthetics*, (2) *Navigation* and (3) *Errors and Help*. A summary of the proposed heuristics is presented below, including heuristics' ID, name and definition.

Design and Aesthetics Heuristics:

(H1) *Clarity*: A Grid Computing application interface should be easy to understand, using clear graphic elements, text and language.

(H2) *Metaphors*: A Grid Computing application should use appropriate metaphors, making the possible actions easy to understand, through images and familiar objects.

(H3) *Simplicity*: A Grid Computing application should provide the necessary information in order to complete a task in a concise (yet clear) manner.

(H4) *Feedback*: A Grid Computing application should keep users informed on the jobs' progress, indicating both the global and the detailed state of the system. The application should deliver appropriate feedback on users' actions.

(H5) *Consistency*: A Grid Computing application should be consistent in using language and concepts. The forms of data entry and visualization of results should be consistent.

Navigation Heuristics:

(H6) *Shortcuts*: A Grid Computing application should provide shortcuts, abbreviations, accessibility keys or command lines for expert users.

(H7) *Low memory load*: A Grid Computing application should maintain the main commands always available. It should offer easy to find elements, functions and options.

(H8) *Explorability*: A Grid Computing application should minimize navigation and should provide easy, clear, and natural ways to perform tasks.

(H9) *Control over actions*: A Grid Computing application should offer ways to cancel a running task or process. It should allow undo and/or changes of actions.

Errors and Help Heuristics:

(H10) *Error prevention*: A Grid Computing application should prevent users from performing actions that could lead to errors, and should avoid confusions that could lead to mistakes.

(H11) *Recovering from errors*: A Grid Computing application should provide clear messages, hopefully indicating causes and solutions of errors.

(H12) *Help and documentation*: A Grid Computing application should provide an easy to find, easy to understand, and complete online documentation. It should provide contextual help and glossary of terms for novice users.

Table 1 presents the mapping between Grid Computing 12 heuristics and Nielsen's 10 heuristics [13].

TABLE I. MAPPING BETWEEN GRID COMPUTING HEURISTICS AND NIELSEN'S HEURISTICS

Grid Computing Heuristics		**Nielsen's Heuristics**	
ID	***Definition***	***ID***	***Definition***
H1	Clarity	N2	Match between system and the real world
H2	Metaphors		
H3	Simplicity	N8	Aesthetic and minimalist design
H4	Feedback	N1	Visibility of system status
H5	Consistency	N4	Consistency and standards
H6	Shortcuts	N7	Flexibility and efficiency of use
H7	Low memory load	N6	Recognition rather than recall
H8	Explorability	N3	User control and freedom
H9	Control over actions		
H10	Error prevention	N5	Error prevention
H11	Recovering from errors	N9	Help users recognize, diagnose, and recover from errors
H12	Help and documentation	N10	Help and documentation

Heuristics H1 and H2 particularize Nielsen's N2 heuristic. Heuristics H8 and H9 denote Nielsen's N3 heuristic. Both N2 and N3 heuristics where detailed and particularized based on the characteristics of the Grid Computing applications, their evolution from scripts to Web interfaces, and the new (heterogeneous) type of users they have.

Heuristic H3 particularizes Nielsen's N8 heuristic, emphasizing the complex tasks that Grid Computing users have to deal with. Heuristic H4 particularizes Nielsen's N1 heuristic into the context of Grid Computing applications, detailing specific feedback requirements. As there are not yet widely recognized standards for Grid Computing applications, heuristic H5 particularize Nielsen's N4 heuristic, stressing the dominance of the consistency over standards. Heuristic H6 provides more specific means than Nielsen's N7 heuristic, and heuristic H7 specifies more precisely Nielsen's N6 heuristic, based on the characteristics of the Grid Computing applications.

Finally, heuristics H10, H11 and H12 put Nielsen's heuristics N5, N9 and N10 (respectively) into the context of Grid Computing applications.

IV. Validating Grid Computing Usability Heuristics

The 12 proposed Grid Computing usability heuristic were checked against Nielsen's 10 heuristics, using *GreenView* and *GreenLand* as case studies. The potential of the Grid Computing heuristics was also checked in usability evaluations of Grid Computing applications as intercultural collaboration platforms.

A. Case Study: GreenView

GreenView is an environmental application that uses high-resolution satellite measurements in climate related studies, modeling the pollution and the impact that urban spaces have on vegetation (Fig. 1). As *GreenView* authors acknowledge, the development of environmental applications based on Grid infrastructures and dedicated to non-technical experts is a challenging task [14].

GreenView v3.1 was examined by two groups of 4 evaluators each. All 8 evaluators had similar (medium) experience in heuristic evaluations (with Nielsen's heuristics), but no experience in usability evaluation of Grid Computing applications. They all had comparable (low, if some) experience in using Grid Computing applications.

The first group performed a heuristic evaluation of *GreenView*, using only the 12 new (Grid Computing) heuristics (based on the full heuristics' specification). The second group performed a similar heuristic evaluation, but using only the Nielsen's 10 heuristics. Table 2 shows the number of usability problems identify by each group of evaluators.

TABLE II. Number of Usability Problems Identified in *GreenView*, by Heuristics

Group 1: Using Grid Computing Heuristics		**Group 2: Using Nielsen's Heuristics**	
ID	***Number of problems***	***ID***	***Number of problems***
H1	3	N2	1
H2	1		
H3	0	N8	3
H4	2	N1	2
H5	1	N4	1
H6	1	N7	0
H7	0	N6	0
H8	1	N3	1
H9	1		
H10	3	N5	2
H11	2	N9	1
H12	2	N10	1
Total:	**17**	**Total:**	**12**

When using Grid Computing heuristics, more usability problems were captured than using Nielsen's heuristics. Analyzing the total of 29 problems identified by the 8 evaluators, we found that:

- 11 problems (38%) were identified by both groups of evaluators,
- 12 problems (41%) were identified only by the group that used Grid Computing heuristics,
- 6 problems (21%) were identified only by the group that used Nielsen's heuristics.

The results seem to prove that Grid Computing heuristics work better than Nielsen's heuristics. However, the question that arises is: why 6 usability problems were not identified using Grid Computing heuristics? There are two possible reasons:

(1) Grid Computing heuristics were not able to identify these problems, either because there are no appropriate heuristics, or because the heuristics are not properly specified.

(2) Evaluators using Grid Computing heuristics subjectively ignored the problems.

The problems identified only by Nielsen's heuristics were associated to heuristics N8 - *Aesthetic and minimalist design* (3 problems), N1 – *Visibility of system status* (1 problem), N5 – *Error prevention* (1 problem), and N10 – *Help and documentation* (1 problem). The set of Grid Computing heuristics provides the tools that can potentially identify all these problems: H3 - *Simplicity*, H4 - *Feedback*, H10 - *Error prevention*, and H12 - *Help and documentation*, respectively. So, the first hypothesis is unlikely to be the true.

All 6 problems identified only by Nielsen's heuristics were qualified with relatively low severity scores (an average of 2.5 or less, on a five point scale). The second hypothesis seems to be the correct one.

In order to validate the second hypothesis, a usability test was designed and performed, with 5 users. The test was focused on the 6 usability problems identified only by Nielsen's heuristics. All these problems were not in fact perceived as real problems by users, so the second hypothesis was validated.

Analyzing the 12 problems identified only by Grid Computing heuristics, most of them were qualified as sever: 8 of 12 problems had an average severity of 2.5 or superior, on a five point scale. Moreover, 3 of 12 problems had (very high) average severity (3.25).

B. Case Study: GreenLand

GreenLand is an environmental application that processes high-resolution *Landsat* satellite images in order to obtain thematic maps of specific elements, such as land, water, air, and vegetation (Fig. 2).

GreenLand v1.2 was examined by two groups of 3 evaluators each. All 6 evaluators had similar (medium) experience in heuristic evaluations (with Nielsen's heuristics), but no experience in usability evaluation of Grid

 ISBN: 978-1-61208-117-5

Computing applications. They all had comparable (low, if some) experience in using Grid Computing applications.

As in the previous case study, the first group performed a heuristic evaluation using only the 12 new (Grid Computing) heuristic (based on the full heuristics' specification). The second group performed a similar heuristic evaluation, but using only Nielsen's 10 heuristics. Table 3 shows the number of usability problems identify by each group of evaluators.

TABLE III. NUMBER OF USABILITY PROBLEMS IDENTIFIED IN *GREENVIEW*, BY HEURISTICS

Group 1: Using Grid Computing Heuristics		Group 2: Using Nielsen's Heuristics	
ID	***Number of problems***	***ID***	***Number of problems***
H1	4	N2	6
H2	2		
H3	2	N8	4
H4	5	N1	2
H5	5	N4	5
H6	0	N7	1
H7	0	N6	0
H8	2	N3	2
H9	3		
H10	0	N5	0
H11	1	N9	1
H12	2	N10	1
Total:	**26**	**Total:**	**22**

As in the previous case study, more usability problems were captured using Grid Computing heuristics than using Nielsen's heuristics. Analyzing the total of 48 problems identified by the 6 evaluators, we found that:

- 14 problems (29%) were identified by both groups of evaluators,
- 22 problems (46%) were identified only by the group which used Grid Computing heuristics,
- 12 problems (25%) were identified only by the group which used Nielsen's heuristics.

The results proved once again that Grid Computing heuristics work better than Nielsen's heuristics. Problems identified only by Nielsen's heuristics were in fact qualified as minor: only 2 of 12 problems had an average severity over 2, in a five points scale. Most of them were also discharged by a usability test, performed with 4 users.

As in the previous case study, most of the 22 problems identified only by Grid Computing heuristics were qualified as sever. Most of them had an average severity of 2.5 or superior, on a five point scale. Moreover, 1 problem had an average severity of 4, and 3 problems had average severities of 3.67!

C. Grid Computing Applications as Intercultural Collaboration Platform

Grid Computing applications are usually intercultural collaboration platforms, and heuristic evaluations of both *GreenView* and *GreenLand* were cross-cultural challenges. Two European projects were evaluated using the 12 Grid Computing heuristics proposed by a Latin American (Chilean) team. Heuristic evaluators were also Chileans. There were both cultural (Latin American vs. European) and language (English vs. Spanish) barriers.

No significant culture–related problems where highlighted during the heuristic evaluations performed on both *GreenView* and *GreenLand*. The 12 usability heuristics proved to be effective tools when evaluating Grid Computing applications as intercultural collaboration platforms [15].

V. CONCLUSIONS

Grid Computing has nowadays a wide range of applications. Even if the current use of Grid Computing is at the hand of experts and researchers with extensive knowledge, it is expected that in the future the technical knowledge of grid users will decrease. That is why we think the usability of Grid Computing applications will soon become a main issue.

Research usually focuses on Grid Computing based application development from a technical point of view; there is a need for new evaluation methods or at least usability evaluations should be particularized for Grid Computing environments.

A set of 12 specific usability heuristics for Grid Computing applications was proposed. The new heuristics were validated through two case studies. Their potential was also checked in usability evaluations of Grid Computing applications as intercultural collaboration platforms.

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] R. Otaiza, C. Rusu, and S. Roncagliolo, "Evaluating the Usability of Transactional Web Sites", Proc. 3rd International Conferences on Advances in Computer-Human Interactions (ACHI 2010), IEEE Press, pp. 32-37, ISBN-13: 978-0-7695-3957-7, 2010.

[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, pp. 324-328, ISBN: 978-1-4244-3351-3, 2009.

[3] I. Foster, C. Kesselman, and S. Tuecke, “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”, International J. Supercomputer Applications, vol. 15(3), pp. 200-222, 2001.

[4] J. Linderoth, “An Introduction to the Computational Grid”, University of Wisconsin, 2007.

[5] M. Norman, “Types of grid users and the Customer-Service Provider relationship: a future picture of grid use”, University of Oxford, 2006.

[6] International Organisation for Standardisation, ISO 9241, “Software Ergonomics Requirements for office work with visual display terminal (VDT)”, Geneva, 1998.

[7] J. Nielsen, “Usability Engineering”, Morgan Kaufmann Publishers, 1993.

[8] T. Doyle, “GridPP: the UK’s contribution to the international collaboration building a worldwide Grid, the LHC Computing Grid: is the system usable?” University of Glasgow, 2006.

[9] B. Hagemeier and M. Kozlovszky, “TG9 – Grid Portals and User Interfaces”, Research Center Juelich, 2008.

[10] A. Walker, T. White, R. Bruin, and K. Austen, “Usable grid infrastructures: practical experiences from the eMinerals project”, University of Cambridge, 2008.

[11] 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), IEEE Press (accepted paper).

[12] R. Meszaros, I. Lagzi, Z. Barcza, and G. Gelybo, “GreenView Application Specifications”, Eotvos Lorand University, Hungary, 2008.

[13] J. Nielsen, “Ten Usability Heuristics”, Available at www.useit.com/papers/heuristic/heuristic_list.html (last access 12/10/2010).

[14] D. Gorgan, T. Stefanut, V. Bacu, and D. Mihon, “Grid based Environment Application Development Methodology”, Published by Springer in LNCS 5910, pp. 499-506, 2010.

[15] 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, pp. 179-182, ACM 978-1-4503-0108-4/10/08.

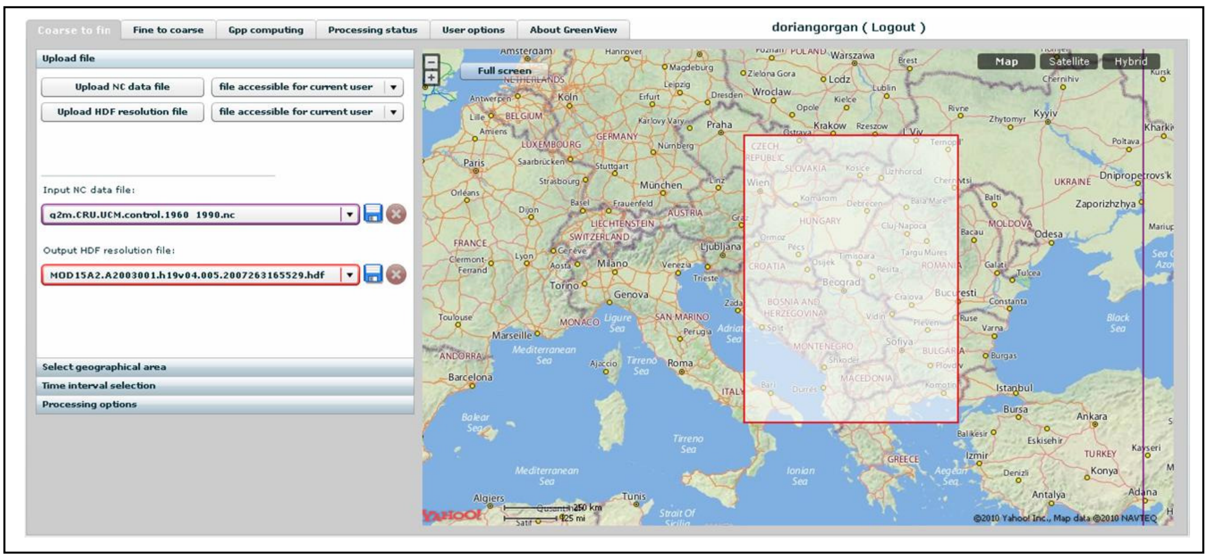


Figure 1. *GreenView* application shotscreen.

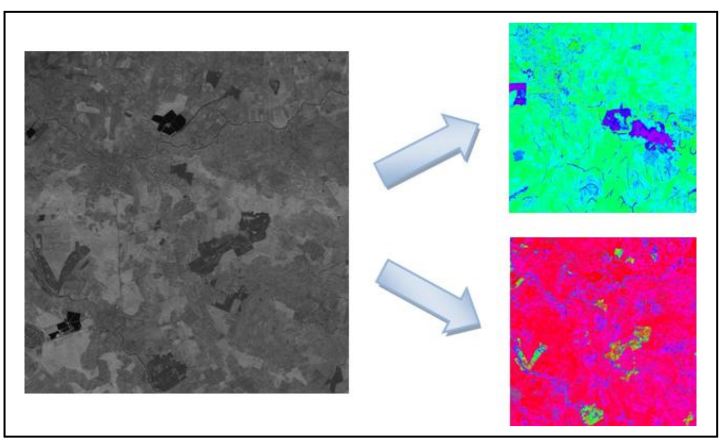


Figure 2. *GreenLand* thematic maps processes.

A Methodology to Establish Usability Heuristics

Cristian Rusu[1], Silvana Roncagliolo[1], Virginica Rusu[2], Cesar Collazos[3]

[1]Pontificia Universidad Catolica de Valparaiso, Valparaiso, Chile
e-mail: cristian.rusu@ucv.cl, silvana@ucv.cl

[2]Universidad de Playa Ancha, Valparaiso, Chile
e-mail: virginica.rusu@upla.cl

[3] IDIS Research Group, Universidad del Cauca, Popayan, Colombia
e-mail: ccollazo@unicauca.edu.co

***Abstract* — 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? A methodology to develop usability heuristics for emerging applications is proposed. The methodology was fully checked in the case of Grid Computing applications, and partially applied in the case of Interactive Digital Television and Virtual Worlds.**

Keywords - usability; usability evaluations; usability heuristics; grid computing; interactive digital television; virtual worlds.

I. INTRODUCTION

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, into the context of new interaction paradigms? How can it be measured? How should we develop for (better) usability? The traditional usability engineering concepts and evaluation methods should be re-examined. There is a need for new evaluation methods or at least for the use of traditional evaluations in novel ways [1]. Frameworks of usability evaluation, including appropriate methods or combination of methods should be established, in order to get more effective and efficient evaluations on new interaction paradigms.

The paper proposes a methodology to establish new usability heuristics. Section 2 highlights the necessity of new heuristics for applications based on emerging information technology, and describes the proposed methodology. Section 3 shows preliminary results of applying the methodology: usability heuristics for Grid Computing, Interactive Television, and Virtual Worlds. Section 4 presents preliminary conclusions and future works.

II. DEFINING NEW USABILITY HEURISTICS

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 [2]. Usability evaluation methods are commonly divided into inspection and testing methods. Inspection methods find usability problems based on the expertise of usability professionals. Testing methods find usability problems through the observation of the users while they use (and comment on) a system interface [3].

Heuristic evaluation is a widely used inspection method. A group of evaluators inspect the interface design based on a set of usability heuristics [4] [5]. 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 significant.

Usability inspections, including heuristic evaluation, are well documented and many publications describe the usage of the methods. Literature usually focuses on describing the advantages and disadvantages of usability evaluation methods but not on how to develop new methods and/or usability heuristics.

Over the last couple of years there is an increasing interest on usability of applications based on emerging information technology [6]. Measuring the usability of such applications was a challenging task. Traditional usability evaluation methods, especially usability heuristics became short.

A methodology to establish new usability heuristics, for specific applications, was gradually defined. The methodology includes 6 stages:

- STEP 1: An *exploratory* stage, to collect bibliography related with the main topics of the research: specific applications, their characteristics, general and/or related (if there are some) usability heuristics.
- STEP 2: A *descriptive* stage, to highlight the most important characteristics of the previously collected information, in order to formalize the main concepts associated with the research.
- STEP 3: A *correlational* stage, to identify the characteristics that the usability heuristics for specific applications should have, based on traditional heuristics and case studies analysis.

 ISBN: 978-1-61208-117-5

- STEP 4: An *explicative* stage, to formally specify the set of the proposed heuristics, using a standard template.
- STEP 5: A *validation* (experimental) stage, to check new heuristics against traditional heuristics by experiments, through heuristic evaluations performed on selected case studies, complemented by user tests.
- STEP 6: A *refinement* stage, based on the feedback from the validation stage.

STEP 1 explores the specific applications that require new usability heuristic.

STEP 2 re-examines the very meaning of usability and its characteristics, in the context of the examined applications.

If literature provides no specific and/or related usability heuristics, Nielsen's 10 well known and extensively used heuristics are used as a basis at STEP 3.

The standard template used at STEP 4 is the following:

- *ID, Name and Definition*: Heuristic's identifier, name and definition.
- *Explanation*: Heuristic's detailed explanation, including references to usability principles, typical usability problems, and related usability heuristics proposed by other authors.
- *Examples*: Examples of heuristic's violation and compliance.
- *Benefits*: Expected usability benefits, when the heuristic is accomplished.
- *Problems*: Anticipated problems of heuristic misunderstanding, when performing heuristic evaluations.

STEP 5 evaluates the set of heuristics defined at STEP 4 against Nielsen's heuristics, in specific case studies. The application is evaluated by two separate groups of evaluators, of similar experience, in equal conditions. One group uses only the set of heuristics defined at STEP 4, while the second group uses only Nielsen's heuristics. Usability problems founded by the two groups are then compared. Three categories of problems are expected:

- (P1) Problems identified by both groups of evaluators,
- (P2) Problems identified only by the group that used the set of heuristics defined at STEP 4,
- (P3) Problems identified only by the group that used Nielsen's heuristics.

New heuristics works well when (P1) and/or (P2) include the highest percentage of problems. Question arises with problems (P3). Why these problems are not identified when using the new set of heuristics? There are basically two possible reasons:

(1) New heuristics are not able to identify these problems, either because there are no appropriate heuristics, or because the heuristics are not properly specified.

(2) Evaluators using new heuristics subjectively ignored the problems.

Hypotheses (1) and (2) may be validated or rejected by complementary evaluations and/or user tests.

STEP 6 refines the set of heuristics defined at STEP 4. Stages 1 to 6 may be applied iteratively. Specific usability checklist may also be developed, detailing usability heuristics and helping heuristic evaluations practice.

III. Applying the Methodology in Practice

A. Usability Heuristics for Grid Computing Applications

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, which are not necessarily at the same geographical location. The process is transparent for users, allowing the use of resources as a single supercomputer.

Nowadays many Grid Computing based projects offer access to their services through Web applications, by Web portals. It is expected that in the future the technical knowledge of grid users will decrease. That is why usability Grid Computing applications' usability will become a main issue.

The methodology described in the previous section was applied in order to establish specific usability heuristics for Grid Computing Applications. A set of 12 new heuristics was developed, grouped in three categories: (1) *Design and Aesthetics*, (2) *Navigation*, and (3) *Errors and Help*. A summary of the proposed heuristics is presented below, including heuristics' ID, name and definition.

Design and Aesthetics Heuristics:

(CGH1) *Clarity*: A Grid Computing application interface should be easy to understand, using clear graphic elements, text and language.

(CGH2) *Metaphors*: A Grid Computing application should use appropriate metaphors, making the possible actions easy to understand, through images and familiar objects.

(CGH3) *Simplicity*: A Grid Computing application should provide the necessary information in order to complete a task in a concise (yet clear) manner.

(CGH4) *Feedback*: A Grid Computing application should keep users informed on the jobs' progress, indicating both the global and the detailed state of the system. The application should deliver appropriate feedback on users' actions.

(CGH5) *Consistency*: A Grid Computing application should be consistent in using language and concepts. The forms of data entry and visualization of results should be consistent.

Navigation Heuristics:

(CGH6) *Shortcuts*: A Grid Computing application should provide shortcuts, abbreviations, accessibility keys or command lines for expert users.

(CGH7) *Low memory load*: A Grid Computing application should maintain the main commands always available. It should offer easy to find elements, functions and options.

(CGH8) *Explorability*: A Grid Computing application should minimize navigation and should provide easy, clear, and natural ways to perform tasks.

(CGH9) *Control over actions*: A Grid Computing application should offer ways to cancel a running task or process. It should allow undo and/or changes of actions.

Errors and Help Heuristics:

(CGH10) *Error prevention*: A Grid Computing application should prevent users from performing actions that could lead to errors, and should avoid confusions that could lead to mistakes.

(CGH11) *Recovering from errors*: A Grid Computing application should provide clear messages, hopefully indicating causes and solutions of errors.

(CGH12) *Help and documentation*: A Grid Computing application should provide an easy to find, easy to understand, and complete online documentation. It should provide contextual help and glossary of terms for novice users.

Following the proposed methodology, the set of Grid Computing usability heuristics was specified, validated and refined in a three-cycle iterative process [7] [8]. New heuristics proved to work better than Nielsen's heuristics in two case studies. Problems (P1) and (P2) were dominant in both cases. Problems (P2) scored 41% in the first case study, and 46% in the second one.

B. *Usability Heuristics for Interactive Television*

Interactive Digital Television (iTV) exceeds the analog TV in several aspects: capacity, better use of the spectrum, greater immunity to noise and interference, better sound and picture quality, potential for transmission of data simultaneously, saving power transmission. However, the main iTV advantage is that the user may interact with the application. Interactivity allows the user to be an active part of the programming, providing the ability to access or extend the information presented, combining multimedia content (audio, video, text), to participate in forums and to control the sequence of information presented [9].

Stages 1 to 4 of the proposed methodology were performed for iTV applications [10]. A set of 14 specific usability heuristics were developed. Stages 5 and 6 are still to be performed. Heuristics were grouped in three categories: (1) *Design and Aesthetics*, (2) *Flexibility and Navigation*, and (3) *Errors and Help*. A summary of the proposed heuristics is presented below, including heuristics' ID, name and definition.

Design and Aesthetics Heuristics:

(ITVH1) *Match between the system and the real world*: iTV should use words, phrases and concepts familiar to the user; the sequence of activities should follow user's mental processes; information should be presented in a simple, natural and logical order; metaphors should be easy to understand; important controls should be represented on screen; there should be an intuitive mapping between them and the real controls.

(ITVH2) *Aesthetic and minimalist design*: iTV should have simple, intuitive, easy to learn and pleasing design; the system should be free from irrelevant, unnecessary and distracting information; icons should be clear and buttons should be labeled; the use of graphic controls should be intuitive; the need for scroll should be minimized; navigation facilities should be present at the bottom of the screen.

(ITVH3) *Consistency and standards*: iTV should use terminology, controls, graphics and menus consistent throughout the system; there should be a consistent look and feel for the system interface; iTV should be consistent with the related standard TV programs, and colors should be consistent between the two systems.

(ITVH4) *Visibility of the system status*: Feedback on system status should be continuously provided.

(ITVH5) *Physical constraints*: Screen should be visible at a range of distances and in various types of lighting; the distance between targets (e.g. icons) and the size of targets should be appropriate; size should be proportional to distance.

(ITVH6) *Extraordinary users*: iTV should use appropriately color restricted; it should be suitable for color-blind users.

Flexibility and Navigation Heuristics:

(ITVH7) *Structure of information*: iTV should have a hierarchical organization of information, from general to specific; related pieces of information should be clustered together; the length of text should be appropriate to the display size and interaction device; the amount of information should be minimized; page titles and headlines should be straightforward, short and descriptive; textual content should be kept to a maximum of two columns.

(ITVH8) *Navigation*: iTV should provide navigational feedback (e.g. showing a user's current and initial states, where they have been, and what options they have for where to go) and navigational aids (e.g. find facilities).

(ITVH9) *Recognition rather than recall*: Help and instructions should be visible or easily accessible when needed; relationship between controls and their actions should be obvious; input formats and units of values should be indicated.

(ITVH10) *Flexibility and efficiency of use*: iTV should allow for a wide range of user expertise; it should also appropriately guide novice users.

(ITVH11) *User control and freedom*: iTV should provide "undo" (or "cancel") and "redo" options; exits should be clearly marked (when users find themselves somewhere unexpected); facilities to return to the top level should be provided, at all stages.

Errors and Help Heuristics:

(ITVH12) *Error prevention*: iTV should offer a selection method provided (e.g. from a list) as an alternative to the direct entry of information; user confirmation should be required before carrying out a potentially "dangerous" action (e.g. deleting something).

(ITVH13) *Help users to recover from errors*: Error messages should adequately describe problems; they should assist in diagnosis and suggest ways of recovery in a constructive way; error messages should be written in a non-derisory tone and refrain from attributing blame to the user.

(ITVH14) *Help and documentation*: iTV should offer clear, direct and simply help, expressed in user's idiom, free from jargon and buzzwords; help should be easy to search, understand and apply.

C. Usability Heuristics for Virtual Worlds

A virtual world is a computer-based simulated persistent spatial environment that supports synchronous communication among users who are represented by avatars [11]. There is interaction between avatars and between avatar and environment. Each virtual world has its own rules.

Stages 1 to 3 of the proposed methodology were applied to some Virtual Worlds case studies. Stage 4 is currently undergoing. To the date, a set of 16 usability heuristics were established. A brief summary, including only heuristics' ID and name, is presented below.

(VWH1) *Clarity*.
(VWH2) *Simplicity*.
(VWH3) *Feedback*.
(VWH4) *Consistency*.
(VWH5) *Low memory load*.
(VWH6) *Flexibility and efficiency of use.*
(VWH7) *Orientation and navigation*.
(VWH8) *Camera control*.
(VWH9) *Visualization*.
(VWH10) *Avatar's customization*.
(VWH11) *World interaction*.
(VWH12) *Law of physics*.
(VWH13) *Communication*.
(VWH14) *Error prevention*.
(VWH15) *Help users to recover from errors.*
(VWH16) *Help and documentation.*

IV. CONCLUSION AND FUTURE WORKS

Heuristic evaluation is a well known and widely used usability inspection method. As it may miss domain specific problems, the use of appropriate heuristics is highly significant.

A methodology to establish new usability heuristics is proposed. The methodology facilitates the development of both usability heuristics and associated usability checklists. The methodology was applied and validated for Grid Computing applications. It is being currently applied for Interactive Digital Television applications and for Virtual Worlds. Future works should include more experiments and validation.

A right balance between specificity and generality should be follow. If heuristics are too specific, they will probably become hard to understand and hard to apply. General heuristics, complemented by specific usability checklists, will probably work better, most of the time. However, there is a need for new usability heuristics especially for applications based on emerging information technology brings new challenges.

As most of the studies recommend, heuristic evaluations should always be complemented by other usability evaluations, especially usability tests.

ACKNOWLEDGMENT

The work was highly supported by the *School of Informatics Engineering* of the *Pontifical Catholic University of Valparaiso – Chile*, especially by the members of *"UseCV" Research Group in Human – Computer Interaction*, and by the *IDIS Research Group* of *University of Cauca - Colombia*.

REFERENCES

[1] 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, pp. 324-328, ISBN: 978-1-4244-3351-3, 2009.

[2] International Organisation for Standardisation, ISO 9241, "Software Ergonomics Requirements for office work with visual display terminal (VDT)", Geneva, 1998.

[3] R. Otaiza, C. Rusu, and S. Roncagliolo, "Evaluating the Usability of Transactional Web Sites", Proc. 3rd International Conferences on Advances in Computer-Human Interactions (ACHI 2010), IEEE Press, pp. 32-37, ISBN-13: 978-0-7695-3957-7, 2010.

[4] J. Nielsen, "Usability Engineering", Morgan Kaufmann Publishers, 1993.

[5] J. Nielsen, "Ten Usability Heuristics", Available at www.useit.com/papers/heuristic/heuristic_list.html (last access 12/10/2010).

[6] C. Rusu, V. Rusu, and S. Roncagliolo, "Usability Practice: The Appealing Way to HCI", Proc. 1st International Conferences on Advances in Computer-Human Interactions (ACHI2008), IEEE Press, pp. 265-270, ISBN-10: 0-7695-3086-9, 2008.

[7] C. Rusu, S. Roncagliolo, G. Tapia, D. Hayvar, 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, pp. 179-182, ACM 978-1-4503-0108-4/10/08.

[8] C. Rusu, S. Roncagliolo, G. Tapia, D. Hayvar, V. Rusu, and D. Gorgan, "Usability Heuristics for Grid Computing Applications", Proceedings of The 4th International Conferences on Advances in Computer-Human Interactions (ACHI 2011), Gosier, France, 2011.

[9] C. Collazos, J. Arciniegas, V. Mondragón, and X. Garcia, "Directrices de diseño y evaluación de la televisión interactiva desde una perspectiva de usabilidad", Proc. Interaccion 2008, Albacete, Spain, pp. 257-260, 2008.

[10] C. Collazos, C. Rusu, J. Arciniegas, and S. Roncagliolo, "Designing and Evaluating Interactive Television from a Usability Perspective", Proc. 2nd International Conferences on Advances in Computer-Human Interactions (ACHI 2009), IEEE Press, ISBN: 978-1-4244-3351-3, pp. 381-385, 2009.

[11] J. Yoonhyuk and K. Hyunmee, "User goals in social virtual worlds: A means-end chain approach", Comput. Hum. Behav., vol. 26(2), pp. 218–225, 2010.

Exploring a Map Survey Task's Sensitivity to Cognitive Ability

Kofi Whitney, Georgi Batinov, Les Miller, Sarah Nusser
Iowa State University
Ames, Iowa USA
{kwhitney,batinov,lmiller,nusser}@iastate.edu

Kathleen T. Ashenfelter
U.S. Census Bureau
Washington, D.C., USA
kathleen.t.ashenfelter@census.gov

***Abstract*— The present work discusses an exploratory study aimed at understanding how users' cognitive abilities influence performance and method during a series of address verification tasks. College students were given a paper map and asked to verify seven residential addresses scattered throughout a neighborhood. This approach, as opposed to using a mobile device as the verification medium, allotted participants more freedom with respect to address verification style and map interaction. The study methodology and results are discussed. The key contribution of the work described in the paper has been the identification of map usage behaviors that are sensitive to visualization and perspective taking.**

Keywords-human-computer; interaction, individual differences; location-based; usability component.

I. Introduction

Individual differences research in computing focuses on the physical, cognitive, psychological, social, and cultural distinctions of users in various settings. Individual differences have been shown to influence behavior and performance in desktop computing scenarios [4,6,10,11,12,27]. Similar findings have been presented for field studies using mobile devices, e.g., Nusser and Murphy [25] and Nusser [26] suggest that a user's spatial-visualization ability is especially pertinent to task performance.

A pilot study was conducted to look at the role that cognitive ability plays on paper map usage with respect to a series of address verification tasks [34]. These tasks were modeled around those typically performed by U.S. Census Bureau employees. The results of the pilot study implied that task performance was sufficiently related to participants' spatial-visualization and perspective-taking ability. Some interesting participant behaviors were also observed that seemed applicable to location-based software design. The favorable outcome of the pilot encouraged us to refine the methodology in terms of sample size, protocol, and setting. The study discussed in this paper is a culmination of these refinements.

The goals of this study were: (1) to demonstrate that user performance during address verification is sensitive to cognitive abilities for which participants can be tested; (2) to collect data on address verification and map usage behaviors—some of which might be tied to these cognitive abilities; and (3) to identify behaviors that might be incorporated into the design of a mobile, map-based prototype—to be evaluated in a subsequent study. These findings could be relevant to a variety of applications where users must orient themselves with respect to their geographic position in order to complete a task.

II. Background

Several investigators have examined strategies to improve the usability of map-based software. Cox [9] looked at very low-level user actions to identify strategies when working with Geographic Information Systems (GIS). Malczewski and Rinner [24] evaluated decision making based on GIS usage. Haklay and Zafiri [15] utilized GIS usage snapshots. In a recent study on a web-GIS system, Ingensand and Golay [18] found several different strategies and pointed out "users performed differently depending on their strategy". Fern et al. [14] used data mining techniques to extract strategies from data logged during software usage.

Individual differences have been recognized as an important aspect of human performance. Benyon et al. [4] note that individual differences help explain the variation in strategies among computer operators. Spatial ability has been found to be the most important of the individual differences with regard to predicting computer performance [10,11,12,33]. Spatial ability is seen as being composed of five subcomponents by Carroll [7] and Lohman [22]: visualization, speeded rotation, closure speed, closure flexibility, and perceptual speed. Visualization is the most often cited spatial ability related to computer performance [6,29,30,31,32,33,35]. Pak et al. [27] note that the importance of spatial ability depends on the task difficulty while Ackerman [1] suggests that task type is more critical.

Carroll [8] suggested that spatial visualization involves manipulation of spatial configuration in visual short-term memory. Baddeley [2,3] has modeled working memory to include verbal (phonological loop) and visual (sketchpad) components. Luck [23] shows that visual memory is limited and that performance drops systematically when individuals have more than three or four items to remember. Spatial orientation has been distinguished from spatial visualization [20] and shown to influence the way that the user visualizes self within the geographic space defined by a map. Klatzky [19] discusses the distinctions between allocentric and egocentric spatial representations. Burgess [5], Lafon et al. [21], andIgloi et al. [17] suggest that these two spatial representations are acquired and exist in parallel, a position

 ISBN: 978-1-61208-117-5

recently corroborated by electroencephalographic evidence in Plank et al. [28].

III. Methodology

A. *Screening*

Three cognitive tests were administered during the screening phase: spatial visualization (VZ-2) [13], visual memory (MV-2) [13], and perspective-taking (PT) [20]. Twenty-seven college students were selected to participate from a tested pool of over 100. The intent was to create a sample consisting of students with high combined scores and students with low combined scores. We expected this partitioning to allow us to observe greater differentiation among participants.

B. *Materials*

Topologically Integrated Geographic Encoding and Referencing System (TIGER\Line) Shapefiles were combined using ESRI's ArcGIS Desktop© to create the map. This map contained two layers of information: (1) a street layer that provided all of the streets and their respective labels and (2) an address layer that depicted each residential address as a small dot accompanied by an address number. The decision to provide such sparse map to the participants was motivated by two issues. First, our goal was to use maps similar to those used by Census Bureau staff. Second, the bare and abstract presentation of the map encouraged participants to enrich the map with detail that supported their actions. Address numbering and placement errors were deliberately added to the map to make the verification tasks more challenging.

Each participant was given a clipboard with the paper map attached to the front side in landscape orientation and a randomized list of the seven target addresses attached to the back. A multi-colored pen was provided so that participants could edit the map and list with the appropriate level of detail. Participants were outfitted with an audio device that was used throughout the exercise to record think-aloud comments and to capture responses to an exit questionnaire. The questionnaire probed participants on the effects of setting, map design, planning, task difficulty, and previous knowledge/experience. Observers used coding sheets throughout the exercise to capture supplementary data. The map, list, audio recording, and coding sheets were collected for later analysis. Additionally, these materials were used to reconstruct participant routes. These routes were input into Google® Earth so that the individual travel distances could be estimated.

C. *Residential Area: Grid vs. Non-Grid*

The field exercise took place in the cross section of the residential area depicted in *Figure 1*. The western half is designated the *grid section*. It is made up of streets that are homogeneous and closely aligned with the cardinal directions. The three-way intersections of this area are exclusively comprised of T-junctions; the 4-way intersections are similarly perpendicular. These features give this area an orthogonal, uniform structure. The resulting blocks are approximately rectangular in appearance.

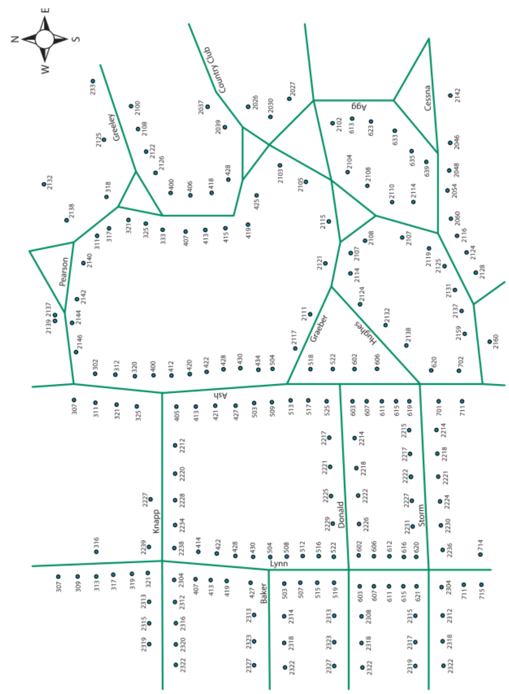


Figure 1. Field exercise map given to participants (right-side up).

The eastern half of the area contains non-uniform streets that seldom run parallel with the cardinal directions. The three-way intersections formed by these streets are Y-junctions, rather than T-junctions. The four-way intersections occur at varying angles. The large triangular medians are another notable feature—they are formed when three Y-junctions interconnect. This area is not orthogonal in nature and is designated the *non-grid section*.

It is thought that these contrasting halves would add variation to the verification scenarios encountered by participants.

D. *Field Exercise*

The participants selected for the field exercise were taken, one at a time, to the residential neighborhood. Each session contained one participant and one observer. The exercise began at a uniform starting location, where the observer explained the task flow, the think-aloud protocol, and the possible outcomes of verification. Each participant was instructed to verbalize his or her cognitive processes and thoughts related to the exercise. The participant was then asked to complete three training scenarios on a simplified map containing only two streets. At the end of the training session, the participant was given an answer key and received feedback from the observer on verbalization. The participant and observer then returned to the starting location, where the participant received the exercise map and the list containing seven addresses to be verified.

The participant was told that each of the seven target addresses required one of these basic address-verification actions: (1) *add-to-map*, (2) *move-on-map*, (3) *delete-from-map*, and (4) *confirm-on-map*. The grid and non-grid sections of the map each contained three addresses that required address-verification actions *1*, *2*, and *3* (no replace); the major street that divided these sections contained a single address requiring address-verification action *4*. Participants were told to edit the map at their discretion; the only requirement was that they clearly convey the address-verification actions chosen. After answering any final questions, the observer would no longer communicate with the participant, other than to encourage a person who had fallen silent or to request more detail with regard to a participant's response.

E. Method of Analysis and Variables Used

The performance variables, *total time*, *distance traveled*, and the number of *errors made*, were tested against the cognitive test scores using pair-wise correlations. Participants' behaviors were categorized via the analysis of qualitative data found in the observer coding sheets, the coded think-aloud transcripts, participant annotations on the provided maps and target address lists, and participants' responses to the field exercise questionnaire. Behavioral variables (excluding questionnaire data) that could be quantified across participants were tested against the cognitive test scores and the performance variables. Pair-wise correlations and a two-tailed Welch's *t* test were used when appropriate. The available records did not allow for coding of some measures, so the number of observations per variable can be fewer than 26. The significant coded variables are described below.

1) Variables Found in Transcriptions.

- *Address error pre-detection* – The number of times a participant recognized address errors at the start of the exercise; they had not yet seen the actual address.
- *Nearest address selection* – The number of times a participant chose their next target address based on its proximity to the one previously verified.
- *Cardinal heading usage* – The number of times a participant described their heading in terms of cardinal direction (north-south-east-west). This was interpreted as a proxy for an allocentric frame of reference.

2) Variables Found in Maps/Lists.

- *Target streets highlighted on map* – True if the participant highlighted the street labels associated with the target addresses.
- *Map verification annotations* – True if the participant annotated target addresses on the map in excess of what was required to indicate a solution.
- *List verification annotations* – True if the participant annotated target addresses on the list in excess of what was required to indicate a solution.
- *Route sequence on list* – True if the participant enumerated each target address to indicate the route sequence.

IV. Results

A. Correlation Among Cognitive Test Scores

Table I contains the pair-wise correlation coefficients of the 26 participants' three cognitive test scores, along with the p-value of a test determining whether the true correlation is zero. The results offer moderately strong evidence of a correlation between spatial visualization and the two other cognitive tests. There is suggestive evidence that visual memory and perspective-taking scores may also correlate.

TABLE I. Cognitive Test Score Correlation.

Cognitive Test	Cognitive Test	*n*	*r*	*p*
Spatial Visualization	Visual Memory	26	0.54	0.00
Spatial Visualization	Perspective-taking	26	0.44	0.02
Perspective-taking	Visual Memory	26	0.36	0.07

B. Correlation of Cognitive Test Scores with Performance Variables

Participants took from 30 to 66 minutes to complete the field exercise and traveled between 1.77 and 3.02 km (1.10 and 1.88 mi). Total time was negatively correlated with scores on spatial visualization ($r = -0.44$, $p = 0.02$) and perspective-taking ($r = -0.51$, $p = 0.01$). This indicates that participants with higher spatial visualization or perspective-taking ability tended to finish the exercise faster. Distance traveled was negatively correlated with spatial visualization test scores ($n = 21$, $r = -0.65$, $p = 0.00$). Visual memory test scores showed no significant correlation to the performance variables.

Clearly, there are other participant characteristics that could influence time, e.g., walking speed. We didn't investigate the connection between physical characteristics and cognition scores. However, the fact that the total distance traveled by the participants was negatively correlated with spatial visualization leads us to believe that the relationship between time and cognitive abilities is more likely influenced by the fact that lower ability participants walked farther than by any physical characteristics.

C. Cognitive Test Scores & Coded Variables

Spatial visualization test scores are positively correlated with *address error pre-detection* ($n = 21$, $r = 0.44$, $p = 0.05$) and *nearest address selection* ($n = 21$, $r = 0.45$, $p = 0.04$). This indicates that participants with high spatial visualization ability (1) identified more map errors at the onset of the exercise and (2) consistently chose their next target address based on its proximity to the one previously verified. Perspective-taking test scores are positively correlated with *address error pre-detection* ($n = 25$, $r = 0.49$, $p = 0.01$) and *cardinal heading usage* ($n = 23$, $r = 0.51$, $p = 0.01$). This suggests that participants with higher perspective-taking ability (1) identified more map errors at the onset of the exercise and (2) were more likely to describe their heading from a cardinal, allocentric frame of reference (north-south-east-west) rather than an egocentric one (forward-backward-right-left).

A two-tailed Welch's *t* test was used to test for associations between the cognitive test scores and the variables found in the collected maps and lists (Table II). Spatial visualization test scores were negatively associated with *map verification annotations*, suggesting that participants with lower spatial visualization ability added more supporting detail to the map. Participants with lower perspective-taking ability exhibited similar behavior with regard to the target address list—as shown by the negative association between perspective-taking test scores and *list verification annotations*. Additionally, perspective-taking test scores showed negative association with *target streets highlighted on map*, indicating that participants with lower perspective-taking ability tended to identify the streets that the target addresses were on; they then highlighted the street labels on the map. Visual memory test scores showed a positive association with *route sequence on list*, implying that participants with higher visual memory tended to enumerate a sequential route order on their target list of addresses.

TABLE II. ASSOCIATION OF COGNITIVE TEST SCORES WITH MAP AND LISTS VARIABLES (WELCH'S *T* TEST)

Behavior	Cognitive Test	$Y_1 - Y_0$ *	$SE(Y_1-Y_0)$*	*p*
Target streets highlighted on map	Spatial Visualization	-4.25	1.54	0.01
Map verification annotations	Perspective-taking	-4.45	1.52	0.02
List verification annotations	Perspective-taking	-4.31	1.90	0.05
Route sequence on list	Visual Memory	3.29	1.11	0.01

* Y_1 is the mean of cognitive test scores for all who exhibited the behavior and Y_0 is the mean of cognitive test scores for all who did not exhibit the behavior.

D. Additional Behaviors Observed

Behaviors that were sporadic and difficult to capture could not be adequately linked to cognitive test scores, however, they may be worthy of consideration in the follow-up study.

- Participants preferred either a north-up or track-up map orientation and some north-up users seemed to temporarily switch orientations in confusing areas of the neighborhood.
- Participants covered addresses and map elements that were unrelated to their current target address.
- Participants used color on the maps and lists to indicate and differentiate their various actions and decisions.
- Participants inferred detailed relationships between cardinal directions and street numbering patterns, e.g., "even-numbered addresses are on the north on east-west streets and on the west on north-south streets".

V. DISCUSSION

A. Relation of Participant Performance to Cognitive Test Scores

Our findings support the hypothesis that cognitive test scores are related to participant performance on a map survey task. Spatial visualization scores were a strong prior indicator of performance, being significant against distance traveled, total time and *nearest address selection*. Perspective-taking ability was also correlated with total time, *nearest address selection*, and *address error pre-detection*. Conversely, we did not find visual memory to be linked to performance metrics. Errors were not statistically relevant.

Predictably, distance traveled and total time are significantly correlated; however, *nearest address selection* was not correlated with distance traveled. This could result from two or more participants traveling different distances despite their similar *nearest address selection* scores. Another explanation may be that, in some cases, a tendency to choose the next target address based on distance is inadequate when a given route will cover multiple addresses—a holistic approach should be taken in this case.

Address error pre-detection was positively correlated with perspective-taking scores. One explanation is that in order to excel on the perspective-taking test, one must assess the relative placement of a target to its surroundings. This same ability might be applied to target addresses with respect to the map, allowing these participants to hone in on discrepancies. Furthermore, this finding, when taken together with the fact that perspective-taking scores are correlated with *cardinal heading usage*, suggests that people with higher perspective-taking ability are capable map users.

These results provide evidence that the address verification task was sensitive to cognitive abilities. The literature further indicates that spatial visualization scores predict computer performance [6,29,30,31,32,33,35]. Tailoring interfaces for cognitive differences appears to be a desirable direction in map survey software design.

B. Behaviors Linked to Test Scores and Software Design

Behaviors associated with the cognitive test scores provided some evidence that address verification software can benefit from features sensitive to the respective abilities of a user.

Participants who detected target address errors on the map (missing addresses, for example) without physically examining the target location typically did so at the beginning of the field exercise. This practice might be emphasized in software through an initial planning step. Participants with higher perspective-taking ability tended to prefer describing their movement in cardinal (north-south-east-west) terms. This finding seems to suggest that a software presentation that facilitates or even emphasizes the cardinal directions would be appropriate for users with high perspective-taking ability. A compass rose, for example, helps to fill this void.

Modifications that were made to the provided map and list should also be considered. Participants that added additional address verification annotations to the map had a lower overall spatial visualization ability; thus, textual cues may have served to alleviate deficiency in this ability. A software feature that allows custom tagging of map elements may benefit these users. Analogously, participants who added extra address verification annotations on the target list

 ISBN: 978-1-61208-117-5

had a lower overall level of perspective-taking ability. This group may be assisted through embedded note-taking features or some sort of checklist. Participants who highlighted target streets had a lower overall level of perspective-taking ability than participants who did not. This group may also appreciate the ability to tag map elements via a highlighting tool. Finally, some participants added the order in which they verified addresses to their target list. They had a higher overall level of visual memory than the rest of the sample. Their bookkeeping can be automated in software, or alternatively, they may benefit from a planning tool that links map areas to a sequence of target addresses.

C. Observed Behaviors and Software Design

The behavioral variables found on the maps and lists and cardinal direction were not statistically related to performance metrics but suggested enhancements to the software design space. Some participants would use a pen or their hands to obscure addresses that they had verified. This preference might be accommodated by presenting dissimilar levels of detail for different areas of the map; one example would be a "fish-eye" map viewport. Pan and zoom functions could be extended to allow for more freedom, also addressing user dispositions with regard to map detail. Additionally, pan and zoom "bookmarks" could enable the retention of serendipitous map views. Color-coding can be employed not only to differentiate among survey actions (as used by participants), but also to highlight the odd or even-numbered sides of streets and to convey relationships between these streets and the cardinal directions.

VI. CONCLUSTION

Our study presents evidence that an address verification task, driven by a paper map, is sensitive to the cognitive abilities of the verifier—especially their spatial visualization and perspective-taking abilities. Performance and some behaviors were significantly associated with psychometric test scores, thereby improving the plausibility of a software design that incorporates enhancements sensitive to users' cognitive abilities. The behaviors that were observed also suggest a number of software design considerations. While our tests were specific to address verification, we believe that several of the lessons learned in this study will be applicable to other areas of map-based surveys. In particular we have seen that spatial ability played a role in initial planning (pre-planning), how the participants used the map during way finding and how they used the map in the area around the target address. Most map-based applications will involve one or more of these activities.

Our future work in this area will include the development of a software interface that incorporates the enhancements previously discussed. This software will be the focus of a study that will evaluate the efficacy of these enhancements with respect to the cognitive abilities of the participants. We will also use the data collected during this study to develop decision models as a means of providing a clearer picture of how spatial abilities impact the participant's work.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their comments that led to an improved manuscript. The work was supported in part by a grant from the National Science Foundation (0822002).

REFERENCES

[1] P. Ackerman, "Ability Determinants of Individual Differences in Skilled Performance," In R. J. Sternberg and J.E. Pretz (eds). Cognition and Intelligence: Identifying the Mechanisms of the Mind. Cambridge University Press. Cambridge, UK. 2005, pp. 142-159.

[2] A. D. Baddeley, "Short-term and working memory," In E. Tulving & F. I. M. Craik (Eds.), The Oxford handbook of memory, Oxford, England: Oxford University Press, 2000, pp. 77–92.

[3] A. D. Baddeley Working Memory, Thought, and Action. Oxford University Press. Oxford, England, 2007.

[4] D. Benyon, A. Crerar, and S. Wilkinson, "Individual differences and inclusive design," In: Stephanidis C. (ed.) User Interfaces for All: Concepts, Methods and Tools, LEA, Mahwah, 2001, pp. 21-46.

[5] N. Burgess, "Spatial memory: how egocentric and allocentric combine," Trends in Neurosciences, Vol. 10, No. 12, 2006, pp. 551-7.

[6] F. R. Campagnoni and K. Ehrlich, "Information retrieval using a hypertext-based help system," Transactions on Information Systems, vol. 7, 1989, pp. 271–291.

[7] J. B. Carroll, Human cognitive abilities: A survey of factor-analytic studies, New York: Cambridge University Press. 1993.

[8] J. B. Carroll, "Psychometric tests as cognitive tasks: a new structure of intellect," Princeton, N.J.: Educational Testing Service, Research Bulletin, 1974, 74-16.

[9] Cox, A. "Multimodal GIS Interfaces in Crisis Management –Usability Study," http://www.geovista.psu.edu/publications/2005/ DAVE, 2005, Accessed June 17, 2010.

[10] N. Dahlbäck, K. Höök, and M. Sjölinder, "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 1996.

[11] D. Egan, "Individual differences in human-computer interaction," In: M. Helander (ed.), Handbook of Human-Computer Interaction, Amsterdam: Elsevier Science Publishers, 1988, pp. 543-568.

[12] D. E. Egan and L. M. Gomez, "Assaying, isolating and accommodating individual differences in learning a complex skill," In R. Dillon (Ed.), Individual differences in cognition, New York: Academic Press, 1985, pp. 173-217.

[13] R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, "Manual for kit of factor-referenced cognitive tests," Princeton, NJ: Educational Testing Service, 1976.

[14] X. C. Fern, C. Komireddy, V. Grigoreanu, & M. Burnett, "Mining problem-solving strategies from HCI data," ACM Transactions on Computer-Human Interaction, Vol. 17, No. 1, 2010, DOI10.1145/1721831.1721834.

[15] M. Haklay and A. Zafiri, "Usability engineering for GIS: learning from a snapshot," http://eprints.ucl.ac.uk/16799/1/16799.pdf, 2007, Accessed June 17, 2010.

[16] C. Gibson and S. M. Nusser, 2006 Census Test Evaluation 3: Assessing the Role of Spatial Abilities in Address Canvassing Activities, Census Bureau Report, September 1, 2006.

[17] K. Igloi, M. Zaoui, A. Berthoz, and Laure Rondi-Reig, "Sequential egocentric strategy is acquired as early as allocentric strategy: parallel acquisition of these two navigation strategies," Hippocampus 19, 2009, pp. 1199–1211.

[18] J. Ingensand and F. Golay, "Task-oriented usability evaluation of a web-GIS for a real world community," http:// www.urisa.org/files/Ingensand-Golay-Usability_evaluation.doc, 2009, Accessed June 17, 2010.

[19] R. L. Klatzky, " Allocentric and egocentric spatial representations: definitions, distinctions, and interconnections," In C. Freksa, C. Habel, & K. F. Wender (Eds.), Spatial cognition - An interdisciplinary approach to representation and processing of spatial knowledge (Lecture Notes in Artificial Intelligence 1404), 1998, pp. 1-17.

[20] M. Kozhevnikov, M. A. Motes, B. Rasch, and O. Blajenkova, "Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance," Applied Cognitive Psychology, vol. 20, 2006, 397-417.

[21] M. Lafon, M. Vidal and A. Berthoz, " Selective influence of prior allocentric knowledge on the kinesthetic learning of a path," Experimental Brain Research, 194, 2009, pp. 541–552.

[22] D. F. Lohman, "Spatial ability and g," In I. Dennis & P. Tapsfield (Eds.), Human Abilities: Their Nature and Measurement, Mahwah, NJ: Lawrence Erlbaum Associates, 1996, pp. 97-116.

[23] S. J. Luck, "Visual short-term memory," In S.J. Luck and A. Hollingworth (eds). Visual Memory. Oxford University Press. Oxford, UK, 2008, pp. 43-85.

[24] J. Malczewski and C. Rinner, "Exploring multicriteria decision strategies in GIS with linguistic quantifiers: a case study of residential quality evaluation," Journal of Geographical Systems, Vol. 7, no, 2, 2005, pp. 249-268.

[25] E. D. Murphy and S. M. Nusser, "Evaluating user interfaces for accommodation of individual differences in spatial abilities and way-finding strategies," In Proceedings of the UAHCI 2003: Second International Conference on Universal Access In Human-Computer Interaction, vol. 4, 2003, pp. 1005-1009.

[26] S. M. Nusser, "Digital capture of geographic feature data for surveys," Proceedings of the 2005 Federal Committee on Statistical Methodology Research Conference. Available: http://www.fcsm.gov/05papers/Nusser_IXA.pdf., 2005.

[27] R. Pak, W. A. Rogers and A. D. Fisk, "Spatial ability subfactors and their influences on a computer-based information search task," Human Factors, vol. 48, 2006, pp. 154-165.

[28] M. Plank, H. J. Müller, J. Onton, S. Makeig, and K. Gramann, "Human EEG correlates of spatial navigation within egocentric and allocentric reference frames," In Hölscher et al. (Eds.), Spatial Cognition VII, Lecture Notes in Computer Science, 6222, 2010, pp. 191-206.

[29] F. Seagull and N. Walker, "The effects of hierarchy structure and visualization ability on computerized information retrieval," International Journal of Human Computer Interaction, vol. 4, 1992, pp. 369–385.

[30] M. Sein, L. Olfmann, R. Bostrom and S. Davis, "Visualization ability as a predictor of user learning success," International Journal of Man – Machine Studies, vol. 39, 1993, pp. 599–620.

[31] K. M. Stanney and G. Salvendy, "Effects of diversity in cognitive restructuring skills on human–computer performance," Ergonomics, vol. 37, 1994, pp. 595–609.

[32] K. M. Stanney and G. Salvendy, "Information visualization: Assisting low spatial individuals with information access tasks through the use of visual mediators," Ergonomics, vol. 38, 1995, pp. 1184–1198.

[33] K. Vicente, B. C. Hayes and R. C. Williges, "Assaying and isolating individual differences in searching a hierarchical file system," Human Factors, vol. 29, 1987, 349-359.

[34] K. W. Whitney, G. I. Batinov, S.M. Nusser, L. L. Miller, and K. T. Aschenfelter, (In press) Understanding the User's Approach to Location-based Field Tasks. To be presented at the 23rd International Conference On Computer Applications in Industry and Engineering. 2010.

[35] H. Zhang and G. Salvendy, "The implications of visualization ability and structure preview design for web information search tasks," International Journal of Human-Computer Interaction, vol. 13, no. 1, 2001, pp. 75–79.

A Graphical Interface for User Authentication on Mobile Phones

Mohammad Sarosh Umar
Department of Computer Engineering,
Aligarh Muslim University, Aligarh,
India
saroshumar@zhcet.ac.in

Mohammad Qasim Rafiq
Department of Computer Engineering,
Aligarh Muslim University, Aligarh,
India
mqrafiq@hotmail.com

***Abstract* — Recently, mobile phones have become an important tool to carry out financial transactions besides the normal communication. They are increasingly being used to make payments, access bank accounts and facilitate other commercial transactions. In view of their increased importance there is a compelling need to establish ways to authenticate people on the mobile phones. The current method for authentication uses alphanumeric username and password. The textual password scheme is convenient but suffers from various drawbacks. Alphanumeric passwords are most of the times easy to guess, vulnerable to brute force attacks and are easily forgotten. With financial transactions at stake, the need of the hour is a collection of robust schemes for authentication. Graphical passwords are one of such schemes which offer a plethora of options and combinations. We are proposing a scheme which is simple, secure and robust. The proposed graphical password scheme will provide a large password space and at the same time will facilitate memorability. It is suitable to implement on all touch sensitive mobile phones.**

Keywords- User authentication; graphical password; mobile phone security; usability.

I. INTRODUCTION

Mobile phones have made their presence felt in different walks of human life. Today's technically advanced mobile phones are capable of not only receiving and making phone calls, but can very conveniently store data, take pictures and connect to the internet. They have also become a powerful tool to conduct commercial and financial transactions. They are increasingly being used to make payments, such as at retail shops, public transport, paid parking areas and also to access the bank accounts via internet. In view of this the security and safety of mobile phones have become paramount to prevent unauthorized persons from conducting any unwarranted transactions through the phones. Conventional method of authentication remains mainly text based as it has been around for several decades and also because of ease of implementation. However, text based passwords suffer from various drawbacks such as they are easy to crack through dictionary attacks, brute force, shoulder surfing, social engineering etc. The “small dictionary” attack is so successful that in Klein’s case study [1], about 25% of 14,000 passwords were cracked by a dictionary with only 3 million entries. Following the same method used by Van Oorschot and Thorpe [12], such a dictionary can be exhausted by a 3.2 GHz Pentium^{TM}4 machine in only 0.22 second. Graphical passwords, which require a user to remember and repeat visual information, have been proposed to offer better resistance to dictionary attack. Psychological studies support the hypothesis that humans have a better capability to recognize and to recall visual images than alphanumeric strings [3], [4] and [5]. If users are able to remember more complex graphical passwords (i.e., from a larger password space), an attacker has to build a bigger dictionary, thus spend more time or deploy more computational power to achieve the same success as for textual passwords. In this paper, we will demonstrate a graphical grid-based password scheme which will aim at providing a huge password space along with ease of use. We will also analyze its strength by examining the success of brute force technique. In this scheme we will try to make it easy for the user to remember and more complex for the attacker.

II. RELATED WORK

Many papers have been published in recent years with a vision to have a graphical technique for user authentication. Primarily there are two methods, having recall and recognition-based approach respectively. Traditionally both the methods have been realized through the textual password space, which makes it easy to implement and at the same time easy to crack.

The study shows that there are 90% recognition rates for few seconds for 2560 pictures [2]. Clearly the human mind is best suited to respond to a visual image. A recall-based password approach is VisKey [6], which is designed for PDAs. In this scheme, users have to tap spot in sequence to make a password. As PDAs have a smaller screen, it is difficult to point exact location of spot. Theoretically, it provides a large password space but not enough to face a brute force attack if number of spots is less than seven [7].

Figure 1: VisKey SFR

A scheme like *Passfaces* in which user chooses the different relevant pictures that describes a story [8] is an image recognition-based password scheme. Recent study of graphical password [9], says that people are more comfortable with graphical password which is easier to remember. In Recall-based password, user has to remember the password.

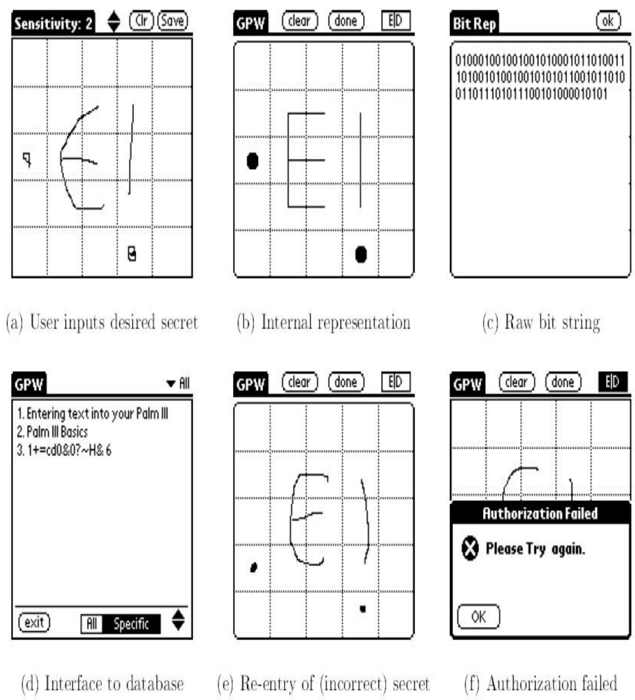


Figure 2: DAS scheme.

Jermyn et al. [10], proposed a technique, called "Draw-a-secret (DAS)", which allows the user to draw their unique password (Figure 2). In the DAS scheme, stylus strokes of the user-defined drawing are recorded and the users have to draw the same to authenticate themselves. DAS scheme also allows for the dots as well as shown in one of the examples in Figure 2.

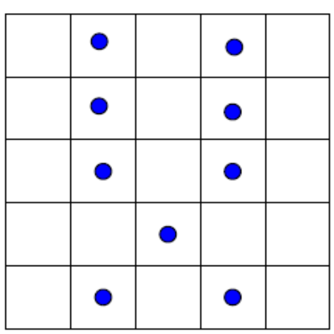


Figure 3: Example of a password in DAS has only dots.

But research shows that people optimally recall only 6 to 8 points in pattern [11], and also successful number of recalls decreases drastically after 3 or 4 dots [12]. Our main motivation will be to increase password space. The user can choose the geometrical shape of their choice for the device like PDA having graphical user interface that will also optimize that password storage space. In our scheme, we will allow users to draw some geometrical shape with some fixed end points and by putting dots at different location but it will give some filed triangle in such a way that chances of remembering those positions will be better.

III. DRAWING GEOMETRY

Drawing geometry is a graphical password scheme in which the user draws some geometrical object on the screen. Through this scheme we are targeting devices like mobile phones, notebook computers and hand-held devices such as Personal Digital Assistants (PDAs) which have graphical user interface. Since these devices are graphical input enabled we can draw some interesting geometries using stylus.

In this scheme there will be *mxn* grids and each grid is further divide into four parts by diagonal lines as shown in Figure 4. We have considered 4x5 grid keeping in mind the typical screen size of the PDAs these days and its width height ratio. Depending on the screen size it can be changed with justifiable number of rows and columns.

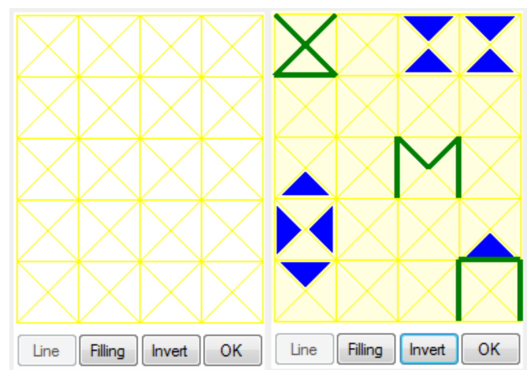


Figure 4: Grid provided to user and some simple geometrical shape drawn by user.

On taking the size (4x5) we have total of 5x4 = 20 blocks and each block has four triangles so total number of possible triangles is: (20 blocks) x (4 triangle/block) = 80 triangles. Similarly each block has 4 small diagonal lines so total lines in that way (20 blocks) x (4 lines/block) = 80 lines. Also we do have some lines which are a result of joining adjacent points horizontally and vertically. That will give 4x6=24 (horizontal) and 5x5=25 vertical lines which makes a total of 24+25 = 49 (horizontal and vertical) lines. In that way we will have total of

$$p\,(5,4) \Rightarrow 80 + 80 + 49 = 209 \qquad (1)$$

These 209 objects can be used to choose password by drawing some of these objects in an easy and efficient manner. A password is considered to be the selection of certain lines and triangles. When a triangle is selected it is filled with some color and when a line is selected the color of that line changes (gets highlighted). Any combination of the selection of lines and triangles will form a password as shown in Figure 5. In this way, highlighted lines and filled triangle will provide us larger password space. Filling

 ISBN: 978-1-61208-117-5

triangle and highlighting work can be done by using stylus of PDAs either by putting dot in triangle or by dragging the stylus crossing that line. As research shows that if the number of dots increases to difficult to remember those it is also increases. In this scheme we fill the triangle highlighted lines makes geometric shape which is to be recalled not the dots. More over we give another option which converts all highlighted lines to un-highlighted and vice-versa and the same for filling triangle by single click a button "Invert" a button which at least double the password space within practical limit of password length. A line which is not inclined at an angle of 45° or 0° or 90° i.e. the line which is not parallel to diagonal, horizontal as well as vertical lines. (Let's call them *non-parallel* lines) These non-parallel lines can also be drawn by joining two points after enabling those drawing by clicking the button given labeled line "Line" which enables user to draw non-parallel lines. As we can see that crossing the same lines again cancels the effect of highlighting, Figure 6, in general we can say that crossing even number of times the same line will cancel the highlighting effect. The users don't need to recall the strokes but the resulting geometry. By using inversion operation as shown in Figure 7 the user can deselect all currently highlighted lines and triangles and select all the unselected lines and triangles.

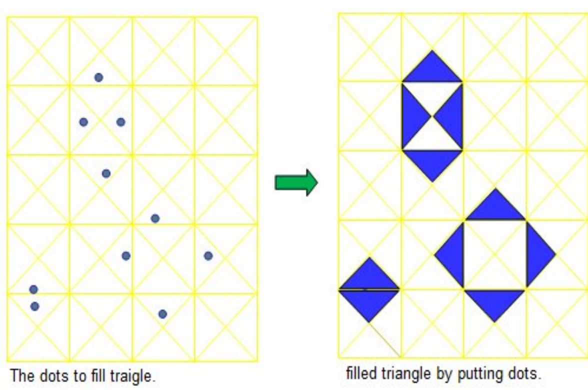


Figure 5: Drawing solid triangle

Note that the inversion does not take place for non-parallel lines. Figure 8 shows a password made by using parallel and non-parallel lines. To draw that we have button stylus able to draw those lines by dragging stylus from one point to another. The start point and end point of such line will be decided by actually where stylus touches the screen and where it leaves it. As illustrated in Figure 8 if stylus touches the screen at any location say coordinate *(x,y)* where two vertical line *va* and *vb* (nearest vertical lines from point P at a distance half cell width) such that *va* _ *x* < *vb* and *ha*_ *y* < *hb* the nearest point of region P will be considered. Same strategy will be adopted for end point where stylus release screen. If lines drawn by user are parallel but procedure adopted by user to draw is as of nonparallel, in that case the scheme will automatically detect that and even if parallel lines are drawn by non-parallel method of drawing it will be considered as parallel lines.

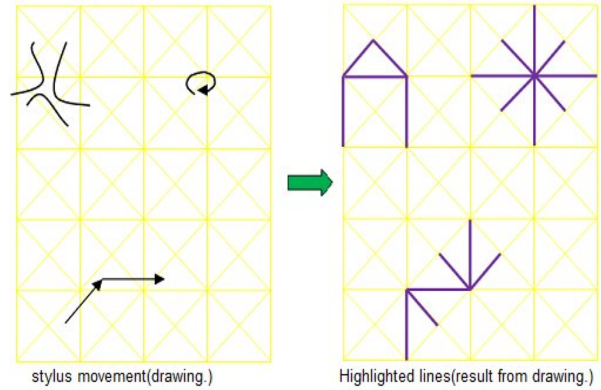


Figure 6: Drawing lines

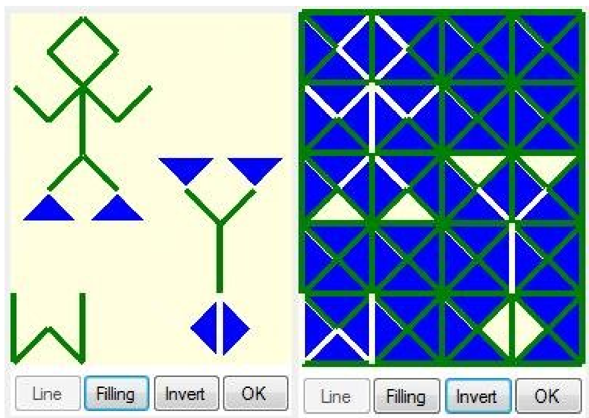


Figure 7: Inversion of drawn geometry

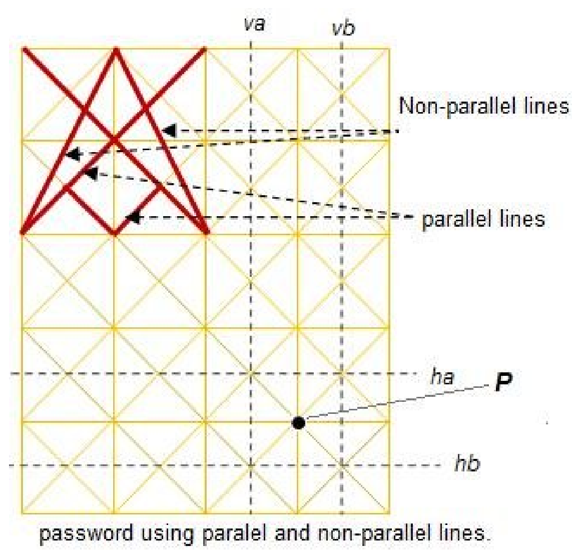


Figure 8: Example of non-parallel lines

The grid shown on screen is for the user's convenience. Password drawn on invisible grids is shown in Figure 7 also illustrates the inversion.

IV. TEXT SIMULATION

The above mentioned technique can also be used to write any textual password in graphical manner. In the example shown the word "IMAGINE" is written vertically to accommodate more letters on the screen, still letter E is missing (purposely) as shown in Figure 9. If the password contains more words then multiple screens (say frames) can be used to accommodate them.

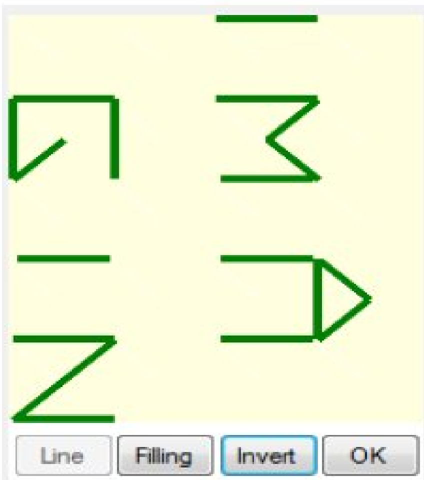


Figure 9: Example of textual password

This can allow users to use textual passwords in graphical way. This letter(s) can be drawn in any direction and at any letter can be entered at any position on screen as per the user's convenience.

V. EXTENSION FOR POSITION INDEPENDENCE AND MULTISTAGE

As of now we have considered that the shapes as well as its location constitute the password, together. If the user has written letter 'A' but fails to recall the position of the 'A' even then the password will be incorrect. This scheme can be extended to accommodate such cases. The location of the figure can be ignored if the shape is correct (as illustrated in Figure 10). The same shape pattern at two different location circled should be treated as same. Obviously doing so the password space decreases but by increasing number of grid this can be compensated. As we have seen that text can be drawn but size of the PDAs limits the grid size. We can have multiple stages for drawing shapes i.e. one shape in first frame followed by next frame and so on.

The user can select the *more* button provided (not shown any where) to go into next fresh blank frame on which more letters or shape can be drawn. As we could not write full word IMAGINE but by doing so (multistage) we can write first few letters say IMA in first frame and rest GINE in second frame. Multistage increases the time required to enter the password but also it gives us huge password space like my password word GRAPH is simulated in geometry the way it can be entered or chosen by user increased like GR and APH or GRA and PH etc for two stage, though stages will be less normally but by not fixing the number of stage we get advantage of high password space.

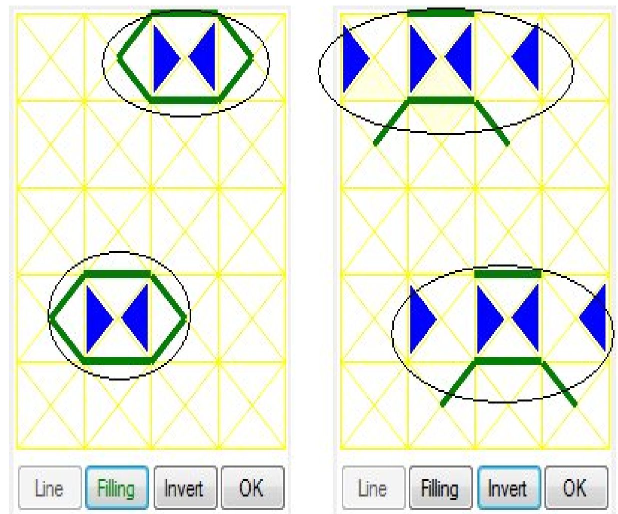


Figure 10: Example of position independence

VI. STORAGE OF PASSWORD

Since there is no need to store any image therefore only password need to be stored as we have seen in case of grid size (4x5) there are 209 possible objects if non-parallel lines are not considered, if we numbered every object from number 0, 1, 2,… , 208 then 209 bits are sufficient to store such password. An extra bit should be kept for inversion whether the password is inverted or not to avoid more calculation while entering the password. For including non-parallel lines, each non-parallel line can be stored by storing the coordinates of two points (start point and end point). The first fix number of bits will represent how many such lines are there and then the coordinates of end points of each line (10 bits for each). So if number of non-parallel lines is *np* then total password length by taking 10 bits for representing each non-parallel line is given in Figure 11.

Required bits to store password;

$$= 209 + 1 + 10 + np * 10;$$
$$= 220 + np * 10.$$

So this scheme does not take much space to store the password as many graphical schemes take [10].

VII. SECURITY ANALYSIS

As we have seen in eqn.(1) that we have 209 objects each can be either highlighted or unselected, individually. Considering only the 209 objects and excluding the non-parallel lines then we have a total of $2^{209} = 8.2275X10^{62}$ possibilities which is huge in terms of password space.

So it is very robust from security point of view even after excluding non-parallel lines. If we consider non-parallel lines also the additional 220 lines will be added which will be also either highlighted or unselected so in that case total password possible $2^{(209+220)} = 2^{429} = 1.386X10^{129}$. It is clear that the password space will increase exponentially with increase in rows or columns as shown in table above. It will be possible for device with a bigger screen (like ATM) to have many more columns and rows.

Due to this larger password space it is very difficult to carry out brute force attack on this password. With this scheme even if user decides to have the graphical representation of the text, he will not be susceptible to dictionary attacks. We have computed above password space in simple case with only 4x5 grids and single stage password entry. Since we have not made any special assumption for text simulation in this scheme, the password space remains same even if we use it as textual password scheme.

VIII. CASE STUDY

A small case study was conducted and twenty five users were chosen randomly comprising undergraduate and graduate students. A short questionnaire was developed and the users were requested to try this scheme and share their experience with us. Interestingly, the users rated it 8.5 on a scale of 10.0 on ease of use which demonstrates the acceptable usability of the scheme. 80% of the users (5) found it easier to remember passwords in the form of graphical figures. Twenty one users out of twenty five could reproduce their passwords after an interval of one week. A rigorous study may be conducted to further explore the proposed scheme.

IX. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a graphical password scheme in which the user can draw simple geometrical shapes consisting of lines and solid triangles. The user need not remember the way in which the password is drawn but only the final geometrical shape.

This scheme gives large password space and is competent in resisting brute force attack. Moreover, the way of storing the password requires less memory space as compared to the space required by other existing graphical authentication schemes.

This scheme is less susceptible to shoulder surfing as the screen of the hand held device is visible to the user only. However, when employed on PCs and ATM machines it is susceptible to shoulder surfing. To make it more robust and handle the problem of shoulder surfing, the geometrical shape will have to be drawn by assigning an order to the various components i.e. triangles and lines. This consideration will limit the scheme's vulnerability to shoulder surfing and will further expand the password space.

REFERENCES

[1] D. Klein, "Foiling the Cracker", A Survey of, and Improvements to, Password Security in The *2nd USENIX Security Workshop*, pp. 5-14, 1990.

[2] L Standing, J. Conezio, and R. N. Haber, "Perception and Memory for Pictures: Single-trial Learning of 2500 Visual Stimuli". *Psychonomic Science*, 19(2): 73-74, 1970.

[3] A. Paivio, T. B. Rogers, and P. C. Smythe, "Why are pictures easier to recall than words?", Psychonomic Science, 11: 137-138, 1968.

[4] G. H. Bower, M. B. Karlin, and A. Dueck, "Comprehension and memory for pictures". Memory and Cognition, 3, 216-220, 1975.

[5] L. Standing, "Learning 10,000 Pictures". Quarterly Journal of Experimental Psychology, 25: 207-222, 1973.

[6] SFR-IT-Engineering, http://www.sfr-software.de/cms/EN/pocketpc/viskey/, Accessed on January 2011.

[7] M. D. Hafiz, A. H. Abdullah, N. Ithnin, and H. K. Mammi, "Towards Identifying Usability and Security Features of Graphical Password in Knowledge Based Authentication Technique", in Proceedings of the Second Asia International Conference on Modelling & Simulation, IEEE Computer Society, pp. 396-403, 2008.

[8] D. Davis, F. Monrose, and M. Reiter, "On User Choice in Graphical Password Schemes", in Proceedings of *13th USENIX Security Symposium*, pp. 151–164, 2004.

[9] J. Thorpe and P. C. van Oorschot, "Graphical Dictionaries and the Memorable Space of Graphical Passwords", in Proceedings of *13th USENIX Security Symposium*, pp. 135–150, 2004.

[10] I. Jermyn, A. Mayer, F. Monrose, M. Reiter, and A. Rubin, "The Design and Analysis of Graphical Passwords", in Proceedings of *8th USENIX Security Symposium*, pp. 1-14, 1999.

[11] R. S. French, "Identification of Dot Patterns From Memory as a Function of Complexity", *Journal of Experimental Psychology*, 47: 22–26, 1954.

[12] S. I. Ichikawa, "Measurement of Visual Memory Span by Means of the Recall of Dot-in-Matrix Patterns", *Behavior Research Methods and Instrumentation*, 14(3):309–313, 1982.

[13] X. Suo, Y. Zhu, and G. S. Owen, "Graphical Passwords: A Survey", in Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC 2005), IEEE Computer Society, pp. 463-472, 2005.

[14] K. Chalkias, A. Alexiadis, and G. Stephanides, "A Multi-Grid Graphical Password Scheme", in 6th International Conference on Artificial Intelligence and Digital Communications (AIDC 2006), Greece, pp. 80-90, 2006.

[15] J. Thorpe and P. C. van Oorschot, “Towards Secure Design Choices for Implementing Graphical Passwords”, in Proceedings of the 20th Annual Computer Security Applications Conference (ACSAC’04), IEEE Computer Society.

[16] J. Thorpe, P.C. van Oorschot, and A. Somayaji, “Pass-thoughts: Authenticating With Our Minds”, Proceedings of Workshop on New security paradigms, Lake Arrowhead, California, pp. 45 – 56, 2005

[17] P. L. Lin, L. T. Weng, and P. W. Huang, “Graphical Passwords Using Images with Random Tracks of Geometric Shapes”, in Proceedings of 2008 Congress on Image and Signal Processing, IEEE Computer Society, pp. 27-31, 2008.

[18] S. Chiasson, P. C. van Oorschot, and R. Biddle, “Graphical Password Authentication Using Cued Click Points”, ESORICS 2007, 12th European Symposium on Research in Computer Security, Dresden, Germany, September, pp. 24-26, 2007.

[19] M. W. Calkins, “Short studies in memory and association” from the Wellesley College Laboratory. Psychological Review, 5:451-462, 1898.

[20] M. A. Borges, M. A. Stepnowsky, and L. H. Holt, “Recall and Recognition of Words and Pictures by Adults and Children”. Bulletin of the Psychonomic Society, 9:113-114, 1977.

Total cells	No. of Rows(i)	No. of Columns(j)	Parallel lines and triangles. (*p(i,j)*)	No. of Non-Parallel lines. (*n(i,j)*)	Total lines and triangles.	Password space (without non-parallel lines)	Password space (including non-parallel lines)
9	3	3	96	44	140	7.922×10^{28}	1.393×10^{42}
12	4	3	127	80	207	1.701×10^{38}	2.057×10^{62}
16	4	4	168	140	308	3.741×10^{50}	5.215×10^{92}
20	5	4	209	220	429	8.227×10^{62}	1.386×10^{129}
24	6	4	250	320	570	1.809×10^{75}	3.864×10^{171}
25	5	5	260	340	600	1.852×10^{78}	4.149×10^{180}
30	6	5	311	490	801	4.712×10^{93}	1.333×10^{241}
36	6	6	372	700	1072	9.619×10^{111}	5.060×10^{322}
49	7	7	504	1288	1792	5.237×10^{151}	2.791×10^{539}

Figure 11: Variation of password space with increase in number of grids.

ISBN: 978-1-61208-117-5

PuppetAnimator: A Performative Interface for Experiencing Shadow Play

Yue Shi, Yue Suo, Shang Ma, Yuanchun Shi
Key Laboratory of Pervasive Computing
Dept. of Computer Science, Tsinghua University
Beijing, China
shi-y04@mails.tsinghua.edu.cn, suoyue@tsinghua.edu.cn, msjianghu@gmail.com, shiyc@tsinghua.edu.cn

***Abstract*— We present PuppetAnimator, a novel interface that enables non-expert users to experience funny shadow play and create puppet animation by manipulating the character with simple pointing device. Inspired by the real shadow puppet performance, we employ the "sticks from hands to the puppet" metaphor for interaction and design a set of motion modes including Translation, Swing, Drag, Rotation and Two-joint Drive. Based on motions modes' combination, several puppet templates are prebuilt for the user to customize the skin with images and control channels with input devices. Then the user can manipulate the puppet to arrange actions. Features such as recording the user's performance, replaying and saving as video files are also provided. This paper describes the system design rationale and implementation. We demonstrate PuppetAnimator with several puppet actions and pilot study shows that with the system, users can easily make animation of a shadow puppet's action for storytelling.**

Keywords-multi-point; performative interface; shadow puppet

I. INTRODUCTION

Chinese shadow play is a characteristic form of art in which shadow puppets are manipulated in front of the illuminated backdrop for storytelling and entertainment. During the performance, each puppet is animated to perform various actions by several sticks held by a puppeteer. Being an ancient drama, shadow play also generates a modern cartoon style which is very popular among people.

However, making and performing a real shadow puppet is a very difficult task which keeps non-expert users from experiencing this form of art. The process to make shadow puppet animation is also time-consuming even with the aid of current animation software. When a user wants to make a shadow puppet style animation to tell a story or import into other system as a material, he has to insert several key frames in a time line, pose the gesture of the characters in each frame with a lot of clicks and drags but not be able to see the action until the last minute. If a tool can enable the user to experience the shadow play and at the same time, quickly create shadow puppet animation in the way similar to puppeteer's, the editing process will be made easier, more straightforward and interesting.

In recognition of the above issues, we have developed PuppetAnimator (Figure 1), a novel interactive system that allows non-expert users to experience shadow play and create shadow puppet style animation by performing the actions of the puppets with point devices. With the system, users can design virtual puppets, arrange their actions and

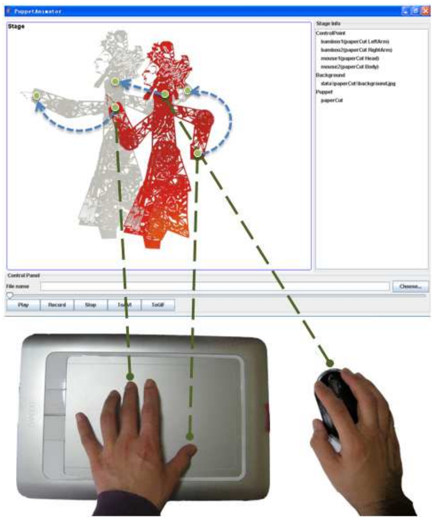


Figure 1. Up: the screenshot of PuppetAnimator. Down: the approach to interact with PuppetAnimator. Input points (green dots) from different devices can be mapped to the points of force application (green dots with white circles) on the puppet to perform an action (blue dashed lines with an arrow).

record the ideas as animation clips. The system provides a set of motion modes which are elemental actions in the real shadow play. With the combination of those modes, puppet templates can be designed that simulate the physical behavior of a real puppet in respond to users' inputs. In the performance stage, we make use of a real performance metaphor as the interaction method - holding point input devices like sticks to control the puppet. To make it more accessible to novice users, we choose simple and inexpensive point devices such as mouse and touch pad.

Our work is a kind of performance animation that provides an approach to enable junior level users to perform their own shadow play. The first contribution of our work is the interface to support this which is easy to learn and straightforward to use. The devices we use to generate input are available to everyone. Second, the set of motion modes we designed constitutes the basic tool box for the puppet manipulation. And the interaction method based on it is demonstrated with several actions such as walking, jumping and dancing. The work is aimed to take effect in informal circumstances such as improvised storytelling and animated GIF production.

II. RELATED WORK

Designing interface to manipulate a puppet and generate skeletal animation is a long standing problem that many techniques and systems are provided.

Some systems have enabled users to control a puppet. In PuppetWall [4], a puppet can follow a stick held in a user's hand. Video puppetry [3] allows users to manipulate paper puppets as control. Although articulated puppets have been taken into consideration, the current interaction is limited: a user can only rotate the segment to change its orientation while shadow puppet manipulation needs more complicated motion modes such as multi-jointed motion and simple pendulum movement. Moreover, there are some interaction techniques looking at interacting with digital objects, especially in tabletop systems. In Andrew's work [1], contact points with other input information detected by the touch screen are combined with physical simulation. As far as we know, none of these systems provides the off-the-shelf interaction techniques that are fine-grained enough to manipulate a shadow puppet.

A lot of software enables the creation of skeletal animation. For example, Flash's bone tool [11] can add virtual bones to the animation character from segments of limbs. To animate the bone structure, a user needs to insert a number of key frames and in each frame, drag the character in a new position. These tools are powerful in creating skeleton animation with rich effects. However, in informal application, this approach is uneconomical in time and not straightforward since the character cannot act until finished.

Performance animation as another method has also been studied. Various input devices such as human body with sensors, joysticks, and stylus are used for the animation creation. Sageev [9] establishes a tangible interface to control the 3D human character's segment orientation. Matthew [7] designs an interface where user can use input sketch to set the motion of a character. Here in the solution of our problem, we take the "sticks from hands to the puppet" metaphor as our method.

Being a kind of the articulated puppet, Chinese shadow puppet is used as an interaction metaphor for building logical narrative [2]. In addition, how to generate Chinese shadow play animation has also been considered. Yi-Bo [10] develops the technique to mode the image of the shadow puppet and renders an animation. Tsai-Yen [8] develops a motion planning technique. In his work, a pendulum model is used for the lower body motion, which is also integrated in PuppetAnimator. While their work focuses on the algorithms to give compliant motion for the shadow puppet, our system considers the problem to provide an easy user participation interface.

III. DESIGN FOR PUPPETANIMATOR

PuppetAnimator makes use of the shadow play performance workflow. To use it, first thing to do is to authorize a virtual shadow puppet character. Next, perform the shadow puppet's action in the "stage panel" of the interface. At last, replay or save the result into animation files. We use the articulated figure model to describe shadow puppet character. Each puppet consists of a set of hierarchical rigid segments such as head, body, arms and legs as well as joints to link them together. To reproduce the feeling of manipulating real shadow puppets, we analyze shadow puppet's common animation actions, decompose them according to the difference of segments and design a set of motion modes to represent them. Based on those motion modes, puppet templates can be designed including the definition of segments linking relationship and the motion mode each segment follows. User can further add images for the skin of the segments and map point device input channels to the points of force application on segments.

Figure 2. A puppeteer uses sticks to manipulate a puppet [12].

A. Shadow Puppet Motion Modes

In traditional Chinese shadow play, a puppet is animated with several sticks whose ends are held by a puppeteer (Figure 2). The ends of the sticks link to the points of force application on the puppet's head, body, arms or legs. By operating the sticks cooperatively in hand, the puppeteer can perform various puppet actions. Usually three sticks are held, one on the body for the translation of the whole puppet and two on the hands for the movement of arms. Sometimes another stick may be added for the rotation of the head. The segments linking directly to sticks will drive others under the laws of physic. As the puppet is held upright, gravity also affects its movement. For example, the resultant force of gravity and dragging by the body will cause the two legs to swing like a pendulum in use for representing walking action in the drama.

We design a basic set of element motion modes that a simulation system should support to achieve common puppet's actions. Each motion mode can drive one or two segments to move in respond to an external force created by point input or neighbor segments. The motion modes are *Translation*, *Swing*, *Drag*, *Rotation* and *Two-joint Drive* (Figure 3). By configuring those motion modes to certain segments of the puppet and using contemporarily, a user can arrange various actions through simple multipoint input.

1) Translation: To change a puppet's position on the stage is a basic operation in shadow play. Usually there is a stick dragging the body which then drives other segments to translate iteratively while keeping their relative position and orientation constant. In the system, we design Translation as a motion mode to enable a segment to move without rotation following its parent segment or input point. (Figure 3a).

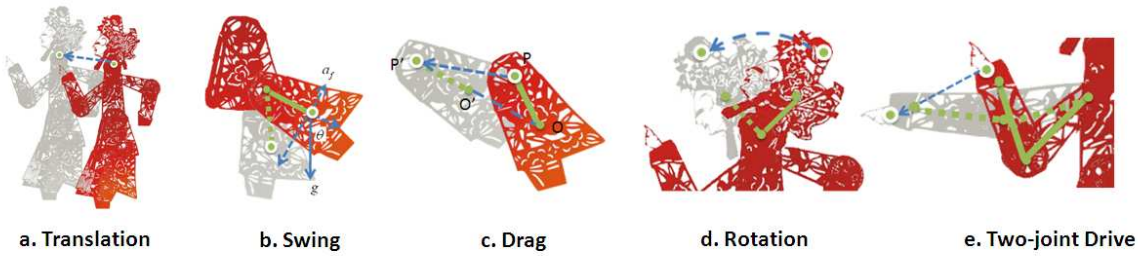


Figure 3. Motion Modes. External force (blue line) acts on the segment through the point of force application (green dot with a white circle).

2) Swing: Gravity and friction creates single pendulum motion to a segment (Figure 3b). In PuppetAnimator, Swing is designed to produce gravitational effect on the shadow puppet. We use the simple pendulum model to describe its motion which makes the segment looks like suspending on its parent. This mode can be installed on the lower boy of the puppet for the actions such as jumping when the legs are off the ground or walking with two legs swaying back and forth.

3) Drag: If moment is taken into account when a segment is pulled, not only its position but also its orientation will be changed. Drag is designed to achieve this effect (Figure 3c). When driven by it, a segment will move at a minimum while the distance between the center and the point of force application remains constant. We adopt the RNT algorithm [6] to model this mode. It can play a role in calculating the instant position and orientation change when a segment with mass is pulled.

4) Rotation: For a segment, if it is attaching to a fixed point when an external force works on it, it will rotate around the fix point. We design Rotation to enable this effect (Figure 3d). It causes the segment to rotate for the angle that best matches the external force. This can be applied to the head of a shadow puppet for the nod action.

5) Two-joint Drive: The arms are the most flexible segments that a puppeteer will take a lot of effort to control. In the performance, there is a stick linking to the each hand and driving the lower and upper arms to move cooperatively. To support the users of our system to achieve flexible arm actions, we design Two-joint Drive (Figure 3e). When a force pulls from the end, the position and orientation of two linked segments will be recalculated according to the mode. As a problem of Inverse Kinematics, Cyclic Coordinate Descent (CCD) algorithm [5] is chosen to implement it.

Based on the set of the motion modes described above, we can assemble them into a puppet template. First, design a skeletal structure. Second, choose several motion modes and assign them to some segments describing how to respond to an external force. We have provided several puppet templates of carton humans and animals. The skin image of the segments and point input devices are left for the user to customize in XML files.

B. Multipoint Input

We refer to the stick manipulation in real shadow play for the interaction in our system. Point input devices such as mouse and touch pad are used instead of the sticks to control the virtual shadow puppet. Each channel of the input point is mapped to a point of force application located in a segment (Figure 1). When a user moves the mouse or his fingers on the touch pad, messages of the movement will be sent to the PuppetAnimator. Then the message will be mapped as an external force to pull the point of force application over a distance. The motion mode will calculate the new position and orientation for the segments involved. Besides, the puppet can be turned around by clicking right mouse button.

Take the puppet template shown in Figure 1 as an example, we use the Translation mode to act on the body and bind the input from a mouse to it. Meanwhile, two touch points from a touchpad are in charge of creating external forces onto two arms which are driven by Two-joint Drive. Besides, two legs are under the gravity action. Other segments are also linked and driven by Translation or Drag. To perform with this shadow puppet, a user can drag the mouse to make it move with two legs swaying. He can also slide his two fingers on the touchpad controlling two arms to achieve various gestures.

C. User Interface

The main window of our system (Figure 1) has three panels. The upper left one is the stage panel, the performance area where puppets are played. Its background can also be customized. The right one is a list of stage configuration information. The control panel is located at the bottom which includes time line and buttons to control the process of the performance such as play, record, stop, save as AVI and save as GIF.

To make a little animation, first press the record button which starts the recording process, then perform the puppet's actions and press the stop button to end the recording. After that a user can replay what he performed just now or save it as an AVI or GIF format file.

IV. Results and Discussion

The prototype of PuppetAnimator is implemented in Windows using Java. The motion modes are implemented as a separate engine library. The interface is built on Java Swing and displays puppets driven by the motion mode engine. Besides, we use source codes from two open source projects, AVIOuputStream (http://blog.hslu.ch/rawcoder/2008/08/12/writing-avi-videos-using-pure-java) and GifDecoder (http://www. fmsware.com/stuff/gif.html) to encode AVI and GIF format file. We make use of famous cartoon characters as examples and design six puppet templates as well as five actions: falling from the sky, talking gesture, dancing with two arms, climbing and mouse stealing the oil. Each puppet template can be loaded to the system as a plug-in. The system can receive point input messages through UDP in XML format. Two types of senders, mouse and touch pad, are supported.

A pilot study is performed to obtain an initial feedback from the users. 8 graduate students in our lab were invited to join the user study. They were novices at performing shadow puppet. First they had five minutes to get familiar with PuppetAnimator. Then they were asked to make five action animations following the plot of story. Each action involves two or three kinds of motion modes. For each task, they could try as many as 5 times and select the best one they liked to submit. They were also asked to fill out a questionnaire when finished. The questionnaire included questions about the motion modes, interface and manipulation method. The user study was performed on a desktop computer with a mouse and a touchpad to generate multipoint input. The type of the touchpad was Wacom Bamboo Fun CTH-661, which could detect two touch points. To access touch points' position from the touchpad and send message to PuppetAnimator, a stand alone program, derived from the source code of Bamboo-TUIO (http://code.google.com /p/bamboo-tuio/), was written.

We receive positive feedback on the ability enabled by PuppetAnimator to experience shadow puppet manipulation and little animation production. The participants were exciting that they could control the puppet directly. Observations showed that they got started quickly and were able to utilize the motion modes to arrange puppet's actions. The set of motion modes can fulfill these actions' requirement. We also found some user chose to use fingers from two hands, one from each, to manipulate the puppet's two arms. When an action required three input points with two fingers from one hand touching on the touchpad and the other hand holding the mouse, users would spend more time getting familiar with the operation and suggest that using another touch device instead of mouse would make it easier. Questionnaire showed that users agree on applying only a few fingers to control and employing common point device makes the interface more accessible. They also made some suggestions to the features of the system. At the beginning, some confused the mapping of input points and points of force application, so they suggest that adding visualization feedback for points could give them stronger feeling of physical manipulation, which we modified in the updated version.

V. CONCLUTION

PuppetAnimator is an ongoing effort in building tools for non-expert users to experience a Chinese shadow play. We experimented with a prototype system to achieve several puppets' actions and the positive feedback gave us the confidence for PuppetAnimator's application. The real shadow puppet performance requires high level for the professional puppeteer who needs to take effort to practice their skills. Though our work has satisfied users for their basic requirement to manipulate a shadow puppet in their computer, the issue to enable a novice user to reach the expert level comprehensively and easily with the help of interaction tools still remains open. Future work may include extending the motion modes set to support more complicated actions, which aims to reach more natural, efficient interaction effects. A formal study including multi-user cooperation is also considered to draw a stronger conclusion about the effectiveness of our approach.

ACKNOWLEDGMENT

This study is supported by National High-Tech Research and Development Plan of China under Grant No.2009AA01Z336.

REFERENCES

[1] A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk, "Bringing physics to the surface," Proc. of the 21st annual ACM symposium on User interface software and technology (UIST'08), pp. 67-76.

[2] A. Paiva, M. Fernandes, and A. Brisson, "Children as affective designers-i-shadows development process," Humaine WP9 Workshop on Innovative Approaches for Evaluating Affective Systems(2006)

[3] C. Barnes, D. E. Jacobs, J. Sanders, D. B. Goldman, S. Rusinkiewicz, A. Finkelstein, and M. Agrawala, "Video puppetry: a performative interface for cutout animation", ACM SIGGRAPH Asia 2008, No. 124.

[4] L. A. Liikkanen, G. Jacucci, E. Huvio, T. Laitinen, and E. Andre, "Exploring emotions and multimodality in digitally augmented puppeteering", Proc. of the working conference on Advanced visual interfaces (AVI'08), pp. 339-342.

[5] L. T. Wang and C. C. Chen, "A Combined Optimization Method for Solving the Inverse Kinematics Problem of Mechanical Manipulators," IEEE Transactions of Robotics and Automation 7. 4 Aug. 1991, pp. 489-499.

[6] M. Hancock, F. Vernier, D. Wigdor, S. Carpendale, and C. Shen, "Rotation and Translation Mechanisms for Tabletop Interaction," First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06), pp. 79-88.

[7] M. Thorne, D. Burke, and M. Panne, "Motion doodles: an interface for sketching character motion", ACM SIGGRAPH 2004, pp. 424-431.

[8] T. Li and S. Hsu, "An authoring tool for generating shadow play animations with motion planning techniques," International Journal of Innovative Computing, Information and Control, Vol.3, No.6(B), Dec.2007, pp. 1603-1614.

[9] S. Oore, D. Terzopoulos, and G. Hinton, "A desktop input device and interface for interactive 3D character animation," Proc. Graphics Interface, pp. 133-140.

[10] Y. Zhu, C. Li, I. F. Shen, K. Ma, and A. Stompel, "A new form of traditional art: visual simulation of Chinese shadow play," International Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH 2003 Sketches & Applications), pp. 1-1.

[11] Flash Bone Tool, http://www.adobe.com/devnet/flash/articles/concept_bone_tool.html. Last access date December 14, 2010.

[12] wangronghua888.51.com

Exploring Temporal Ego Networks Using Small Multiples and Tree-ring Layouts

Michael Farrugia*, Neil Hurley†, Aaron Quigley‡
*School of Computer Science & Informatics
University College Dublin
Email: mike.farrugia@gmail.com
†School of Computer Science & Informatics
University College Dublin
Email: neil.hurley@ucd.ie
‡Chair of Human Computer Interaction
School Of Computer Science
The University of St. Andrews
Email: aquigley@cs.st-andrews.ac.uk

***Abstract*—Many of the current dynamic network visualisations methods or techniques rely on node-link force-based models that were originally developed for visualising static network snapshots. In this study, we diverge from this traditional layout approach and develop a layout for ego networks that places the time dimension in the foreground, by turning time into an element of shape. In addition to this we develop an interactive system that enables the visualisation of multiple networks simultaneously by employing small multiples. Using the proposed layout and analytical system as a grounding visual structure, we visually characterise dynamic network events in 3 different networks; the evolution of the biotechnology field, a phone call data set and a network of passenger connections of an airline. From this analysis we propose a range of ego network visual motifs that can be used as templates to identify and characterise events that are occurring in a dynamic network.**

***Keywords*-Dynamic networks, Ego networks, Small Multiples, Graph Drawing**

Social networks are dynamic systems in a state of constant change. People are born, form friendships, find partners, lose friends and die. Companies are created, merge, spilt and close. Traditionally, mainly because of the difficulties in collecting data, networks were studied by analysing a single snapshot of the network taken at a particular time. Today, dynamic network data is becoming increasingly available from sources such as social networking sites, telecommunication networks and customer transaction databases. Leading researchers in the field of network science have identified the study of *dynamic networks* as one of the next primary challenges in network theory [4].

From the very early days of social network analysis, visualisation has been a key element to the development of the field [21]. With access to dynamic network data, researchers in information visualisation are working to enhance the social scientist toolbox by developing tools and visualisation techniques that facilitate the study, understanding and exploration of network dynamics and the processes they represent.

Where dynamic social network visualisation has been studied the focus has been predominantly on new nodes entering the network, large scale networks, and the use of animation techniques to convey the dynamics within the network [5], [8], [15], [26]. Recent studies [19] comparing static and animated dynamic network tasks, have shown that contrary to common expectation, static displays are generally more effective in terms of both time and accuracy, when performing analytic tasks.

By contrast, instead of visualising dynamic networks using animation techniques we develop a system based on a *small multiples* visual interface. The principle behind small multiples design is to place several small visualisations near each other to facilitate comparison and pattern identification while maximising the ink to information ratio [35]. By displaying multiple networks simultaneously on the same page, we seek to overcome the weaknesses of animation while preserving the ability to understand the temporal nature of the network dynamics.

The motivation and tasks for this study originated from studying interactions between biotechnology companies, over a span of 10 years. The initial visualisation approach relied on using multiple node-link diagrams each representing a single year displayed on the same page as small multiples. This visualisation approach soon became impractical due to the large size of the data and scale free nature of connections. The resulting diagrams were complex, difficult to understand and poor according to most graph drawing aesthetic measures [13]. Additionally, the visualisation was cluttered because it was confined to the small space dictated by the small multiples design. Our next attempt refocussed on the analyst's tasks and asked the question, "how can the patterns of connections and events be mapped into simple visual images that can be generalised as visual motifs for dynamic network data".

We ground our study of dynamic social network visualisation in the tasks and problems faced by social scientists.

 ISBN: 978-1-61208-117-5

This grounding has allowed us to focus on the concept of an *ego network*. An ego network is a sub-network that focuses on an individual actor who is the subject of the network. The focal point of the network is called the "ego", whereas the other actors he interacts with are called "alters". In an ego network, only actors that are directly connected to the ego form part of the network. We describe an *event* as a change in the graph structure that occurs at a particular point in time.

A *network motif* acts as a visual fingerprint of a connection pattern. While motifs for static networks are commonly employed, especially in biology, the equivalent for dynamic networks do not exist. In [36], Welser et al address ego motifs to characterise roles in social groups. In order to visualise the temporal element of activity, they use a separate visualisation called an authorline, to supplement ego motif networks with temporal data. Here, we address the need to create a simple intuitive layout that can be used to describe temporal information in ego network motifs, a visualisation that is able to tell a story in time with a single diagram.

Our method to achieve this is through a tree-ring layout algorithm which converts several temporal aspects of data into visual shape elements. This form of visualisation provides a network analyst both a visual language for describing network motifs and the ability to focus on important changes in the links between actors in the network. New nodes and hence new edges are an important facet of dynamic network structure but so too are the nature, rhythm and pattern of interactions these new entities bring. Figure 1 shows the design of our small multiples system that combines a number of different network diagrams including traditional node link diagrams, matrix views and tree-ring ego networks, in one interface.

In this paper, we make two main contributions. The first contribution is the development of a new visualisation based upon a tree-ring metaphor that fits well within a small multiple design and is more structured and intuitive than traditional force based node link diagrams for displaying dynamic networks. The second contribution is the design of a small multiples based network exploration system that displays multiple network visualisations at the same time.

Traditionally small multiple visualisations focus on displaying attribute based visualisations such as scatter plots or bar charts. Here the small multiples system is used in favour of animation to visualise dynamic networks. The small multiples are intended to assist with the exploration of different relationships between actors in a network and help identify different roles and types of actors based on their individual connection patterns. With the help of this visualisation technique and interactive system we study 3 different data sets and provide analytical insights into the networks and the data within the networks. We also show examples of how the ego network motifs can be used to identify divergent or atypical patterns in the data.

I. Dynamic network tasks

Prior to proposing any new visualisation design for dynamic networks, one has to clearly understand the questions and tasks that the visualisations are intended to assist with. Past research in information visualisation has grouped user tasks into taxonomies according to the type of data in question. Robert Amar et al [2] and Plaisant et al [25] have extended the general study by Shneiderman [31] to attribute data and graphs respectively. As yet, there is no published taxonomy tailored towards dynamic network tasks, so in this section we attempt to describe the most common dynamic network concepts and tasks drawn from previous studies of such data.

Network structures can change in a variety of ways. However, the dynamic processes behind network evolution can be categorised in three overarching categories; *formation*, *rewiring* and *dissolution* [30]. These three general principles apply to both the atomic elements of the network (actors, links) and compound structures within the network, such as groups or communities [29].

If we consider the whole network structure, the formation of the network describes how the network was formed to exhibit the structure that it currently holds. If a time component is added to any structural feature in a static network, the temporal formation of that feature can be explored. This is the most basic means of enhancing a network with temporal information.

A more significant aspect of change occurs when there is rewiring within a structure manifested as a change in the links between actors. The term rewiring groups all changes that alter the connections within the current structure, possibly changing the type of the structure. For example, rewiring can include a split or a merge between members of a community. Rewiring is significant in the context of network evolution because these rewiring changes can raise important questions and lead to insights into the processes that are underlying the change.

Dissolution describes the process of any structural component or network member disappearing from the network structure. An example of dissolution is a scenario in which an actor in the network does not continue to from a part of the network after a certain time period. The propagation of a dissolution effect on the network is sometimes referred to as churn [12].

While the above types of change relate to changes in network structure, the properties of the members or actors of the network can also change. This type of change is important especially if it occurs as a result of the connections in the network. Christakis and Fowler investigate this type of change in their studies on obesity [10] and smoking habits [11]. In the biotechnology field evolution study by Powell et al [30], the authors use the term *multivocal* to describe the behaviour of actors that perform different roles

 ISBN: 978-1-61208-117-5

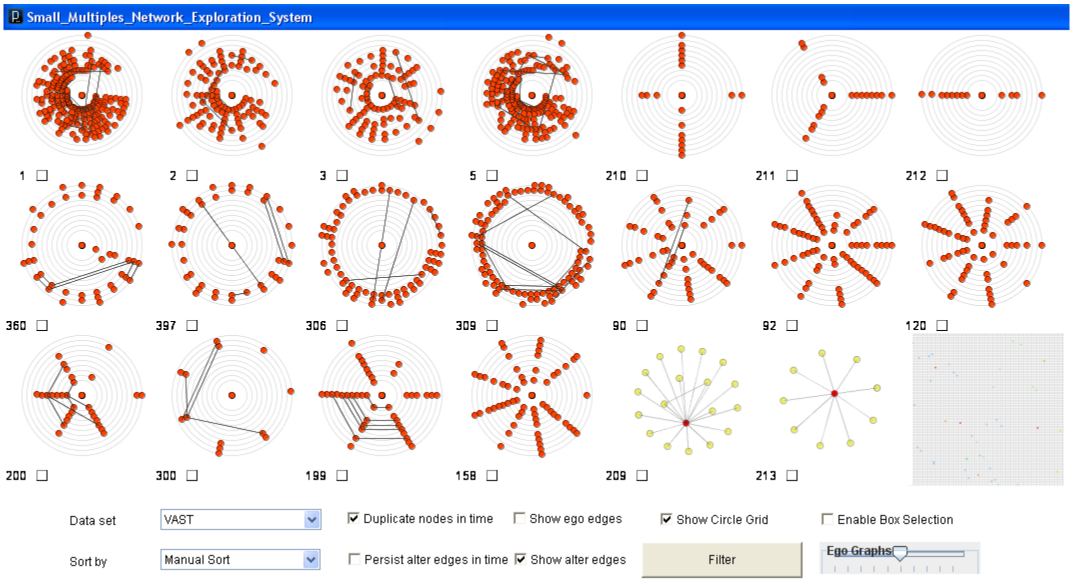


Figure 1. Small Multiples Network Exploration System

throughout the data collection period.

In dynamic social network visualisation, many of the current techniques focus on showing the formation of the structure over time [23]. In this study we fill in a gap for the task of analysing actor rewiring. Our visualisation addresses the question, "how can we visually describe the choice and change of partner preference over time". Along with this we also try to facilitate the understanding of how nodes change their properties or exhibit multivocal behaviour over time.

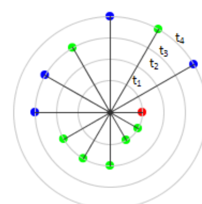


Figure 2. Tree ring ego network layout principle

II. Tree Ring Layout

The proposed node layout algorithm for dynamic ego network motif exploration was inspired by tree rings. Dendrologists study tree rings to determine the age of a tree and the amount of new growth of a tree in a year. In the tree ring layout, the ego node is first placed in the centre of the drawing (black node in Figure 2). From the central ego node a number of concentric circles, akin to tree rings, are drawn. Each concentric circle represents a time period in the network data (t_1 to t_4 in Figure 2). The alters are then placed on the concentric circles according to the time when the interaction between the alter and the ego takes place. The earlier the interaction, the closer to the centre the alter is placed. All alters are placed equidistantly from each other based on the total number of nodes in the ego network through the networks' timespan. In Figure 2 the first group of nodes that interact with the ego are the red node and two green nodes on ring 1. On ring 2, only green nodes interact with the ego, while on rings 3 and 4, a mixture of green and blue nodes interact with the ego.

The alter nodes start being positioned from the angular position zero with respect to the ego. In Figure 2 this position is occupied by the red node on time ring 1. The desirable visual characteristic here is to position the majority of alters that interact with the ego at one time point, close together in a logical ordering. The ordering of the alters within the same time period can be changed using the layout parameters (see Section II-B).

Each alter has a unique angular position with respect to the ego position in the centre. The uniqueness of this constraint avoids overlapping of the same nodes across different time periods. Assigning a unique angular position to each node also enables nodes to be replicated at each time point, when there is an interaction with the ego. For example, if a node interacts with the ego at time t_1 and time t_4 then the node can be shown both on the ring representing time t_1 and the one showing time t_4. The node is shown at the same angular position thus aligning it in a straight line with the first time the node was displayed. An example of this can be seen in Figure 3 label 2, where the orange node is replicated across 4 time periods.

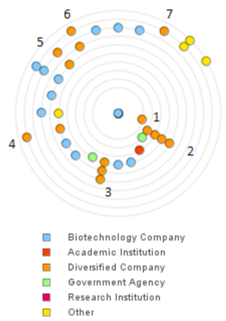


Figure 3. Example of the ego network of a biotech company

Rings do not necessarily have to represent a year in time or a uniform time pattern. A single ring can contain an aggregate number of time periods in it. If each ring is representing a time aggregate, then any node placed on the ring means it interacted with the ego during that time period. For instance, if the data set contains a range of a 100 years each time ring can be used to aggregate a uniform number of years, in this case 10 rings each representing 10 years. Alternatively, a period of years can be compressed on one ring and the others expanded in the rest of the rings. In this example, the first 90 years can be represented in the inner ring while the last 10 years can be displayed in their own individual rings.

A. Motivational Example

Figure 3 is an example of an ego network from the data analysed in the biotechnology case study (discussed in detail in Section IV-A). The explicit edges from the ego to the alters are omitted to evidence the node positioning of the layout and the meaning behind the ordering. The first interaction of the ego, a dedicated biotechnology company represented by a blue node, is with a diversified company, represented by an orange node (label 1). Notice how this node has the smallest angular position with respect to the ego. In the second year there is no more interaction with this company, but the ego interacts with a new diversified company (label 2). This relationship continues for the next four years. In year four, the ego interacts with another diversified company (label 3), for a duration of 3 years. Within this time period the relationship with diversified company 2 terminates. In the fourth year the ego also interacts with another diversified company (label 4), however the relationship has a two year gap which could have possibly been filled by the other diversified companies labelled 5, 6 and 7. In the fifth year when company number 2 is dropped three diversified companies 5, 6, 7 are introduced.

B. Layout Parameters and Interaction

The general layout description specifies that all the nodes that interact with the ego during a time period should be displayed on the same tree ring, however it does not impose an ordering of nodes on the same ring. The order of the nodes could follow a temporal structure. If each time ring represents a time aggregate, for instance a year, the nodes can be ordered by the day when they first interact with the ego. Alternatively, nodes can be ordered using an attribute of the data set, such as, the type of the node. This has the effect of visually clustering together nodes of the same visual or logical characteristic.

Nodes can also be ordered based on graph drawing aesthetic criteria, particularly edge crossings between the alters. Since each node in the tree ring layout occupies its own layer, a layer crossing minimising algorithm similar to Sugiyama's approach [33] or Brandes et al [7], can be used to reduce edge crossings. Alternate intra layer node layout techniques may also be adopted.

A number of parameters can be used to control the visual encodings of the elements in the visualisation. Explicit edges such as those connecting the ego to all the alters can be displayed or hidden (this is the case in most of the figures in the paper). Alter nodes and edges can be set to persist through the time period from the first time they interact with the ego. When persistence is enabled the elements in the network aggregate over time to produce a composite view of the network. The edges between alters can also be turned on or off or altered in intensity based on either edge importance or time properties.

In tree rings, the size of each ring is proportional to the growth of the tree during the year. This concept can also be visually applied to the layout by mapping the size of each concentric circle to the activity in the time period. This can be used to allow for the nodes to be spread out further in time. Intermediate time rings can also be introduced in between other time periods to improve the readability of dense time periods.

Interactive techniques can also be used to overcome issues with the selection and aggregation of time periods displayed in the ego network. If there are more time periods than the layout can support in a small multiple view, a brushing interactive motion on an ego network can let the analyst scroll thorough time periods to display activity in other time steps. A movement from the ego out towards the periphery would scroll forward in time, while a movement from the periphery towards the centre could scroll back in time. The time periods that are out of range can be aggregated on the first or last ring if required. Once this motion is complete for the reference network it can be easily applied to all the other small multiples.

III. Small multiples interactive system

A generic small multiples framework was developed to allow different types of visual representations in each individual small multiple. We extend this paradigm that is traditionally used to visualise attribute based data [20], to visualise network data. Figure 1 shows an example of a display where the majority of the diagrams are tree ring diagrams and the last 3 diagrams on the bottom right show traditional node link and matrix visualisations respectively.

When designing the small multiples based system we follow the design mantra of Shneiderman et al [31] of overview first, filter and details on demand as a guiding framework for interaction and exploration. In the overview stage the small multiples can be reduced in size to maximise the number of data elements that can be displayed and provide an overview of the data. In this stage different aggregations and abstractions of the network can also be used to visualise the overall structure of the network and its components. From these visualisations the analyst can interact with the system to select individual visualisations or filter the data by the properties of the network in order to focus on the elements of interest. To get more detail on individual elements within the visualisation, individual elements in the visualisation can be selected to obtain further information.

Each small multiple image consists of a visual representation in a small window and all images are displayed in a grid similar to a photo album. The number and size of the images can be controlled by the user at run time to balance between the amount of information displayed and the detail in each image. The image sizes range from 150x150 pixels up to 300x300. When displaying the ego networks at their smallest size one can get an *overview* of a 400 node network in 10 frames. All the images in a small multiples display are traditionally of the same size and this system follows this design pattern.

The ego networks can be sorted by different criteria such as; the node degree at a time period, node attributes, activity within the time period and the relative age of the ego. When paging through the list of multiples, the user can select a number of small multiples by selecting a checkbox near each image. Upon confirming the selection the display can be reconfigured to display the selected ego networks next to each other on the same page. This interactive features shows how *filtering* can be used from the small multiples display to enable the comparison and study of interesting networks. If more *details* are required for an individual image, the analyst can select the image and display that image at full size instead of showing that image in a small multiple view.

Apart from interacting with the visualisation as a whole, one can also interact directly with the node link visualisations. If a node is double clicked and that node exists in other ego networks on the same page, then the node is highlighted in all the other small multiples. To evidence this highlighting, all the the other node types can be faded and shown in the same colour. If the node is right clicked, the small multiples on the page change to show all the ego networks of which the clicked node is a member. A group of nodes can also be selected to change the small multiples displayed to the ego networks of the selected nodes. It is also possible to interact with the individual node elements to view information tooltips and/or node labels, depending on the available data.

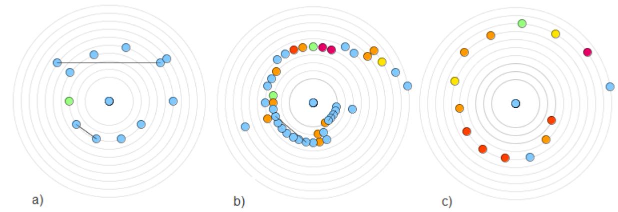


Figure 4. Homophily and multi-connectivity ego motifs

IV. Case Studies

In this section we analyse three different data sets to characterise the types of ego networks that are present in each data set and study the capabilities of using a tree-ring ego layout to develop ego network motifs. The three case studies are from thee diverse fields, the first one is from social science, the second one is a scenario based synthetic data set created for the VAST 2008 competition and the third data set is a commercial data set of airline passengers.

A. Biotechnology company data case study

In the first case study we use data from previous work in the social sciences, that studies the dynamics of "logics of attachment" in the evolution of the biotechnology field [30]. This data set contains examples of the analytical task that our visualisation sets out to assist with, i.e. changing patterns of connection.

The data used for this case study was compiled by Powell et al primarily from *BioScan*, an independent industry directory. Six types of companies are represented: dedicated biotechnology companies (2736), academic institutions (279), public research firms (177), government institutes (201), diversified companies (778) which include large pharmaceutical corporations and diversified health corporations, financial institutions and other types of companies (523). A connection between any two organisations exists whenever there is a news report that mentions a collaborative tie, alliance, contractual agreement or an exchange of resources between two or more organisations. The sample data selected covers a period of from 1988 to 1998 and the data is aggregated on a yearly basis.

The authors of the paper report that the biotechnology industry is characterised by many short term connections between companies. Once the connection is formed and the goal for which the connection was sought is achieved, then the two companies do not remain in contact with each other. For this reason, in this particular analysis the links between alters are only shown if they happen during the same time period.

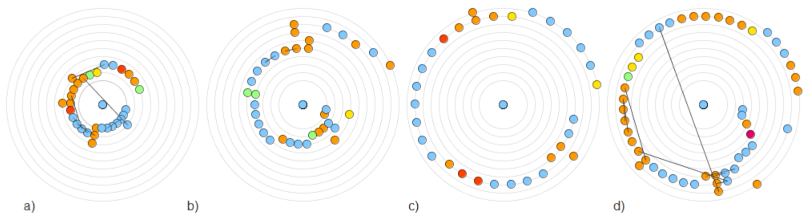


Figure 5. Different company evolution patterns

1) Network motifs and patterns: Powell et al discuss four types of partner attachment principles, *Accumulative advantage*, *Homophily*, *Follow the herd* and *Multiconnectivity*. We focus on three of these principles and explore other types of events that can be easily visualised with the tree ring visualisation. In the case study examples, each type of company is represented with a different colour and the explicit links between the ego and the alters are not displayed.

The principle of homophily [27] states that actors have a higher probability of making connections with actors that are similar to them. This sort of behaviour can be explained with the common saying "Birds of a feather flock together". On the contrary, multiconnectivity explains the behaviour when actors connect with different types of partners, thus diversifying their connections. Accordingly, organisations that exhibit homophily are more likely to connect to companies of the same type, therefore of the same colour. Conversely, multiconnectivity can be identified when companies connect to a variety of different companies, thus connecting to nodes of different colour.

Figure 4a shows an ego motif of a biotech company (blue node) exhibiting homophilious behaviour by primarily interacting with other biotech companies. Figure 4b shows an example of a biotech firm that starts out exhibiting homophilious relationships, but expands to interact with different types of companies in the later years. This pattern is a common pattern in the data set as the importance to forge connections with different types of companies increases with time. Ties with different types of companies are important for survival in such a competitive market, since different ties give access to different resources in the network. In Figure 4c the ego has a connection with different types of nodes, showing a good example of multiconnectivity.

The pattern of accumulative advantage [3] shows how actors with a high number of connections have a higher likelihood of getting new connections. This property is a feature of the entire network that is difficult to visualise when considering individual ego networks, as these do not generally provide a complete network perspective. Using the developed small multiple system, we can visualise the change in degree between the most connected nodes and the rest of the nodes. When ordering the ego networks by degree, a clear difference emerges from the first few nodes with a high degree of interconnections and the rest of the ego networks.

In an aggressive domain such as the biotechnology industry, companies tend to forge alliances, merge with other companies, or acquire new companies to gain access to technology and resources. Ego evolution visualisations can be useful to identify these behavioural patterns by following the temporal outline of the visualisation. Figure 5 shows four examples of company evolution. In Figure 5a the company has a concentrated amount of activity in the early years but this activity ceases soon after. This is visually indicated by the concentration of nodes placed on the first three inner circles. This pattern is indicative of a company being bought out and replaced by another company, or a company failing and ceasing to exist. In Figure 5b the company has a consistent level of activity throughout all the years, visually indicated by a general spread of nodes on all the circles. In Figure 5c the company is a late comer, as there is no activity at all in the first 6 years and all the activity is concentrated in the last 4 years. This is of particular interest because of the high level of activity without prior build up. This pattern might be indicative of an old company opening under a new name, which will explain the significant number of connections in the last few years with no activity in the early years. A similar pattern of connection can be seen in Figure 5d, however here there is some activity in the early years, which quickly increases in later years. This pattern could be indicative of an acquisition or a merger with a bigger company that already has many connections.

The properties of actors in a network can change if data is collected over a long period of time. This change can be important both in terms of the new type of interaction and also to understand if the change occurred as an effect of network connections. In this case study, companies can fulfill different operations and take a different role that changes their type of operation, thus changing their colour. Figure 6 shows some examples of nodes that exhibit this behaviour. These nodes can be easily identified because each node is replicated on the same radius with the outermost circle. As the nodes on the radius are aligned in a straight line it is easy to identify the change in node colour of the same node.

 ISBN: 978-1-61208-117-5

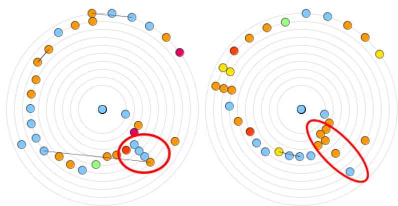


Figure 6. Actors changing type over time

B. VAST 2008 Phone Call Network

In the second case study we use the VAST 2008 phone call data set [24] that consists of a set of phone records between families in control of a controversial religious organisation. Each record has the time of the call, duration and the location of the cell tower from where the call was made. The data spanned a 10 day period and contained 400 unique cell phones.

The main analytical task in the VAST mini challenge was to characterise the change in the network occurring in the 10 day period of data collection. In this case study we use the tree-ring visualisation within the small multiples system, to categorise and characterise the different types of network actors based on their interaction patterns and their possible change in behaviour over the 10 day period. After having identified the important actors in the data set using ego network visualisation, different visual and analytical tools can be used to investigate how the interactions between these prominent nodes evolves through time [17].

Figure 1 (on page 2) shows examples of some of the most prominent actors in the VAST data set. The first set of nodes (1, 2, 3, 5) are the nodes that have the highest number of connections, however we observe that in the outer circles that represent the later time periods, the number of connections is close to zero. If we study the nodes under the first set (360, 397, 306, 309) we notice that these exhibit a similar but inverse behaviour to the first set of nodes. They are very active nodes but are only active in the last 3 days of the time period.

A very common ego network motif is the one exemplified by nodes 210, 211 and 212 where the ego regularly interacts with the same 2 or 3 alters throughout the whole 10 day period, with little or no interaction between the alters at the same time period. This is typical of scale free networks and is indicative of the same accumulative advantage pattern seen in the biotech network where only few nodes have a large number of connections and most have few connections. Other network ego motifs that are common are those exemplified by nodes 90, 92, 120, 158 where the ego interacts with a number of nodes in a regular pattern with gaps between some time periods.

Nodes 199, 200 and 300 are examples of atypical ego motifs because of the high number of interactions between alters when compared to the number of alter interactions in the other ego networks that was very limited. In the scenario description node 200 is described as an important node. From the visualisation we notice that node 300 has a similar ego network motif to node 200 which prompts us to flag is as suspicious. The geography and call duration can be encoded by colour and node size respectively. Using this representation, actors that are changing their geographic location will exhibit the same multivocal patterns as seen in Figure 6.

C. Airline Passenger Network

In the third case study we visualise a sample of an inferred network of airline passengers [16], [18]. The network is automatically inferred from the bookings records of around 430 unique passengers spanning over a period of travel of 10 weeks. The passengers are uniquely identified using an entity resolution pre-process as a unique identifier was not available for all passengers. A relation between two passengers exists if they are booked together on the same booking.

Unlike the VAST and the Biotech network data sets the airline customer data set contains many cliques between passengers travelling together. Each time three or more passengers are booked on the same booking a link is created between all the passengers in the booking, thus creating a clique connection pattern.

Figure 7 shows five representative patterns from the airline passengers data set. The first network illustrates the case where the ego travels with two parties of 4 and 5 passengers each, towards the end of the 10 week time period and a separate trip with another passenger. The second ego network is typical for most egos in that they travel once in the sampled time period but with a number of people in the same booking.

The third network illustrates an interesting pattern of a passenger travelling with the same group of people on subsequent weeks. This can be seen with the overlapping triangles shown in the top left corner. The same passenger travels with other groups of people during the other weeks.

The fourth pattern is interesting because it goes against our intuition and is quite uncharacteristic (there were only two egos with this pattern). In fact this shows the pattern of an actor who has a lot of trips with many separate people, few of which are linked. Upon further analysis of these actors we realised that this was in fact an anomaly which was showing an instance of actor misidentification in the entity resolution stage. Here more than 4 different passengers were considered to be the same person therefore the individual trips of those four people resulted in multiple sparse links in this ego network motif.

The last pattern shows an actor travelling with two groups of people in different time periods where one of the members in the group overlaps. This almost results in a diamond

 ISBN: 978-1-61208-117-5

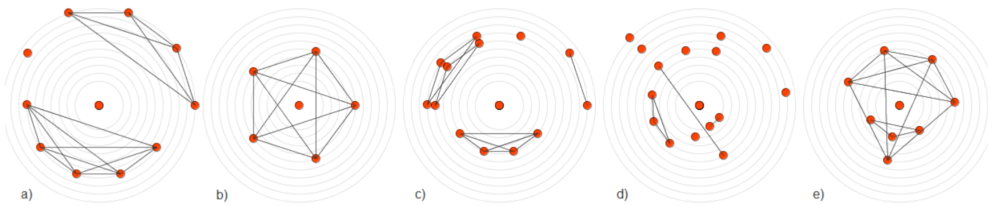


Figure 7. Airline passengers ego network nodes

shaped pattern with the extension of one of the nodes reaching out in a different time slice.

V. Related Work

The layout approach proposed in this paper is a type of radial visualisation that uses a circular pattern to position network nodes. The widespread adoption, variety and applicability of radial visualisations was the subject of a recent survey paper by Draper et al [14]. In this work, the authors categorise different radial layouts based on their visual properties. Our visualisation contains elements of three of these patterns – the star pattern, the tree pattern and, to an extent, the concentric circle pattern. The defining characteristic of the star pattern is the centre of the diagram from which edges originate. In our ego diagrams, the ego is placed in the centre and edges originate from the ego to connect the ego to all the alters. While all leaf nodes are connected to the central node with a straight line, which is a characteristic of star patterns, the inner nodes can be connected between themselves too which is a pattern exhibited in the tree pattern. Thirdly, concentric circles are a critical positioning element in the visualisations where time representation becomes a key issue. The novelty of our approach lies in the combination of a star based layout with an underlying concentric circle field that is used to encode a positional dimension for time.

Circular layouts were amongst the first types of layouts used by social scientists to draw the first social network visualisations [21]. Perhaps the most famous example is Northway's target sociogram [28] that uses a series of concentric circles to represent actor centrality and guide the placement of nodes by placing the most structurally central nodes in the innermost circle. Brandes et al [6], in their work on visualising policy networks, use Northway's original target sociogram principles and elaborate on a method to improve aesthetics when automatically drawing such diagrams. An approach that uses the tree ring metaphor to draw nodes is presented in [34]. In this layout, a hierarchical tree structure is placed along concentric circles representing different time periods.

There is further novelty in our approach of displaying the circularly laid-out ego networks as small multiples, allowing in a single visualisation, the different temporal patterns of different ego networks to be compared. Thus, the 'when' question, which by definition is central to any dynamic analysis is emphasised in each small multiple and differences in temporal patterns are easily discernible by scanning across the multiples. Our work also complements the ideas presented in [32] where the authors use node attributes, including temporal attributes, to guide the node placement.

The idea of using small multiples to visualise networks was explored by Chi and Card in their work on visual spreadsheets [9] where they apply the spreadsheet metaphor for visualisation. In this paper, the authors visualise complete dense scale free networks of the internet over a sequence of steps. The problem with this approach is that large scale free networks tend to be difficult to visualise using node link diagrams with the effect of reducing the readability of the networks. Adamic et al [1], use small multiples to visualise ego networks to explain the types of egos that are present in online Q&A forums. In a recent study [22], the authors use small multiples to visualise network properties such as edge count, density and degree, to facilitate network exploration, but they do not attempt to visualise relationships between nodes in their small multiple displays.

VI. Discussion

The embedding of the temporal dimension in a single network image enables us to use a small multiple design to display different egos simultaneously, showing how different types of egos evolve. The small multiples design allows the analyst to explore and identify patterns and differences in the data, however not all types of visualisation are appropriate for small multiples.

Traditional node link layouts have two disadvantages in this regard. Firstly, as traditional layouts only represent a single snap-shot in time, multiple images are required to explore the network over time. As each network is only showing one time period, the number of images is a multiple of the number of time periods. Identifying change in multiple images requires a mental comparison between multiple images which increases the cognitive load on the analyst. Secondly, certain patterns of connection can create diagrams that are too complex with many overlapping nodes and edge crossings that make the diagram unreadable.

 ISBN: 978-1-61208-117-5

The limitation of space when visualising data in small multiples automatically limits the number of data elements that can be visualised. Better strategies do exist to maximise the amount of information displayed, for instance matrix representations, but the strong metaphor that the tree ring provides makes this layout very intuitive to understand and use. This is particularly beneficial when introducing a new visualisation to users who are not information visualisation specialists but rather specialists in their own fields who need tools to help them analyse their data.

Furthermore, the ‘lost’ space that is intrinsic in a circular shape when bin packing serves as a natural delimiter between different small multiple images. A square or rectangular shape that might enable denser packing might still require the same amount of space in between different images, to make it possible to discern between different images. It is also still not clear how small dense matrices that encode time on one of the axis can be useful to convey the path connection information between the alters. We intend to follow up with user studies on these aspects to compare the different visualisations.

In the three case studies used we make use of two real and one synthetic data set that was modelled on real networks to serve as a realistic scenario for a conference competition. In the three examples, the number of network nodes increases over time in line with what is usually observed in dynamic networks. One natural aspect of concentric ring layouts is that the inner circles are smaller than the outer circles. This visual characteristic fits well with the pattern of network evolution since the increase in the number of nodes over time is in line with the increased amount of space on the outer circles of the visualisation.

When visualising individual ego networks, the network is being decomposed into its constituent parts. This particular network perspective is tailored towards understanding the individual actors in a network rather than the network structure as a whole. Tasks suited to this technique include identifying change in connection patterns, visual characterisation of network actors based on common patterns of interaction over time, and identifying actors that diverge from those network patterns, the outliers. The majority of actors in a large scale network might not be those that an analyst might want to focus on, based on the typical connection properties of most large scale networks. The filtering and ordering interactions built in the interactive small multiples system can help the analyst focus on the more interesting actors.

The applicability of the tree ring layout can be extended and applied beyond a small multiple system. For instance, the visualisation can be used to provide details on demand when combined with a traditional node link layout. In this case, the tree-ring layout can be used to display an overlaid expanded image of a node’s ego network when the node is selected. A sort of “ego lens” feature.

VII. Conclusions and Future Work

In this paper, we describe an intuitive approach to visualise dynamic ego networks using time to guide the positioning of nodes. We take advantage of the compact nature of the generated diagrams to create a system based on small multiples that allows the analyst to explore the network through ego networks. In the 3 diverse case studies, we show how different types of networks can lead to different motifs, we found some similarities and generalities between motifs and also discovered some unexpected patterns that can highlight suspicious activity or issues with data quality.

Based on these positive results we plan to further our studies of this visualisation within a small multiples design. We plan to conduct user studies to study the efficiency of this technique and are also working to provide a similar tool to social scientists to assess the applicability of this technique for other networks in their studies. As part of the user study we would like to study how this visualisation compares to traditional representations along with the effect of unwinding the circle and representing the data in a linear fashion. This should help us get a better understanding of the intuitive positive characteristics of the circular structure of the diagram.

While there are clear benefits to this type of visualisation we are aware that it is not a silver bullet for any dynamic network visualisation problem. The social scientist analysing dynamic networks should be armed with a toolbox of techniques that can be applied according to the problem being investigated and the need to solve it. The advantage that this layout technique is that it is a simple design and easy to understand based on a universally recognised metaphor which makes it a good candidate to integrate into any existing social science toolkits.

References

[1] L. Adamic, J. Zhang, E. Bakshy, and M. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceeding of the 17th international conference on World Wide Web*, pages 665–674. ACM New York, NY, USA, 2008.

[2] R. Amar, J. Eagan, and J. Stasko. Low-Level Components of Analytic Activity in Information Visualization. *Proceedings of the IEEE Symposium on Information Visualization*, pages 111–117, 2005.

[3] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[4] R. J. Barbara, M. Z. Laura, and E. Marshall. Connections - Introduction to Complex Systems and Networks Special Issue. *Science*, 325(5939):405–, 2009.

[5] S. Bender-deMoll and D. McFarland. The art and science of dynamic network visualization. *Journal of Social Structure*, 7(2), 2006.

[6] U. Brandes, P. Kenis, and D. Wagner. Communicating centrality in policy network drawings. *IEEE Transactions on Visualization and Computer Graphics*, pages 241–253, 2003.

[7] U. Brandes, J. Raab, and D. Wagner. Exploratory network visualization: Simultaneous display of actor status and connections. *Journal of Social Structure*, 2001.

[8] J. Branke. Dynamic Graph Drawing. *Lecture Notes In Computer Science*, 2025:228–246, 2001.

[9] E. Chi and S. Card. Sensemaking of evolving web sites using visualization spreadsheets. In *Proc. InfoVis 1999*, pages 18–25, 1999.

[10] N. Christakis and J. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370, 2007.

[11] N. Christakis and J. Fowler. The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358(21):2249, 2008.

[12] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi. Social ties and their relevance to churn in mobile telecom networks. In *EDBT '08: Proceedings of the 11th international conference on Extending database technology*, pages 668–677, New York, NY, USA, 2008. ACM.

[13] G. Di Battista, P. Eades, R. Tamassia, and I. Tollis. *Graph Drawing; Algorithms for the Visualization of Graphs*. Prentice Hall, 1999.

[14] G. Draper, Y. Livnat, and R. Riesenfeld. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics*, 15(5):759, 2009.

[15] C. Erten, P. Harding, S. Kobourov, K. Wampler, and G. Yee. GraphAEL: Graph animations with evolving layouts. *Lecture Notes in Computer Science*, pages 98–110, 2003.

[16] M. Farrugia and A. Quigley. Enhancing airline customer relationship management data by inferring ties between passengers. In *Proceedings of the international conference on Social Computing*, 2009.

[17] M. Farrugia and A. Quigley. TGD: visual data exploration of temporal graph data. In *Proceedings of SPIE*, volume 7243, page 724309, 2009.

[18] M. Farrugia and A. Quigley. Actor identification in implicit relational data sources. In I.-H. Ting, H.-J. Wu, and T.-H. Ho, editors, *Mining and Analyzing Social Networks*, volume 288 of *Studies in Computational Intelligence*, pages 67–89. Springer Berlin, 2010.

[19] M. Farrugia and A. Quigley. A comparative study of animation versus static display methods for visualising dynamic social networks. *Information Visualization*, In print.

[20] S. Few. *Information dashboard design: the effective visual communication of data*. O'Reilly Media, Inc., 2006.

[21] L. C. Freeman. Visualizing social networks. *Journal of Social Structure*, 1(1), Feb. 2000.

[22] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck. Manynets: an interface for multiple network analysis and visualization. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 213–222, New York, NY, USA, 2010. ACM.

[23] Y. Frishman and A. Tal. Online dynamic graph drawing. *IEEE Transactions on Visualization and Computer Graphics*, pages 727–740, 2008.

[24] G. Grinstein, C. Plaisant, S. Laskowski, T. O'Connell, J. Scholtz, and M. Whiting. Vast 2008 challenge: Introducing mini-challenges. *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 195–196, Oct. 2008.

[25] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *BELIV '06: Proceedings of the 2006 AVI workshop on BEyond time and errors*, pages 1–5, New York, NY, USA, 2006. ACM.

[26] L. Leydesdorff, T. Schank, A. Scharnhorst, and W. De Nooy. Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling. *El Profesional de Informacion*, 2008.

[27] M. McPherson, L. Smith-Lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[28] M. Northway. A method for depicting social relationships obtained by sociometric testing. *Sociometry*, 3(2):144–150, 1940.

[29] G. Palla, A. Barabási, and T. Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664, 2007.

[30] W. Powell, D. White, K. Koput, and J. Owen-Smith. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences 1. *American Journal of Sociology*, 110(4):1132–1205, 2005.

[31] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343, 1996.

[32] B. Shneiderman and A. Aris. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):733–740, 2006.

[33] K. Sugiyama, S. Tagawa, and M. Toda. Methods for visual understanding of hierarchical system structures. *Systems, Man and Cybernetics, IEEE Transactions on*, 11(2):109–125, 1981.

[34] R. Theron. Hierarchical-temporal data visualization using a tree-ring metaphor. *Lecture Notes in Computer Science*, 4073:70, 2006.

[35] E. Tufte. *Envisioning Information*. Graphics Press, 1990.

[36] H. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the signatures of social roles in online discussion groups. *The Journal of Social Structure*, 8(2):20036–1903, 2007.

A Case Study of Prototyping a Multimodal User Interface for a Media Annotation Tool

Dominik Ertl, Marie Kavallar, David Raneburger
Vienna University of Technology, Institute of Computer Technology
A-1040 Vienna, Austria
{*dominik.ertl, marie.kavallar, david.raneburger*}*@tuwien.ac.at*

***Abstract*—Media annotation is the process of adding annotations to media, like audio or video data. Annotations are, e.g., emotion descriptions of human emotions. The manual creation of annotations typically requires to repeat small tasks many times. Manual annotation is time-consuming and erroneous because user interfaces (UI) for such annotation tools often lack the possibility of multimodal interaction. In this work, we present a case study were we prototyped multimodal UIs for media annotation. First, we identified time-consuming tasks in the process of media annotation. Then we studied the human-computer interaction, to find out which modality combinations fit well for these tasks. This led us to suitable variants for modality combinations, like speech input, mouse gestures, earcons and an adapted GUI. We used the OpenInterface platform to implement prototypes of these multimodal UI variants for an existing GUI-based media annotation tool. Our prototyping approach allows easy change and adaptation of the multimodal UI. This supports the designer during the multimodal UI development and leads to UIs for media annotation tools that have a well-balanced set of modalities for interaction purposes.**

***Keywords*-Multimodal, Media Annotation, Prototyping**

I. Introduction

Media annotation is the process of adding meaningful annotations like the emotional state of a human to media content, e.g., audio or video data. Research and industry provide dedicated annotation tools for different kind of media, like audio/video streams and files, virtual worlds, etc. Manual media annotation for audio and video is typically done with dedicated time line-based annotation tools like ANVIL [2] or ELAN [3] that usually come with a WIMP (Window-Icon-Menu-Pointer) based UI. Manual media annotation, however, is still an error-prone and expensive task.

In our work, we present a case study where we improved such a time line-based media annotation tool with multimodality. Multimodality — as presented in the work of Reeves et al. [13] — allows to make human-computer interaction more robust.

In our case study we focused on prototyping of such a multimodal UI and the interaction with them. Therefore, we studied the interaction for various modalities and modality combinations like speech input, mouse gestures, a 3D-mouse, etc. This helps to better understand the interaction needs of annotation tool users and therefore, to further improve the process of manual annotation. In our approach we used an existing media annotation tool, the Smart Sensor Integration (SSI) [8], which we coupled the OpenInterface (OI, Lawson et al. [11]) platform for prototyping the UI. This combination supports rapid prototyping of multimodal UIs for media annotation tools and thus, supports the designer in finding a suitable combination of modalities. Figure 1 provides an overview in chronological order of the work done in our case study.

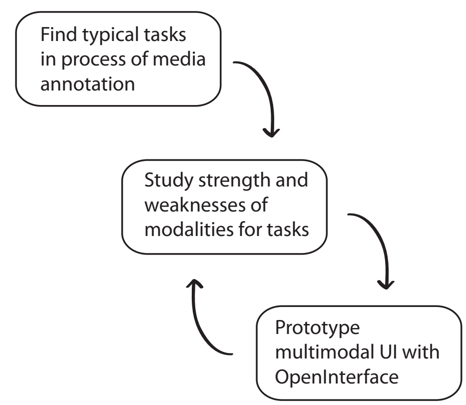


Figure 1. Overview of our Prototyping Approach.

The remainder of this work is organized in the following way: First, we provide background information about media annotation and prototyping of multimodal UIs. Subsequently, we describe how we studied interaction and implemented a multimodal UI prototype for media annotation. Then we discuss the lessons learned, considering benefits as well as drawbacks and pitfalls. Finally, we present related work.

II. Background

The work of Wagner et al. [9] points out that we have to give the computer access to human generated signals and provide adequate models to recognize and interpret behavioral patterns. The annotation process can be done either automatically or manually. Automatic annotation is often based on pattern matching with statistical models (Lavrenko et al. [10]). For example, an audio file is automatically

 ISBN: 978-1-61208-117-5

screened for specific emotions like giggling or crying. If an emotion pattern matches, the automatic annotation tool stores metadata that includes start and end time of the detected pattern, as well as the emotion description. Even if the automatic annotation tools are improving, a human user usually has to check the resulting annotations and adapt them, to achieve a good level of quality. Adapting the annotations has to be done manually if there are no existing algorithms that can detect specific patterns. For example, detecting the color of t-shirts in video files (if this is of interest). In principle, the annotations of media can be stored in a data base. It is then possible to search for specific media sequences via a search mechanism. For our work it does not matter what the annotations are used for, since we mainly focus on the multimodal UI of the annotation tool.

We developed a prototypical multimodal UI based on the SSI [9] tool. SSI is a framework for multimodal signal processing in real-time. It has two functionalities, the acquisition of audio and/or video data and its annotation. In our work, we do not enhance the underlying signal processing tool, but the ModelUI, which is a WIMP-GUI that runs on top of SSI.

For a robust interaction the designer has to optimize the interplay of several modalities. Prototyping a multimodal UI is often helpful to get a deeper understanding of what users need when interacting with a special purpose annotation tool like SSI. Prototyping, moreover, allows to easily attach and detach modalities at design time. In our case study we used OI from Lawson et al. [11], to implement our multimodal UI. OI is a free visual programming environment to prototype interaction. OI enables the UI designer to link components together to design the interaction of a multimodal UI. The documentation of OI is up-to-date and comes with a tutorial and in-depth information about using OI.

III. Multimodal UI for Media Annotation

In this work, we studied and enhanced the SSI annotation tool introduced by Wagner and Andre [8], which is intended to train models for improved recognition of human input behavior. As shown in Figure 1, we identified common tasks for timeline-based media annotation as a first step. Then we iterated over the steps *defining interaction* and *prototyping (implementing) the UI*. In the following, we present our approach in more detail.

A. Tasks in Media Annotation

The decision for the SSI tool as a basis was based on two reasons : First, it is a straightforward media annotation tool that only has a graphical UI. Second, the source code is available for free.

Since some of the code for annotation functionality was mixed up with the UI code, we first had to extract the modality independent functionality from the UI code, to clearly separate the application logic from the UI (see Figure 2, the yellow boxes depict our work). The extracted functionality is typical for timeline-based media annotation tools: start/stop playback of media content with a media player, add/remove annotation elements (segments), edit annotations, add several tracks (for different annotation levels), etc. In Figure 3 we depict the adapted SSI interface where some of the extracted functionalities are marked.

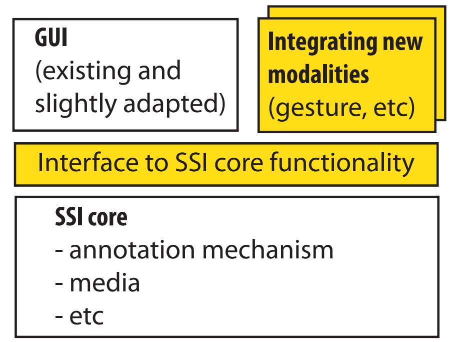


Figure 2. SSI Architecture.

As a second step, we studied the interaction with several media annotation tools, including other tools like ANVIL or ELAN besides SSI. Subsequently we annotated several audio and video files (with a lengths between 1 and 5 minutes) with the existing SSI tool. We measured the time for the different working steps and identified the following set of tasks as the most time consuming ones in terms of human-computer interaction:

- *Create and select annotation tracks.* An annotation track is a container to which annotation segments can be added. An annotation segment has a defined start and endpoint on the time line and contains the annotation value. For example, the annotation track *Emotions* is intended to store annotation segments in a timely order. Each segment contains an emotion description (e.g., laughing, crying, etc.) of humans that appear in this media. In SSI the assignment of emotions to segments on the time line has to be done manually.
- *Segment and reorder segments on the time line.* This includes finding the correct start- and endpoint for an annotation segment. Segments can also be shifted on the time line. Moreover, the user selects/deselects the segments she deals with.
- *Edit segment annotation values.* Creating or editing an annotation value of a segment differs from tool to tool. It is therefore, important to know if the annotation has to be machine readable (e.g., XML format) or just understandable for humans.

Typically, a manual annotation process consists of these small tasks that are repeated over and over again.

 ISBN: 978-1-61208-117-5

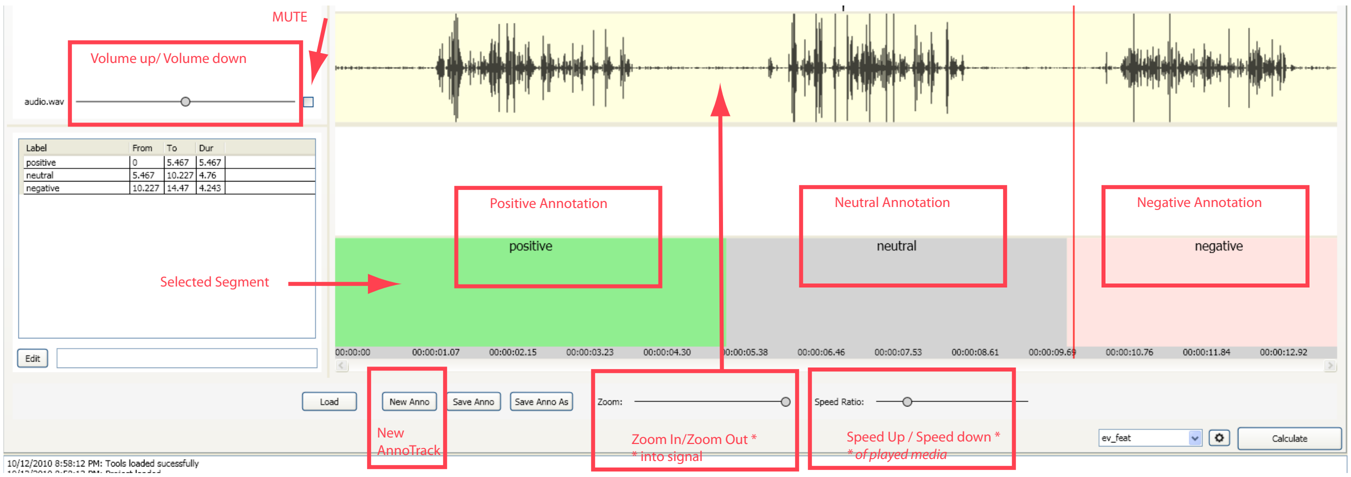


Figure 3. Adapted SSI [9] Graphical UI.

B. Modalities for Interaction

To better understand how interaction can be improved, we studied several modalities with different interaction devices. This way we identified how the inherent strengths and weaknesses of the modalities allow for a robust interaction with media annotation tools.

- *WIMP-based GUI*: This was the existing UI of the SSI tool. A GUI normally supports mouse and keyboard input. The inherent strength of a GUI are the parallel and permanent presentation of content, as well as the *100% recognition rate* (for example you can be sure the user clicked a button). We suggest to use a GUI as the main output modality and as an important input modality.
- *Key binding*: Key bindings are typical short-cuts of keyboard entries that call a certain functionality in the program. This is fairly common in any programming or editing environment. However, it was not included in the SSI toolkit. So we added freely configurable interaction via key bindings. For example, we coupled key *1* to the functionality *set emotional state to positive emotion*.
- *Mouse gestures*: In contrast to normal mouse scrolling and clicking behavior, mouse gestures are like key bindings. We assume that a user that works with an annotation tool often uses mouse and keyboard together. This means that performing a mouse gesture does not require changing the interaction device — which is important for acceptance of a hand-based modality (see Figure 4 for a subset of gestures that we defined). These gesture definitions can also be used for pen-based gestures, WII-mote or other modalities. Such a mouse gesture can be performed within 1-2 seconds and does not require extra skills for anyone who can use a mouse. We used a constrained set of at most 8 gestures for often used commands, because this is a memorable number of gestures.
- *Vocal speech input*: In contrast to the other modalities presented here, speech input does not require an interaction device that is used with one or two hands. Vocal speech input is useful for command-based interaction, like *play*, *add track* and so on.
- *Speech output*: We distinguish between earcons and vocal speech output. Earcons are brief, structured sound patterns that represent an event. They are helpful during media annotation (e.g., an earcon is played if a segment has been added). In contrast, we found that vocal speech output is not helpful in the context of media annotation. First, the serial nature of speech output does not allow to present several tasks in parallel, e.g., to inform a user what she can do next. Second, it hinders a user to concentrate while annotating the media.
- *Jogwheel*: A jogwheel like the shuttleXpress device [4] is a special purpose interaction device that combines buttons with key bindings and scroll wheels. Jogwheels are a good choice for media annotation tools: the issuing of commands via buttons and scrolling on the time line with the wheel(s) is frequently needed.
- *WII-mote*: A WII-mote [6] is a device to perform hand gestures in the air. It is connected via bluetooth to the computer and can be used for a wide number of 3D gestures. However, the process of media annotation requires a lot of precisely given commands that are executed in a chronological order in a relatively short timespan (few seconds between two tasks). People start to get tired just by performing gestures with the WII after a few minutes. So, using a WII-mote did not turn out to be helpful in the process of media annotation — even if the fun-factor is big at the beginning.
- *3D space navigator*: A 3D space navigator [1] is is intended to ease the navigation in a virtual 3D environ-

 ISBN: 978-1-61208-117-5

ment. It also has several buttons like a normal computer mouse. Such a device is mainly useful for CAD-tools or GoogleEarth-like applications. As it is not possible to lock the 3D-maneuverability for 2D scrolling only, using the device is not intuitive compared to the scroll wheel of the jogwheel. We found out that a user often triggered other commands unintentionally, since three axes allow for several possible commands. Therefore, we did not consider the 3D space navigator for the final UI variants.

- *Pen/finger-based gesture*: Pen-based gestures are popular on smart phones and PDAs for navigation and simple commands like starting an application. The media annotation discussed in this work, however, usually involves a desktop or personal computer with a big screen. So, pen-based gestures are out of focus.
- *Multi-touch display*: A multi-touch display is useful for scrolling and zooming in/out, but not for the process of media annotation. Moreover, after a few minutes it gets annoying for the user to have the finger as main input device. Here, we require further studies about satisfaction with multi-touch displays to fulfill longer-lasting work tasks. For this reason we did not consider multi-touch displays for our multimodal UI after studying it for the purpose of media annotation.

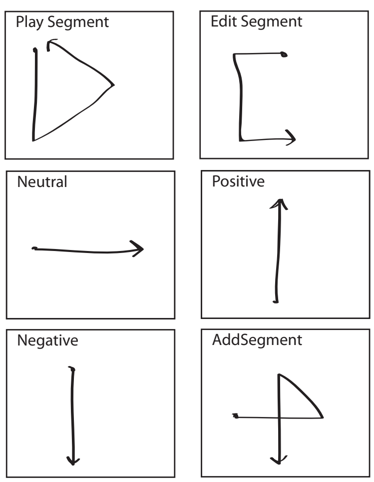


Figure 4. A Set of Gestures Intended for Media Annotation.

C. Prototyping the Multimodal User Interface

We prototyped the multimodal UI with the visual programming platform OI [11]. We carefully assigned the modalities to the defined tasks in order to maximize the advantage of each modality according to Reeves et al. [13]. Furthermore, we only used a constrained grammar set for speech input, to keep the UI consistent. With OI, we defined the interaction of a UI by coupling modalities for human input and output with application components. Besides, *helper* components like data logging units can be included.

In Figure 5 we show the visual programming for one variant of our multimodal UIs. Figure 5 depicts the coupling of our multimodal UI with speech input, mouse gesture and the adapted SSI component. An OI component is either used to forward and transform signals, like the multicast component, a modality component or an application logic component. An OI modality component is an interface to byte code of a modality, e.g., to an external speech input toolkit. We implemented the components SSIOI, LoggerOI, JuliusOI and the according coupling between the components. We executed the interaction pipeline presented in Figure 5 to start the SSI application and its UI with OI.

The *SSIOI* component on the right side of 5 is the interface shown in Figure 3. This SSI component starts the SSI core. Additionally, it starts the GUI of SSI, which is not an OI component. The speech input modality component *JuliusOI* provides the coupling to the Julius [5] speech input toolkit. We configured the recognition grammar for Julius at design time. Furthermore, we included the mouse gesture modality. Here, we did not use one gesture component, but coupled several already existing components together to implement the intended functionality. In principle, a mouse gesture is sent to the SSI component when the user presses the right mouse button, moves the mouse in 2D-space according to a defined gesture and releases the button again. This interaction is realized with the components *DirectXMouseComponent*, *IfThenElse*, *OnInputChanged* and the *GestureRecognizer*. DirectXMouse forwards the x/y coordinates and the pressed button status to a multicast component. The multicast component sends this data to the GestureRecognizer and the IfThenElse component at the same time. The IfThenElse component forwards the status of the mouse button to the OnInputChanged component. The OnInputChanged component sends a continuous signal to the GestureRecognizer, indicating that the right mouse button is pressed. As long as this signal is active, the GestureRecognizer records the gesture. If the right mouse button is released, OnInputChanged sends a signal to stop the gesture recognition, and the GestureRecognizer interprets the recorded data and forwards the best matching gesture to the SSI component.

IV. LESSONS LEARNED

The prototyping of media annotation tools with OI has benefits as well as pitfalls. In this section we will describe those that we encountered during our case study and how they can be generalized for similar projects.

A. Proposed UI variants

We suggest to use the GUI with key bindings, speech input and mouse gestures as modalities of the prototyped UI. Alternatively, if a jogwheel is available , we suggest a variant with jogwheel, keyboard and GUI. The jogwheel is in this context a very interesting device, due to its strength

 ISBN: 978-1-61208-117-5

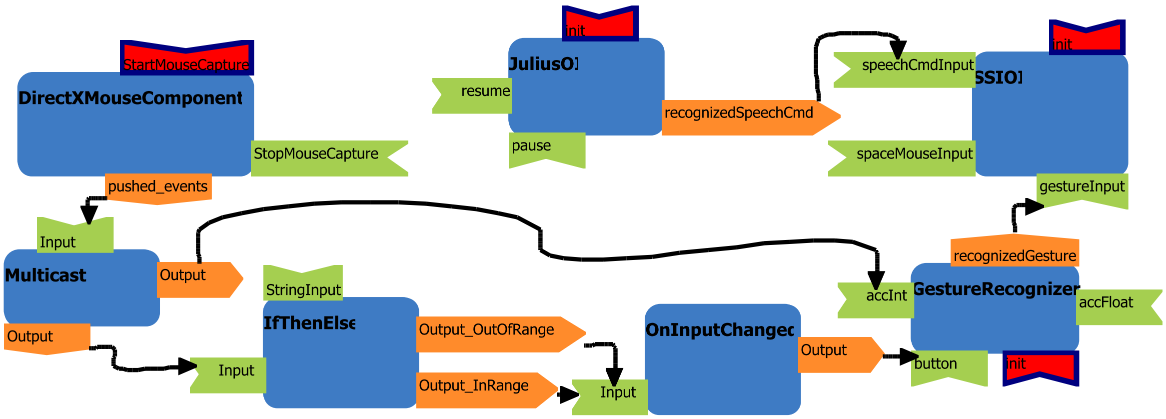


Figure 5. Building a Multimodal UI for Media Annotation with OI.

for scrolling and button-bindings. Both variants should have earcons as output to confirm typical commands.

B. *Benefits*

The open source tool SSI comes with the benefit of relying on existing and tested functionality. Thus we were able to focus on the interaction with this tool. An open source tool may be helpful for any designer, if she wants to define interaction for an application logic that is still under development or only planned to be developed. The open source tool, of course, must have a comparable functionality. This is valid not only for media annotation tools, but any other software with a UI as well.

One of the main tasks for a multimodal UI designer is to combine *suitable modalities* for a dedicated task in a *suitable way*. This also includes taking care of not cluttering a workspace with too many interaction devices. OI is a good support for the designer during the modality selection process, because it offers an easy way to define interaction. The way of visual programming of OI is easy to understand and the online repository provides a lot of already implemented components. The development of new OI components, like the JuliusOI component that we needed, can be completed without much effort by simply following the instructions in the OI documentation. Moreover, OI supports Java and C++, which are two established programming languages.

Another benefit our approach is that the UI designer can change a UI prototype and fine tune it for an improved interaction without much effort. Finally, the stand-alone multimodal UI can be implemented based on the prototyped interaction. The interaction has then already been tested and evaluated with the prototype, thus saving time and money because the prototyping approach is faster than starting from scratch with the *real* UI code.

C. *Drawbacks and Pitfalls*

Since OI is embedded in an Eclipse-based environment it is for prototyping purposes only. This means that the UI prototyped with OI can only be used when started from the OI platform. This drawback is mitigated by the fact that the final UI can be implemented rapidly, based on the already prototyped and thus tested and evaluated, multimodal interaction.

While working with OI, we found some (minor) shortcomings of the program itself. For example, let us consider the coupled components in Figure 5. There is no support to copy and paste this group of connected components between OI interaction diagrams. Support for copy and paste would be very useful, as it further eases the reusability of coupled components, allowing to create *patterns of interaction*. Another feature that would ease the use of OI would be the opportunity to see the current input data at a component's input. This comes in handy, if several input-pipes are connected to one component. Besides, OI could provide more hints while the designer creates the interaction. Such hints could be how to clone an output for the multicast component for example. Even if these are not major issues, they cost the designer some time when defining the interaction.

V. Related Work

Previous work of Oviatt et al. [12] points out that building a multimodal UI is not just connecting several modalities together. This leads to common misconceptions of multimodality. Thoughtless coupling of modalities can be rather counterproductive as it means a multimodal UI may be ineffective or disadvantageous. Instead, the modalities have to be combined carefully, considering which modality to choose for which purpose. In many cases where modalities are combined to a multimodal UI, a straightforward addition of modalities is a good way to couple their different expressivenesses and reduce their inherent drawbacks. Such a predictable creation process does not lead to mysterious properties and totally unpredictable effects of the resulting multimodal UI [7]. Reeves et al. [13] define

 ISBN: 978-1-61208-117-5

six main categories of guidelines for multimodal UI design. These are requirement specification, designing multimodal input and output, adaptivity, consistency, feedback and error prevention/handling. Multimodal UIs have two important goals: to achieve an interaction closer to human-human communication and to increase robustness of the interaction by using redundant and complementary information.

Previous work of Bigbee et al. [15] discusses analytical annotation tools with multimodal interfaces. They present dedicated modalities for next-generation tools like ink (pen-gesture) and eye-gaze tracking that reduce workload in time-consuming and frequent tasks of an annotation process. In our work, we considered their results and enhanced an already existing annotation tool that comes with a GUI with new input functionalities. However, we did not use the modality eye-gaze tracking (not really widespread), but the modality ink.

Related work of Reidsma et al. [14] defines design guidelines for focused and efficient annotation tools. Two already existing tools are used as examples to show how properties of specific annotation tasks affect the design of specialized tools. Among other evaluation criteria the Audio/Video interface has to offer an easy to use technique for playing audio and video sections. Another important requirement for an annotation tool is to have an extendable architecture. The latter was the key criteria for our work, as we wanted to improve the UI. We used the ideas of their work as a basis for our work.

VI. Conclusion

When a user works with timeline-based media annotation tool, she usually repeats of a lot of small tasks like adding annotation segments, moving them on the timeline or editing their content. In this work, we presented a case study of prototyping a multimodal UI for a media annotation tool where we focused on the interaction of such small tasks. In particular, we used the rapid prototyping platform OI to define variants of multimodal UIs where we improved interaction with the open-source media annotation tool SSI. We suggest two variants of such a UI: One with GUI, speech input, earcons and mouse gestures; another with GUI, earcons, speech input and a jogwheel. Moreover, we point out that prototyping a multimodal UI for media annotation tools supports the process of UI generation and leads to robust interaction.

Further research, however, is needed to perform user studies about the usability of the proposed interfaces with focus groups and to further improve the robustness of the multimodal UIs for annotation purpose.

Acknowledgment

We would like to gratefully thank Lionel Lawson, Johannes Wagner, Christian Frisson and Hermann Kaindl for their support in the course of this work.

References

[1] 3d space navigator: Last visited on December 10, 2010. http://www.3dconnexion.com/index.php.

[2] Anvil: Last visited on December 10, 2010. http://www.anvil-software.de.

[3] Elan: Last visited on December 10, 2010. http://www.lat-mpi.eu/tools/elan.

[4] Jogwheel: Last visited on December 10, 2010. http://retail.contourdesign.com/?/products/5.

[5] Julius speech recognizer: Last visited on December 10, 2010. http://julius.sourceforge.jp/en_index.php.

[6] Wii mote: Last visited on December 10, 2010. http://www.mynintendo.de/wii-mote-controller.

[7] N. O. Bernsen. *Multimodality Theory*, pages 5–29. Springer, Berlin-Heidelberg, 2008.

[8] F. J. Johannes Wagner, Elisabeth André. Smart sensor integration: A framework for multimodal emotion recognition in real-time. In *Proceedings of the Fourth International Conference Affective Computing and Intelligent Interaction (ACII 2009)*, 2009.

[9] M. K. D. L. Johannes Wagner, Elisabeth André. SSI/ModelUI - A Tool for the Acquisition and Annotation of Human Generated Signals. In *Proceedings of 36. Jahrestagung fur Akustik (DAGA 2010)*, 2010.

[10] V. Lavrenko, S. Feng, and R. Manmatha. Statistical models for automatic video annotation and retrieval. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, volume 3, pages iii–1044–7, May 2004.

[11] J.-Y. L. Lawson, A.-A. Al-Akkad, J. Vanderdonckt, and B. Macq. An open source workbench for prototyping multimodal interactions based on off-the-shelf heterogeneous components. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems (EICS '09)*, pages 245–254, New York, NY, USA, 2009. ACM.

[12] S. Oviatt. Ten myths of multimodal interaction. *Commun. ACM*, 42(11):74–81, 1999.

[13] L. M. Reeves, J. Lai, J. A. Larson, S. Oviatt, T. S. Balaji, S. Buisine, P. Collings, P. Cohen, B. Kraal, J.-C. Martin, M. McTear, T. Raman, K. M. Stanney, H. Su, and Q. Y. Wang. Guidelines for multimodal user interface design. *Commun. ACM*, 47(1):57–59, 2004.

[14] D. Reidsma and D. Hofs. Designing annotation tools based on properties of annotation problems. In *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research*. Technology, 2004.

[15] L. H. Tony Bigbee, Dan Loehr. Emerging requirements for multi-modal annotation and analysis tools. In *Proceedings in the 7th European Conference on Speech Communication and Technology (EUROSPEECH-2001)*, pages 1533–1536, Aalborg, Denmark, 2001. ISCA Archive.

Interaction Patterns for Designing Visual Feedback in Secure Websites

Jaime Muñoz-Arteaga
Univ. Autónoma de Aguascalientes
Av. Universidad #940. C.P. 20131
Aguascalientes, México
jmauaa@gmail.com

Eduardo B. Fernandez
Dept. of Comp. Science and Eng.
Florida Atlantic University,
Boca Raton, FL, USA
ed@cse.fau.edu

René Santaolaya
CENIDET, Internado Palmira s/n.
Col. Palmira C.P. 62490, Cuernavaca,
Morelos, México
rene@cenidet.edu.mx

Abstract — **In a website, it is essential to offer accessible and secure online services for end users. In general, usefulness and usability aspects are taken into account during design of website, but security issues normally are put aside. The specification of visual feedback helps the analysis and design of websites. This paper proposes a set of best practices of visual feedback for designing websites where the user task can be made secure and usable.**

Keywords-secure website; interaction patterns; visual feedback; software architectures

I. INTRODUCTION

Online services, such as a bank transfer or a virtual meeting, must be executed in a secure environment. In fact, user tasks are constantly exposed to threats either in a simple or complex online service. Here are some examples of threats:

- **Guessing threat**: Here the intruder tries to guess the password that protects the computer network in order to gain access to it.
- **Spoofing threat**: The goal of this attack is to usurp an authorized IP address and to gain access to the victim's system. The IP spoofing attack is often called blind spoofing, and is using against communication services taking advantage of their security vulnerabilities (e.g., rsh, rlogin, and rcp attacks). This allows the intruder to hide the origin of his attack (used in Denial Service attacks). Denial of service attacks typically involve an attacker disabling or rendering inaccessible a network-based information resource.
- **Scanning threat**: The intruder goes about scanning different ports of the victim's system to find some vulnerable points from where they can launch other attacks, (e.g., port-scan). The scanning and the spoofing attacks may be consider more risky, because usually are the preface for other attacks.

A large variety of design techniques for specifying websites exist but they have limited consideration of security aspects [3]. In general, usefulness and usability aspects are taken into account during design of websites but security issues normally are put aside. We consider that the secure aspects of a website can be specified in an explicit manner through its Graphical User Interface (GUI) in order to offer the end user more secure, reliable, and comprehensible online services. In a GUI, it is possible to use some metaphors and colors to notify the end user about detected threats.

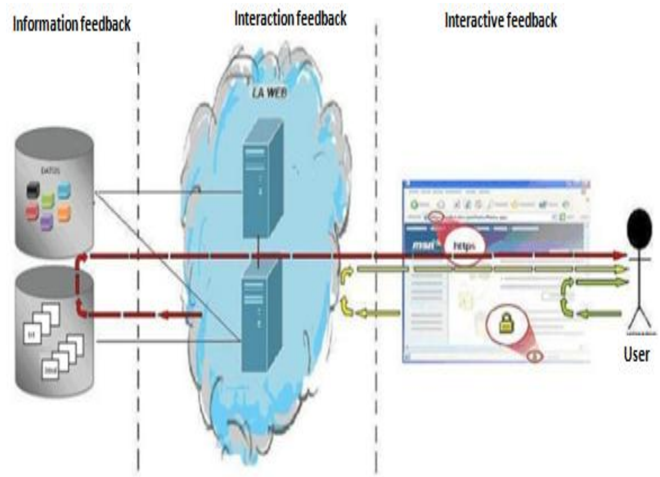


Figure 1. Visual feedback levels in a website.

A website could be considered as a kind of interactive system. The term visual feedback in an interactive system is applied to any graphic form of communication from the system towards the user [9]. Visual, auditory and kinesthetic are different kind of feedbacks that could be used by the system in order offer an easy, comprehensible and usable way to perform user tasks. Visual feedback is predominant in current interactive systems; it can come from different sources in a website. This feedback can be classified in three different categories as in [9]: *information*, *interaction* and *interactivity* levels (see Figure 1). The first category of visual feedback displays the status and the digital content stored in the system. The interaction feedback shows the state of services as available/unavailable. Finally, the interactive feedback is notified immediately to the user with information related to the management of input and output devices used by the user to perform his task. Note that different feedbacks could be closely related, for instance a bank transfer requires multiple visual feedbacks with information about the client's account and details of the service given to the user.

Since visual feedback is predominant in current websites, we consider that the visual feedback could be a meaningful mean to provide security information and to improve security and productivity for a user task. This

 ISBN: 978-1-61208-117-5

paper proposes a set of best practices to assist the user about the security features in a website using visual feedback. For this purpose, this article shows in section two a more detailed analysis of the problem. In section three, we propose a solution in terms of a classification of interaction patterns [5][6] to design the visual feedback to assist the users to make their tasks to be carried out securely. Section four describes several examples of the interaction design patterns proposed in previous sections. Before the conclusion, section five discuses some related work.

II. OUTLINE OF THE PROBLEM

A user could lose control when an online service is requested, for example the exchange of personal data, purchasing and payment of electronic products, or downloading files via Internet banking. Even if users request the services of a website, they may consider it an insecure system, and therefore certain actions generate mistrust and doubt, and even more so when the user is unaware of what security measures are provided by the websites.

When the user interacts with a website, he expects that the GUI gives answers to any of his questions, such as: *What is happening in the system?, why this object is displayed on the screen?, Where am I? And what I can do?*. But how to give a user visual feedback on the security measures of websites?. A good feedback would allow the user to prevent or correct an error caused by a malicious attack.

During the design and development of usability and security aspects for a website we need to take into account:

- Usability aspects are frequently treated in isolation of security aspects [4].
- Taking into account usability and security is considered as a tradeoff for the development team [1].
- It is necessary to use some specification techniques in order to deal with different abstraction levels and diverse perspectives [3].
- In general, formal specification techniques do not address the issues of erroneous behavior of an interactive system, which may have serious consequences for the system and user tasks [2].
- The usage of security aspects is frequently ambiguous for the user. Therefore, the user needs guidance to apply such aspects.

The next sections describe some solutions to the aforementioned problems.

III. INTERACTION PATTERNS FOR DESIGNING VISUAL FEEDBACK

It is quite difficult to develop the external aspect of a system without being immediately stuck into the inherent relation with the internal aspects of system. Taking into account the external aspect during the development of interactive system, it is necessary to work with the presentation, as well as the internal function of system. In addition, a reliable interactive system is not useful for a final user if it is not easy to use, then the security and usability are two significant characteristics in an interactive system. For example the usability flaws of identity management are complex, the structural part require careful thought and redesign of entire systems and standards to fix, but some of these aspects are closely related to a well-designed UI. We point the lack of tools that help identity management systems' developers to mitigate most of the design-challenges particularly those certainly related to the design process of UI's.

It is very important that visual feedback should be displayed through a well-designed user interface. For this goal, current work propose the use of interaction design patterns [5][6], they represent a solution given a recurring problem designing a GUI within a specific context. In addition, the specification of a pattern can communicate the experience and knowledge in a certain domain [5][7].

A good alternative to generating a well designed of security feedback consists of applying interactive design patterns, because it is well known that a pattern represents a proven solution for a recurrent problem within a certain environment. From a computer science perspective, Human-Computer Interaction (HCI) deals with the interaction between one or more users and one or more computers using the GUI of a program [9]. The concepts of traditional HCI can be used to design the interface or improve some interface currently available, considering aspects such as usability. Usability determines the ease of use of a specific technology, the level of effectiveness of the technology, and the satisfaction of the user with the results obtained by using a specific technology to perform specific tasks [9].The interaction design patterns proposed here are classified according to three categories of visual feedback offered by a website.

Informative feedback category: Here is included all information to notify users about available security features, the correct way to use these features, detection of malicious attacks and the internal status of the system.

Interaction feedback category: This category brings together the interaction forms useful to establish the navigation in the windows' interface. In the same way are included the communication forms for the enable or disable of security features, and also, interaction forms to present suggestions of actions to follow when some threat is detected.

Interaction feedback category: This category includes the interactive patterns to specify the security feedback needed to convey information to the end user when the elements of the interface are handled by means of mouse or keyboard.

The interactive patterns of three previous categories form a pattern language (see Table 1); they could be applied to solve the security issues according to the linguistic nature of dialog between the user and an interactive system.

 ISBN: 978-1-61208-117-5

TABLE 1. CLASSIFICATION OF INTERACTIVE PATTERNS

Interaction design patterns for Secure Website	**Informative feedback**	Guessing login
		Accurate Information
		State of secure website
		Protection of personal data
	Interaction feedback	Identification of secure websites
		Restriction of websites for adults
		Activation of online services
	Interactive feedback	Contextual secure feedback.
		Secure website with icons
		Warning with input devices pointer

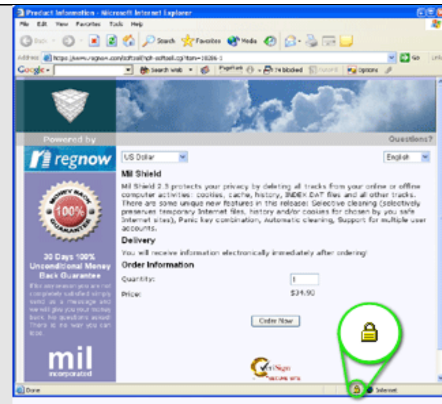
The collection of interaction patterns for secure websites is not exhaustive, someone interested in security and usability aspects could update with new interaction patterns. One of the objective, it is to offer to designer a high level description of visual feedback of different software modules of an interactive system independent of any graphical environment.

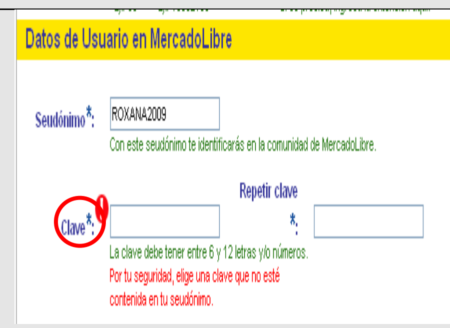
IV. APPLICATION OF INTERACTIVE PATTERNS

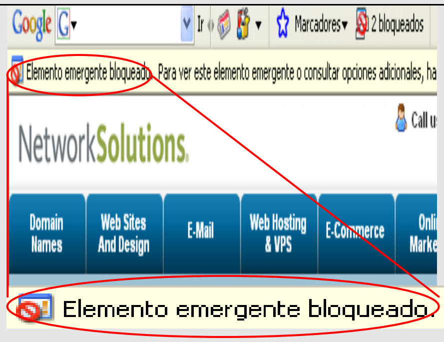
This section describes in detail every category of interaction design patterns proposed in Table 1.

A. Information Feedback

This group of interaction design patterns describes some solutions to display the information in a secure manner. Thus, the designer could offer some guidelines to design visualize the status of security system through preventive or warning messages, it is important to specify that user can decide whether cancel or continue an operation at any moment of interaction. Several component of a GUI could be used, for example a preventive messages, an action buttons and/or give some links with more detailed information.

Name	Accurate information
Problem	User doesn't know if the information obtained from a website is secure.
Context	When user require private information provided by a website.
Force	Provide secure facilitates to get information provided by a website.
Solution	With the information required by user, display some security certificate or icons as part of graphical UI.
Example	 In this example we can see the contents of a e-commerce website by a secure channel of communication. A padlock icon is showed within the GUI.
Consequence	Properly security symbols used in a website make feel user secure about the authenticity of the received information.

Name	Guessing login
Problem	User require a new password to access new services
Context	When a user creates a new account on a website.
Usability Principle	The system could help user tasks to be developed in a secure manner
Solution	If the key provided by user is vulnerable to attack, website should alert and advise the user to change the password.
Example	 Visual feedback of a e-commerce website where the user find advises for getting valid keys in order to create a shopping list online.
Consequence	Give facilities for a better comprehension of user task
Name	Blocking of malicious access
Problem	A website displays a security message but they are specified in terms of internal

	operations.
Context	A secured website should maintain communication with user to set out the actions that the system exerts on the basis of their safety
Force	Display information explicitly to users about current security state of website
Solution	Report clearly and simply the user about processes running the system internally to maintain their security, whether through images, text messages or sounds
Example	 The browser has blocked some facilities of website site in order to protect user task, giving information about the option to activate such facility,.
Consequence	User feels protected by the system.

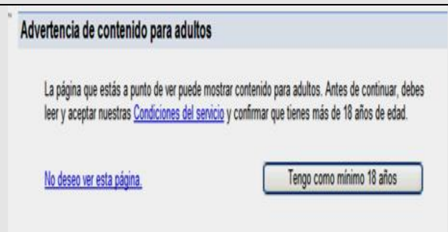
B. Interaction Feedback

The objective of this kind of feedback is display the evolution of communication state between the user and system. This category also includes feedback to the user through the navigation between application windows and the activation of buttons and/or menus for display valid actions.

Name	**Identification of secure websites.**
Problem	User doesn't have any information about the security of current website.
Context	Confidential websites that provide safety information and service online
Force	Notify the security provided by the site
Solution	Show clearly and non confusing information about different security mechanisms provided of current web.
Example	 In this example we can see how the interface is notified by using a secure (https) protocol http. With the lock is notified of the secure connection using SSL certificates.
Consequence	Access to secure online services

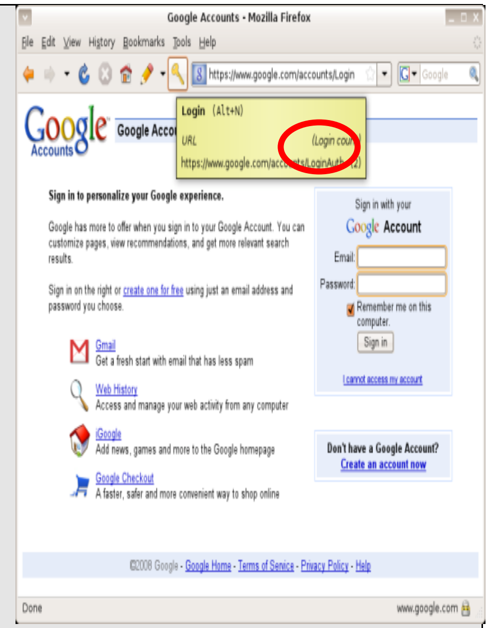
Name	**Navigation on limited areas**
Problem	User feel insecure every time navigate on Internet with reduced space.
Context	Website where the safety information required offering deployed in large quantities and screen space is limited for example in mobile devices.
Force	Help the user to reinforce the security of a website
Solution	Allow the user to view information in several logical drives such as windows, dialog boxes, lists etc deployment., In order to facilitate the exploration of safety information either through direct or sequential navigation
Example	 The example shows how a website helps the user with a dialog box, it make easy to identify and activate secure online services.
Consequence	Navigation is better and allows user a better access to the content of website.

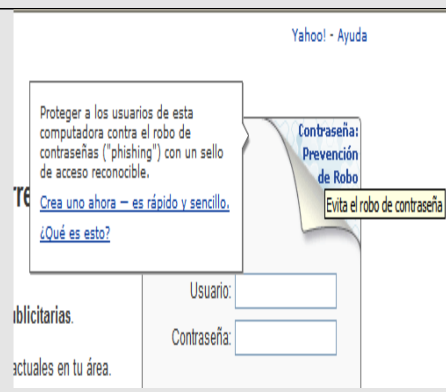
 ISBN: 978-1-61208-117-5

Name	**Restriction of websites for adults**
Problem	How does user prevent to access a website with inappropriate content?
Context	Parental control in order to avoid lose control of kid access of website
Force	User protection.
Solution	Use a warning message (before entering the site) to notify inappropriate content. In case of adult users give the opportunity decide whether cancel or continue to access website.
Example	 This message informs user is trying to enter a website properly for adult. Note that the message gives the opportunity to enter next web page.
Consequence	Share the responsibilities with the system

C. *Interactive Feedback*

Graphical user interfaces of current website increasingly adopt a direct-manipulation style of interaction [8], they give the end user the illusion of directly acting on the objects of interest rather than indirectly accessing them through command buttons and data-entry widgets. Direct manipulation style require immediate feedback and contextual feedback, the management of these feedbacks calls for a much finer grained dialogue modeling than is required for conventional indirect manipulation interfaces. These best practices are taken into account to solve some security issues of website and they are encapsulated in the interactive design patterns.

Name	**Secure website with icons**
Problem	Lack of secure information about user actions
Context	When user action is affected by security issues
Force	Provide safety facilities through user actions.
Solution	Use icons and mouse pointer to describe the confidentiality of information or services online. This feedback is shown when the mouse pointer passes over the item of interest to user and could disappear out of icon.
Example	 Here, the visual feedback provides information about the definition of password when user passes the mouse pointer over the icon with a key.
Consequence	User actions allow learning about solution of security issues in a website.

Name	**Contextual secure feedback.**
Problem	User doesn't have detailed security information in a website
Context	In sensitive text or transactions that provide safety information to user
Force	Use direct interaction style to solve security issues.
Solution	Show security capsules in a website when the user pass the pointer over a meaningful text or objects of a UI.
Example	 In the above example the user informed about preventing password theft through the deployment of information when the

	user passes the mouse pointer over the object composing UI.
Consequence	The user will be informed through their own actions on the site.

V DISCUSSION

In the literature of security engineering [12], several works have been proposed a large number of security patterns regrouped in catalogue [10] [11]. These catalogues cover several aspects of security in order to build reliable software, but visual feedback is not taken into account. Braz et al. [1] have started to take into account the usability with secure factors. In a similar way, current work make emphasis in the visual feedback as a mean at design level for a better understand and comprehension of security issues of a web site. The contribution consists of a set of design patterns to design usable information security feedback combining the concept of user interface patterns [6] and security patterns [10].

VI. CONCLUSIONS AND FUTURE WORKS

This paper proposed a collection of interaction patterns as a specification technique for designing feedback for secure websites with a particular emphasis on visual feedback. The visual feedback can come from different sources in a website to assure the user's task: The first category of visual feedback of information is the group of security patterns that describe in a coherent and continuous way the state in which the user will find the website or any of its internal processes required in a transaction. Visual feedback at the level of user interaction indicates the state of services as available or unavailable to him. The interactive visual feedback captures best practices to assure user actions in detail.

Finally, several aspects could be considered as future work, one of them is the specification of interaction design patterns based on different kinds of feedback such as visual, auditory and kinesthetic.

ACKNOWLEDGMENT

This work was performed thanks to the support from Conacyt (Code Number. 21733), as well as from Promep project No. UAA-CA-48.

REFERENCES

[1] B. Christina, S. Ahmed, and M. David, Designing a Trade-off between Usability and Security: A Metrics Based-Model. Proc. of 11th IFIP TC 13 Conf. on Human-Computer Interaction INTERACT'2007. LNCS Springer, vol. 4663, 2007, pp 114–126.

[2] R. Robert, K. Claire Marie, K. Jhon, and B. Carolyn, Usability Challenges in Security and Privacy Policy-Authoring Interfaces. In: Proc. of 11th IFIP TC 13 Conf. on Human-Computer Interaction INTERACT'2007. LNCS Springer, vol. 4663, 2007, pp. 141–155.

[3] R. Jennifer, J. Carolina, D. Paul, S. Roberto, and N. Kari, G. David, R. Jie, D. Paul, and R. David, Seeing Further: Extending Visualization as a Basis for Usable Security. Proc. of 2nd ACM Symposium on Usable Privacy and Security, SOUPS'2006 ACM Press, 2006, pp. 145–155

[4] V. Roth and T. Turner, User Studies on Security: Good vs. Perfect. In: Proc. of ACM CHI'2007, Workshop on Security User Studies, 2007.

[5] B. Sergy, F. Peter, and Y. Anatoli, User Interface Design Patterns for Interactive Modeling. DVS-IS 2002, LNCS 2545, 2002, pp. 148-158.

[6] V. Martijn and H. Trćtteberg, Interaction Patterns in User Interfaces. Proc. Seventh Pattern Languages of Programs Conference, PLOP 2000.

[7] G. Erich, H. Richard, J. Ralph, and V. Johnson, Design Patterns : Elements of Reusable Object-Oriented Software, Addison-Wesley Professional Computing, 1995.

[8] S. Ben and P.Cathrine, Designing the User Interface: strategies for effective human-computer interaction, Addison-Wesley, 1987. Fourth edition with Catherine Plaisant as co-author 2004, ISBN 0-321-26978-0.

[9] D. Alan, F. Janet, A. Gregory, and B. Russell, Human-Computer Interaction. Prentice Hall, 2004. ISBN 0-13-046109-1

[10] S. Marcus, F. Eduardo, H. Duane, B. Frank, and S. Peter, Security Patterns: integrating security and systems engineering. John Wiley and Sons, 2005.

[11] M. Hafiz, A. Paul, and J. Ralph, Organizing security patterns. IEEE Software, vol 24, 2007, pp. 52-60.

[12] S. Markus, Security Engineering with Patterns: Origins Theoretical Models, and New Applications Editorial Springer ; 2003, LNCS275, ISBN-10: 3540407316

Back-to-Back: A Novel Approach for Real Time 3D Hand Gesture Interaction

Mingming Fan and Yuanchun Shi
Key Laboratory of Pervasive Computing, Ministry of Education
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China
fmm08@mails.tsinghua.edu.cn, shiyc@tsinghua.edu.cn

***Abstract*—In this paper, we present Back-to-Back, a novel real time hand gesture interface for 3D interaction based on double *cameras*. Back-to-Back dexterously makes use of the geometric complement of two back-to-back cameras. Held in hand, Back-to-Back could deduce hand's 3D motion in real time. The basic idea is to extract good corner points from the image sequences captured by two cameras separately and track them while moving. By comparing the motions of two groups of points, the hand's translation and rotation could be deduced accurately as well as other motion parameters. Back-to-Back is a prototype for gestural interaction on mobile devices equipped with two cameras. To further demonstrate its usability, we then analyze the requirements of 3D navigation task and design a strategy to navigate in 3D Space naturally by using Back-to-Back.**

Keywords-Double Cameras, Hand Gesture, 3D Interaction, Natural User Interface, Real Time Interaction.

I. INTRODUCTION

Gesture interaction has been researching for many years. However, robust user's hand motion recognition is still far from enough for real time interaction, especially in unconstrained environment without auxiliary stuff supporting (e.g., color marks; moving in the view field of certain cameras). Instead of designing a set of predefined gestures (e.g., throw, tilt, drawing certain shapes), our approach directly recognizes hand's 3D motions in real time and then uses them for gesture interaction.

Generally speaking, the relative research could be divided into three categories based on the sensors used, which are vision-based approach, accelerometer-based approach and magnet-based approach. The previous researches have restrictions to certain degrees. First, accelerometer based methods are good choices when the aimed gestures have obvious acceleration, such as playing tennis by wii remote [3]. Although accelerometer- based methods could calculate the speed from the acceleration by integral, they could hardly get rid of the drifting errors of integral. What's more, accelerometer-based gesture recognizers usually need to collect a large set of training samples so as to utilize statistical methods, e.g., Hidden Markov Model (HMM), to recognize predefined gestures. Although the Dynamic Time Wrapping (DTW) method [6] needs less training data, it still requires a training phrase. However, using predefined gestures trained from a large amount of samples could hardly meet user-dependent interaction in most cases. As for the camera based interaction, most researches make use of the computer vision algorithms, such as optical flow [1], image differencing, correlation of blocks [9] or corner points tracking [8,10]. Unfortunately, these methods could only detect parts of the 6 Degree of Freedoms (DOF), e.g., translation, rotation, if without the help of auxiliary stuffs, e.g., buttons or color marks. Microsoft Xbox uses cameras and other sensors to detect human body motions in real time for game. However, it requires users stay in the view field of the cameras which restricts the application scenario. Wearing a magnet ring in finger [5], we could interact with devices by simple gestures. However, currently it is only used to recognize predefined simple gestures. In all, making use of hand gestures for real time interaction by cheap, easy-access device is still a challenge.

Back-to-Back is a novel approach to detect 6DOF of hand's motion, which includes translation and rotation by x, y, z axes, by binding two cameras back-to-back (Figure 1). By simply recognizing and comparing the moving directions of the two cameras, Back-to-Back could deduce the current hand's 3D motion without complex computing which is vital for real time interaction especially for mobile devices. What's more, Back-to-Back is cheap and easy to access.

Figure 1. The Prototype of Back-to-Back. It consists of back-to-back cameras. Their unique geometric relationship is the key point we utilize.

As it only utilizes simply computer vision and geometric algorithms, Back-to-Back is highly efficient and real time. Back-to-Back just employs users' natural hand movements and requires no extra learning efforts. Besides, Back-to-Back enables remote control since users do not have to sit at a computer and they could even stand up and walk around while using Back-to-Back. Currently mobile phone tends to equip with two cameras at both the front and back sides, therefore we have tested our idea on NOKIA Symbian S60 3rd phone. Due to current operating system's limitation on support calling two cameras simultaneously, we decide to create our own hardware "back-to-back" to implement and test our idea on Windows OS. We are optimistic about the prospect that our "back-to-back" will be integrated in phone, which will support calling two cameras at the same in the near future, as the 3D gesture interface.

The rest of the paper is organized as: First we will introduce some very close related work and then explain Back-to-Back in detail. Finally, we will give our pilot user study and how to use it efficiently for navigation tasks.

II. Related Work

Camera-based interaction is more suitable for real-time interaction. [8,10] try to detect the hand's motion by detecting and tracking the corner points. The former one uses the PDA which is connected to a PC and all the processing work is done on PC. The method could only detect part of the 3D motion parameters (no rotation around x or y axis). The latter one tries to detect full parameters of hand's 3D motion by using only one handheld camera. However, it tries to distinguish translation and rotation by pressing one key which highly decreases the convenience and naturalness, for users must remember how to press the key while operating.

Fan et al. [4] made a step further. They try to detect hand's 3D motion by only one handheld camera without using auxiliary stuffs. They create several classifiers by analyzing the different geometric characteristics. However, due to the limitation of the one camera, fast translation and low rotation are hard to distinguish with each other.

By analyzing the previous work, we propose Back-to-Back by analyzing the movement directions of two back-to-back cameras. The special structure of two cameras allows us to distinguish translation and rotation simply by comparing motion directions. Back-to-Back is self-contained and needs no calibration. The only thresholds used in Back-to-Back are to filter the noise motion, which is more or less the same while used by different people.

III. Back-to-Back Algorithm

The Figure 2 shows the system flowchart of Back-to-Back. Firstly, Back-to-Back will collect the frames from two cameras and then extract the easy-to-track corner points [7]. Motion Estimation engine will compare the moving patterns of two cameras' corner points. Through this engine, the detected motion parameters will be used to control the Application, like navigation, game, etc.

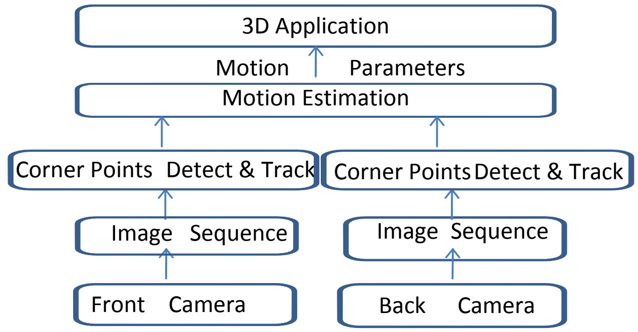


Figure 2. The Flowchat of Back-to-Back Algorithm

A. Corner Points Detection and Tracking

Corner Points have big eigenvalue in the image and are relatively stable to track. The main steps of Corner Points Detection and Tracking are: 1) calculate the minimal eigenvalue for every pixel of the captured image; 2) Perform non-maxima suppression; 3) Reject the Corners Points with the minimal eigenvalue less than a level; 4) Make sure the distances between the corner points larger than a value and remove those too close points; 5) Track the corner points by implementing sparse iterative version of Lucas-Kanade optical flow in pyramids in the consecutive frames [2]. 6) Through the above five steps, we get the corners points' coordinates in the consecutive frames, which will be used in the next step Motion Estimation procedure.

B. Motion Estimation

Motions in 3D space consist of 6 DOFs (translation along and rotation around 3 axes). One challenge facing computer vision research is that certain motions are similar with each other. For example, the captured images both move to the left side while translating camera along X axis to the right side (Left one of Figure 3) and rotate camera by Y axis (perpendicular to the paper plane in Figure 3) to the right side (Right one of Figure 3).

The novel design of setting two cameras back-to-back aims to distinguish these two kinds of movements artfully. As illustrated in Figure3, when Back-to-Back translates to the right side, the images captured by camera ***A*** moves from its Right Side (R) to its Left Side (L). However, the images captured by camera ***B*** moves from its L to R. Therefore, two cameras have different directions of optical flow (Left one in Figure 3). When Back-to-Back rotates in counter-clock around Y axis to the right side, the captured images of both camera A and B both move from their L to R (Right one of Figure 3).

As the Back-to-Back translates up along Y axis, both the captured images of camera A and B will go to the down side of Y axis. However, while rotating around x axis, if the images captured by camera A move to the up side, then the images captured by camera B will move to the down side.

According to the above analysis, translation by x/y axis and rotation by y/x axis can be easily distinguished simply by detecting the directions of optical flows of both cameras.

 ISBN: 978-1-61208-117-5

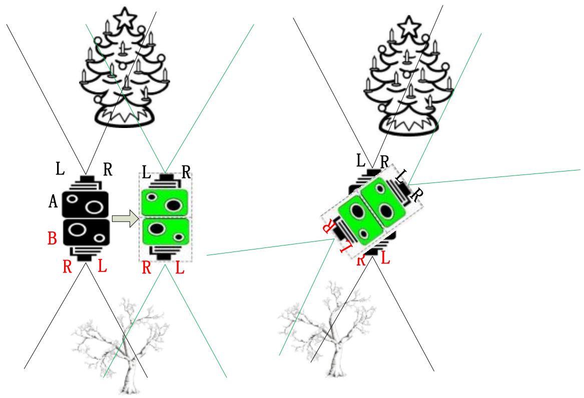


Figure 3. Translation along X axis (Left) and Rotate around Y axis (Right) (Y axis is perpendicular to the paper plane)

Suppose the coordination of the i th corner point in previous frame and current frame are (x_i, y_i), (x'_i, y'_i), then its relative motion is $(x'_i - x_i, y'_i - y_i)$, therefore the relative motion of the camera is $(\frac{1}{N}\sum_{i=1}^{N}(x'_i - x_i), \frac{1}{N}\sum_{i=1}^{N}(y'_i - y_i))$ (N is the number of the tracked corner points). According to the above equation, the motions of two cameras can be calculated and represented as $(\overline{x_f}, \overline{y_f})$, $(\overline{x_b}, \overline{y_b})$. Suppose the parameter rY, rX means rotation around Y, X. ZOOM means translation in Z axis and rZ means rotation around Z axis).

Algorithm 1 Back-to-Back

If $\overline{x_f} * \overline{x_b} < 0$ **and** $|\overline{x_f}| + |\overline{x_b}| >$ *Threshold1*

Then $rY = 0$; *Translate along X axis;*

Else If $\overline{x_f} * \overline{x_b} > 0$ **and** $|\overline{x_f}| + |\overline{x_b}| >$ *Threshold2*

Then $rY = 1$; *Rotate around Y axis;*

If $\overline{y_f} * \overline{y_b} > 0$ **and** $|\overline{y_f}| + |\overline{y_b}| >$ *Threshold3*

Then $rX = 0$; *Translate along Y axis;*

Else If $\overline{y_f} * \overline{y_b} < 0$ **and** $|\overline{y_f}| + |\overline{y_b}| >$ *Threshold4*

Then $rX = 1$; *rotate around X axis;*

$$ZOOM = \frac{1}{N}\sum_{j=1}^{N}\sqrt{(x'_j - \frac{1}{N}\sum_{i=1}^{N}x'_i)^2 + (y'_j - \frac{1}{N}\sum_{i=1}^{N}y'_i)^2} \Big/ \frac{1}{N}\sum_{j=1}^{N}\sqrt{(x_j - \frac{1}{N}\sum_{i=1}^{N}x_i)^2 + (y_j - \frac{1}{N}\sum_{i=1}^{N}y_i)^2}$$

If $ZOOM > 1$, **Then** *zoom out;*

Else If $ZOOM < 1$, **Then** *zoom in;*

$V'_i = (x'_i - O'_x, y'_i - O'_y)$ $V_i = (x_i - O_x, y_i - O_y)$,

$$\theta_i = \arccos(\frac{V'_i - V_i}{|V'_i| * |V_i|})$$

If $rZ = \sum_{i=1}^{N}\theta_i / N >$ *Threshold5*, **Then** *rotate around Z axis;*

(O_x, O_y) is the center of the corner points in the last frame, and (O'_x, O'_y) is the center of corner points in the current frame. The pseudocode of Back-to-Back is listed in Alogirithm1.

The thresholds 1 to 5 are used to filter the noise caused by hand jitter and environment. The specific values are trained from user studies of 11 subjects.

We conducted an informal user study to evaluate the correct rate of Back-to-Back. The test application asks users controlling one cube to match the other one with random position and orientation. Each time it only requires user to do either translation or rotation. The application records the actual operations done by users and later we compare them with the right operations. 11 subjects, nine males and two females, aged from 22 to 26, were hired from local university. They were given about 3 minutes to warm up. The correct rates of 11 subjects are showed in Table 1.

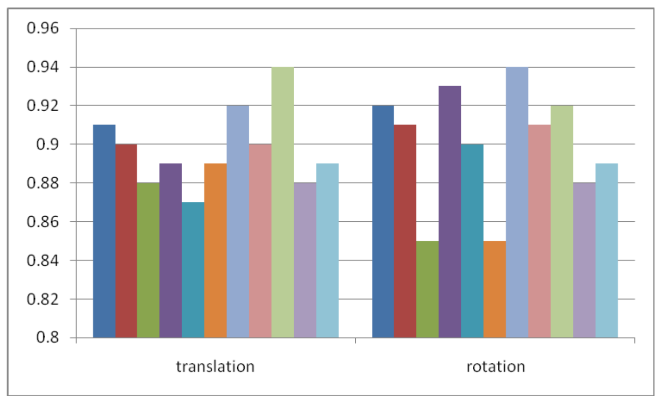


TABLE 1. THE CORRECT RATE OF BACK-TO-BACK

IV. 3D INTERACTION BASED ON BACK-TO-BACK

Back-to-Back could highly distinguish hand's translation and rotation in real time without any requirements of marks or other stylus. Natural gestures could be used in 3D navigation tasks. Users can use the Back-to-Back freely switch between the translation and rotation during navigation.

Keyboard and Mouse operations are highly accurate too. However, it requires users to sit before the computer with two hands operating and navigation in 3D space also requires the coordination of the two hands, however, Back-to-Back can achieve the same goal by one hand.

Previous research tired to employ hand gestures for navigation also requires the key-pressing to switch between the translation and rotation modes which decreases the convenience of operation, because users must keep in their mind when to switch between two kinds of motions (translation and rotation) modes.

Navigating in 3D space has been researched for many years. However, navigating by only applying real time natural hand motion without any other auxiliary stylus is still challenging. The creative design of Back-to-Back gives us a highly accurate approach to distinguish hand's translation and rotation. Translation gesture can be used to control the movement in different directions and rotation gesture can change user's orientation. But there is still another very challenging problem, which is how to avoid the misrecognized motion. Specifically, if only using above algorithm, then our approach will also detect slight zoom while translation or rotation due to effect of the highly dynamic changing background. Therefore a better strategy of distinguishing different motions is needed. A subtle strategy (Algorithm 2) to detect zoom (moving into or out of the 3D space) is described below. Navigation in 3D usually requires continuous moving in one direction (Moving forward), however our hand can only move in a very small area around our body. In order to address this issue, we introduce following strategy: If continuously pushing Back-to-Back forward for a while, system then goes into ZOOM-IN state and we can continuously move into the scene without need of pushing our hand any more. If we want to stop, we just need to pull our hand a little bit back, then system goes into NO-ZOOM state. If we continue to pull back, then system goes into the ZOOM-OUT state.

Algorithm 2 Finite State Automata

States: ZOOM-IN, ZOOM-OUT, NO-ZOOM

Initial: State = NO-ZOOM;

If ZOOM > Threshold6 **Then** Count++;

Else If ZOOM < Threshold7 **Then** Count--;

Else Count = 0;

If State = ZOOM-IN;

If Count < -2 **Then** State = NO-ZOOM; Count = 0;

Else If State = ZOOM-OUT

If Count > 2 **Then** State = NO-ZOOM; Count = 0;

Else If Count > 6 **Then** State = ZOOM-IN;

Else If Count < -6 **Then** State = ZOOM-OUT;

V. 3D NAVIGATION BY BACK-TO-BACK

We demonstrated the usability of Back-to-Back by for navigation in a 3D space. The 3D scene is rendered by DirectX. The first row of Figure 4 shows that when user moves to the left side, the 3D scene also moves to the left. Second row shows that the orientation changes to the right side when user rotates his wrist around Y axis to the right side. Last row shows that when user pushes his hand forward, our viewpoint goes into the space at the same time. By using the above Algorithm 2, users reported that they could easily move forward or backward continually by pushing and pulling their hands. They also pointed out that if they want to rotate a very large angle (e.g. 180), they may not finish this operation in one time. In other words, users need to reset their hand to normal position so as to continue the further rotation. This idea enlightened us to set different states (e.g. Rotating Left State means continually rotating to the left) to satisfy the continual movements requirements (Similar to the Finite State Automata idea).

Figure 4. 3D navigation by using Back-to-Back. The first line shows: when user translates back-to-back to the left, the 3D scene also moves to the left. The second line shows that user rotates camera to the left. The third line shows when user pushes forward, the scene zooms in.

VI. FUTURE WORK

More formal user studies will be conducted to evaluate Back-to-Back's robustness and other performance metrics. We will implement our algorithm on mobile phone, equipped with two cameras. We are also interested in comparing it with phone's other operations: joystick, keyboard and multi-touch.

VII. CONCLUSION

In this paper, we present the design and implementation of a novel 3D hand motion detection algorithm: Back-to-Back. By binding two cameras back-to-back, we can easily detect 6DOF simply by analyzing the moving directions of two cameras. Different from previous approaches, Back-to-Back only use one handheld device to generate 6DOF and does not need complex training phrase. Users could easily manipulate it without any auxiliary stuff or long time pre-learning. We exemplify its usability of Back-to-Back for 3D interaction by discussing how to use it for 3D navigation task.

 ISBN: 978-1-61208-117-5

ACKNOWLEDGMENT

This work is supported by NOKIA and National High-Tech Research and Development Plan of China under Grant No. 2009AA01Z336. We thank all reviewers' careful and insightful comments on both the contents and the format. We also thank our user study's participants for their efforts.

REFERENCES

[1] Ballagas, R., Rohs, M., Sheridan, J., (2005) Sweep and Point and Shoot: Phonecam-Based Interactions for Large Public Displays.. Proceedings of the 23th of the international conference extended abstracts on Human factors in computing systems (CHI EA'05), pp1200-1203

[2] Bouguet, J.V. (1999) "Pyramidal implementation of the Lucas Kanade Feature Tracker Description of the algorithm". Intel Corporation Microprocessor Research Labs.

[3] Castellucci, S., and Mackenzie, I. (2008). Unigest: text entry using three degrees of motion. Proceedings of the 26th of the international conference extended abstracts on Human factors in computing systems (CHI EA'08), pp3549-3554

[4] Fan, M., Zhang, L., Shi. Y. (2008). Hand's 3D Movement Detection with One Handheld Camera. Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST'08), pp255-256.

[5] Harrison,C., and Hudson, S. (2009). Abracadabra: Wireless, High-Precision, and Unpowered Finger Input for Very Small Mobile Devices. Proceedings of ACM Symposium on User Interface Software and Technology (UIST'09), pp121-124.

[6] Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V. (2009). uWave: Accelerometer-based personalized gesture recognition and its applications. Journal of Pervasive and Mobile Computing, Volume 5 Issue 6, December, 2009, pp 657-675

[7] Shi, J., and Tomasi, C. (1994). Good features to track. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR' 94). pp593-600.

[8] Sohn, M., and Lee, G. (2005). ISeeU: Camera-based User Interface for a Handheld Computer. Proceedings of 7th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'05), pp 299-302.

[9] Wang, J., Zhai, S., Canny, J. (2006). Camera Phone Based Motion Sensing: Interaction Techniques, Applications and Performance Study. Proceedings of ACM Symposium on User Interface Software and Technology (UIST'06), pp101-110.

[10] Zhang, L., Fan, M., Shi, Y. (2008) UCam: Direct Manipulation using Handheld Camera for 3D Gesture Interaction. Proceedings of ACM International Conference on Multimedia (MM'08), pp 801-804.

 ISBN: 978-1-61208-117-5

Learning Displacement Experts from Multi-band Images for Face Model Fitting

Christoph Mayer
Intelligent Autonomous Systems Group
Technische Universität München
München, Germany
mayerc@in.tum.de

Bernd Radig
Intelligent Automomous Systems Group
Technische Universität München
München, Germany
radig@in.tum.de

***Abstract*—Models are often used to gain information about real-world objects. Their parameters describe various properties of the modeled object, such as position or deformation. In order to fit the model to a given image, displacement experts serve as an update function on the model parameterization. However, building robust displacement experts is a non-trivial task, especially in real-world environments. We propose a novel approach that learns displacement experts from a multi-band image representation which is specifically tuned towards the task of face model fitting. We provide the fitting algorithm not only the original image but an image representation that reflects the location of several facial components within the face. To demonstrate its capability to work robustly not only in constrained conditions, we integrate the *Labeled Faces In The Wild* database, which consists of images that have been taken outside lab or office environments. Our evaluation demonstrates, that the information provided by this image representation significantly increases the accuracy of the model parameter estimation.**

***Keywords*- face model fitting; computer vision; human-maschine-interaction**

I. Introduction

The interpretation of human face images is a traditional topic in computer vision research. The analysis of human faces provides information about person identity, facial expression or head pose and is the interest of several research groups. Model-based techniques have proven to be a successful method for extracting such high-level information from single images and image sequences. Face models reduce the large amount of image data to a small number of model parameters $\mathbf{p}$. These model parameters describe properties of interest of a single face, such as its position, shape or the visual appearance of its texture. However, in order to allow the extraction of high-level information to work robustly, model parameters that match the content of a single image have to be calculated to ensure that the model corresponds to the image content.

Model fitting is the computational challenge of finding these model configurations. Although a large number of face models and fitting strategies have been proposed, it is still a challenging task in uncontrolled environments when no restrictions towards face pose or lighting and only weak restrictions towards face size or image quality exist. Usually, approaches are evaluated on standard image databases that restrict the image content with respect to background, lighting, head pose or facial expression [19], [20]. Most image data is captured in controlled lab or office environments [12], [13], [14]. Some databases, such as the BioId Database [14] or the VALID database pose less constraints on background and lighting to create realistic image data. However, both databases do not include turned heads and half-profile views.

In contrast, we propose to evaluate face model fitting algorithms in an unconstrained environment. The contributions of this paper are two-fold: Firstly, we integrate the database *Labeled Faces In The Wild* database [18], which contains image material taken in real-world conditions. The images depict persons of public life and include both male and female, spanning a large variety of ethnic backgrounds, age, facial expressions, lighting and image backgrounds. See Figure 1 for some example images on which a face model has been fit automatically. Second, in order to tackle this challenging task, we provide the fitting algorithm not only with the original image data but with multi-band images which highlight facial components like eye brows or lips. These image-bands can be thought of as additional image channels but we restrain from this nomenclature in order to avoid confusions.

This paper continues as follows: Section II provides an overview about related approaches and the scientific background of this paper. Section III introduces our face model and displacement experts as well as our learning approach. Section IV demonstrates results of the experimental evaluation of our approach.

II. Related Work and Background

The interpretation of image data is often arranged in multiple steps with every step relying on the computation results of preprocessing steps. This section discusses related approaches in both important steps of this work: creating

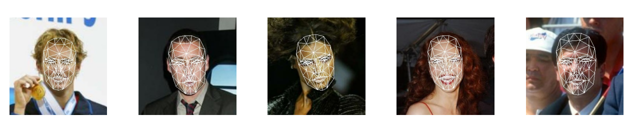


Figure 1. The face model is automatically fitted to every database image.

 ISBN: 978-1-61208-117-5

facial component related multi-band images and model fitting.

A. Computing the Facial Component Feature Bands

Skin color represents an important source of information to various computer vision applications. For a recent survey we refer to Phung et al. [3]. Most approaches identify a subspace of the color space to represent skin color, either by inspecting every single element of the color space or by defining clustering rules [4], [3]. Usually, this mapping is obtained by inspection of a (in some cases very large) number of training images. In contrast, we adapt our model to the specific image content to obtain a more robust estimation of the skin color distribution.

The approach of Hsu et al. is similar to ours such that it expects facial components to be located within an ellipsoid around the skin color area [4]. It constructs eye and mouth maps in order to verify the location of the entire face. A similar approach has recently been published by Beigzahed et al. who manually define rules to construct the eye and mouth maps for determining mouth and eye candidates [5]. We take the reverse approach, since we use an estimation of the face position to determine the location of facial features. Lips classifiers are able to provide useful information for speech recognition, speaker authentication and lip tracking [6], [7].

B. Related Approaches Towards Face Model Fitting

Most fitting strategies fall in one of two categories, utilizing either *objective functions* or *displacement experts*. Objective functions $f(I, \mathbf{p})$ yield a comparable value that determines how accurately a parameterized model $\mathbf{p}$ fits to an image I and are minimized to determine the optimal parameterization $\mathbf{p}_I$. This strategy is either applied on the raw image data with help of a photo-realistic appearance model [8], [10], [11] or rely on the extraction of image features, such as edges and color representation, Haar-like features, or color distribution [9], [21]. The draw-back of these approaches is that the minimization is usually computationally expensive and prone to local noise.

In contrast, displacement experts calculate a parameter update $\Delta\mathbf{p}$ on the initial parameters $\mathbf{p}_0$ to obtain the optimal model parameters $\mathbf{p}_I = \mathbf{p}_0 + \Delta\mathbf{p}$. The drawback of this approach is, that displacement experts can only be accurately learned when the displacement is small. This is only the case if a good initial pose estimate is available. Traditionally, these algorithms are evaluated on a set of well-known databases [12], [13], [14]. The images in these databases are manually annotated with model parameters or single points that reflect the position of eyes, mouth, nose tip or other facial features. The fitting algorithm is presented a subset of these images for training and is evaluated on another subset. Therefore, the fitting algorithm is able to consider only variances that are represented in the images. If, for instance, only frontal view images are included in the data, the algorithm is not able to work on half-profile views.

Cootes et al. [15] propose to utilize images with two feature bands for creating and fitting face appearance models. These feature bands reflect edge directions in two dimensions, where the magnitude indicates the degree of reliance in the orientation estimation. This approach is similar to our approach because not only the raw image data but an image representation with various additional feature bands is considered. Similarly, Stegmann et al. propose to utilize a multi-band image representation [16]. Edges and color bands obtained from converting the image to different color spaces are considered. Kahmaran et al. also take this approach but rely on a different color conversions [17]. In contrast to these approaches, our representation adapts to image conditions and the characteristics of the visible person.

III. Learning Displacement Experts from Multi-band Images

To fit the model and calculate the model parameters $\mathbf{p}_I$ for a single image I, we calculate a parameter update $\Delta\mathbf{p}$ to the current model parameters $\mathbf{p}$ as shown in Equations 1 and 2.

$$\mathbf{p}_I = \Delta\mathbf{p} + \mathbf{p}. \tag{1}$$

$$\Delta\mathbf{p} = \mathbf{g}(I, \mathbf{p}) \tag{2}$$

Since $\mathbf{p}$ is known, the challenge in this approach is the computation of $\Delta\mathbf{p}$. Because this computation has to be performed from the image data I and current model parameters $\mathbf{p}$ only, this requires to obtain robust calculation rules in $\mathbf{g}(I, \mathbf{p})$. To obtain this robustness, our approach provides the displacement expert not only with the original image I but with a multi-band image representation $\mathbf{I}$. The idea is, that since model parameters represent relative positions of facial components in the face, the fitting algorithm benefits from an image representation that specifically highlights these facial components. Although the displacement expert could

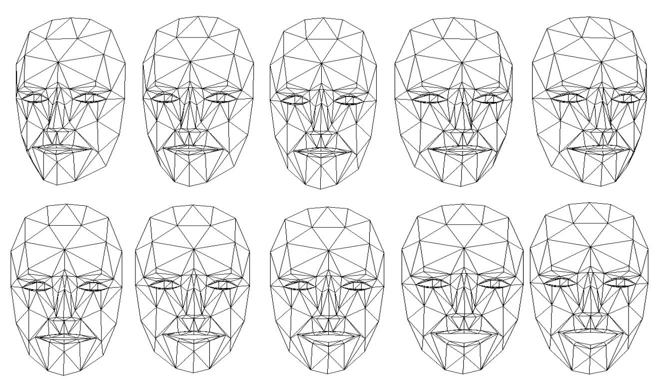


Figure 2. Model parameters change point positions to reflect face pose or shape change

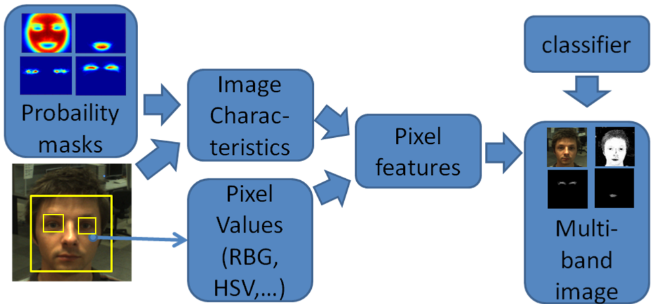


Figure 3. Pixel-based classifiers are trained on adjusted pixel features, that have been adapted to the image content via a set of image characteristics.

be applied to any model parameterization $\mathbf{p}$, we usually apply it to a fixed, initial parameterization $\mathbf{p}_0$:

$$\mathbf{p}_I = \mathbf{g}(\mathbf{I}, \mathbf{p}_0) + \mathbf{p}_0 \tag{3}$$

We apply this fitting approach to the Candide-3 face model, which is a wire-frame model consisting of $K = 116$ anatomical landmarks [2]. The $3K$ dimensional vector $\mathbf{v}$ contains the vertex x-, y- and z-coordinates. The model is controlled by applying the shape deformation $\mathbf{v}_{shape} = Ss + Aa$, a rotation matrix R, a scaling factor c and a translation $\mathbf{t}$ to the basic model structure $\mathbf{v}_{basic}$. The difference between the parameters in s and a is that A contains motion that may appear due to facial expressions whereas S contains vertex motions to adapt the general face structure to the face structure of a specific person. In total, the model vertex coordinates are computes according to Equation 4.

$$\mathbf{v} = cR(\mathbf{v}_{basic} + \mathbf{v}_{shape}) + t. \tag{4}$$

We denote the single vector elements, i.e. single parameters by p_i. The parameter vector is assembled according to Equation 5. Figure 2 depicts some example parameterizations.

$$\mathbf{p} = [c, t^T, r_x, r_y, r_z, a^T, s^T] = [p_1, ..., p_n, ..., p_N] \tag{5}$$

The remainder of this section will show learning displacement experts in two steps: First, the calculation of the multi-band image representation is discussed, and then the procedure to create a displacement expert from this image representation is presented.

A. Computation of Multi-Band Images

We rely on the approach presented by Mayer et al. and refer to their work for a detailed description and evaluation [1]. Still, for the sake of completeness, we will provide a short overview which is visualized in Figure 3. Firstly, characteristics of the current image are calculated that describe its content. Secondly, every pixel is inspected individually and a large amount of pixel features are calculated with help

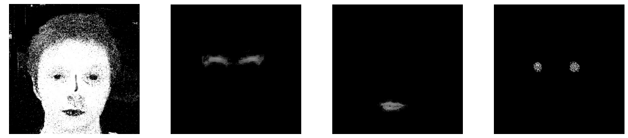


Figure 4. Image bands represent the probability of single pixels to depict certain facial components like skin, eye brows, lips and retinas for a specific image.

of the image characteristics. Since their calculation is based on the image characteristics they are adjusted to the image content and they are called "adjusted pixel features". Finally, pixel-based classifiers are trained on these pixel features that decide for a single pixel whether this pixel depicts a specific facial component or not. Figure 4 depicts examples for four different feature bands. We will inspect every single step in further detail.

The **image characteristics** are calculated by applying a set of object locaters to determine a region of interest (ROI) around the face and the position of the eyes in the image, represented by yellow boxes in Figure 3. The approach of Viola et al. is integrated for this task [22].We generate a set of masks from training images that determine the probability for each pixel within the face bounding box to depict a specific facial component, see Figure 5 for a visualization of some example masks. By aligning the masks to the face bounding box, we map each pixel within the face bounding box to an entry in the probability masks. We denote the pixel by $\mathbf{x}$, its position within the image by $\mathbf{x}_s$ and its color values by $\mathbf{x}_c$. Furthermore, $\mathbf{x}$ is mapped to the mask entry $w^f_\mathbf{x}$ corresponding to the facial component f. We calculate the spatial mean of each facial component according to equation 6 and an estimation of average facial component colors according to equation 7.

$$\mu_f = \sum_{\mathbf{x} \in ROI} w^f_\mathbf{x} \mathbf{x}_s \tag{6}$$

$$\nu_f = \sum_{\mathbf{x} \in ROI} w^f_\mathbf{x} \mathbf{x}_c \tag{7}$$

The **adjusted pixel features** are computed for each pixel separately. They contain the pixel coordinates within the face bounding box and relative to the eye positions. Furthermore, distances $\mathbf{x}_s - \mu_f$ for different facial components are computed in different distance measurement, such as Euclidean distance and Mahalanobis distance. Additionally, the pixel's color relative to the estimated color means are calculated by $\mathbf{x}_c - \nu_f$ with respect to skin color distribution, brow color distribution and lip color distribution.

Training classifiers requires to manually annotate a set of training images with the facial component each pixel depicts. A face locater is applied to determine the face bounding box and the probability matrices are applied to calculate the image characteristics. Although, all pixels in all images' face

 ISBN: 978-1-61208-117-5

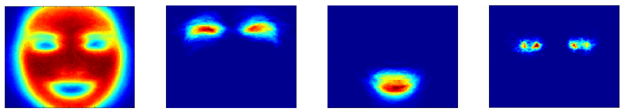


Figure 5. Probability masks are created beforehand to represent the spatial distribution of facial components in the face bounding box. Left to right: skin, eye brows, lips and retinas.

bounding boxes could be used as training data, due to the large amount of data only a subset of them is utilized. For every facial component, a classifier is trained from the extracted data to determine whether a single pixel depicts that specific facial component or not. In this paper we assemble a multi-band image $\mathbf{I} = \{I^{grey}, I^{skin}, I^{brow}, I^{lip}, I^{retina}\}$. Please note again, that this is only a rough overview of this approach. To learn more on how the spatial and color distributions are calculated and which distance measures are applied in detail, we refer to the work of Mayer et al [1].

B. Learning Displacement Experts

Ideally, a displacement expert should always determine the correct parameter update to obtain a perfect model fit $\mathbf{p}_I^\star$. Equation 8 depicts such an ideal displacement expert which simply computes the difference between the ideal model parameterization $\mathbf{p}_I^\star$ and the current model parameterization $\mathbf{p}$. Note, that the ideal model parameterization is usually not known, unless it is specified manually, which prevents this function from being practically applied. However, it will be applied to generate training data to train a further displacement expert $\mathbf{g}^l(\mathbf{I}, \mathbf{p})$ by combining Equation 3 and Equation 8. This displacement expert will be independent of $\mathbf{p}_I^\star$ and instead utilize the image data I as presented in Equation 9.

$$\mathbf{g}^\star(\mathbf{p}_I^\star, \mathbf{p}) = \mathbf{p} - \mathbf{p}_I^\star \tag{8}$$

$$\mathbf{g}^\mathbf{l}(\mathbf{I}, \mathbf{p}_I^\star + \Delta\mathbf{p}) = \Delta\mathbf{p} = \mathbf{g}^\mathbf{l}(\mathbf{I}, \mathbf{p}) \tag{9}$$

The first step is to annotate a set of images with the correct model parameters $\mathbf{p}_I^\star$. Obviously, it is still desirable to have $\mathbf{g}^l(\mathbf{I}, \mathbf{p}_I^\star) = 0$, since no parameter update is required in this case. However, we required training examples that reflect cases, when a parameter update is required. This training data is acquired automatically, by inducing random model parameter variations $\Delta\mathbf{p}$ to obtain new model parameterization $\mathbf{p} = \mathbf{p}_I^\star + \Delta\mathbf{p}$. In these cases the displacement expert is expected to determine the induced parameter variation according to Equation 9. Please see Figure 6 for a visualization of the feature extraction for two example annotations.

Therefore, training data consists of pairs of images and example parameterizations $< \mathbf{I}, \mathbf{p} >$, that are labeled with the induced parameter error $\Delta\mathbf{p}$. The displacement expert will be trained to determine $< \mathbf{I}, \mathbf{p} > \rightarrow \Delta\mathbf{p}$, which allows to fit the model according to Equation 1. In order to simplify

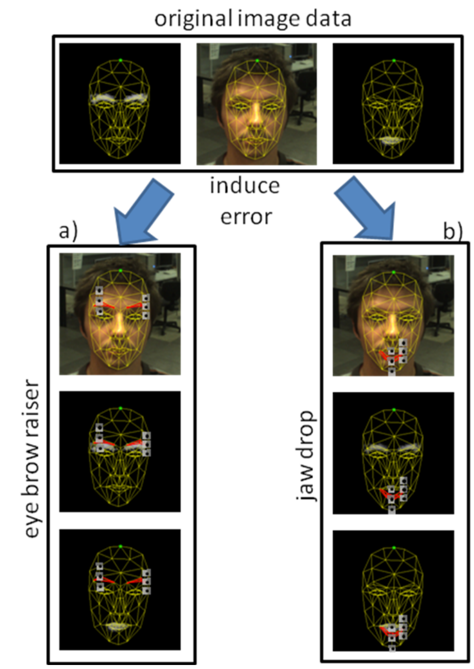


Figure 6. Features are extracted at points influenced by single parameter changes. Additional features are extracted along the direction of point motion from all image bands Due to space limitations, only selected image bands are presented here. Top: image bands of the original image with manually fit model. a) error induced in the eye brow raiser parameter. b) error induced in the jaw drop parameter.

this learning problem, we train single displacement experts for each single parameter separately. Equation 10 formalizes this step.

$$\Delta p_i = g_i^l(\mathbf{I}, \mathbf{p}_I^\star + \Delta p_i) \tag{10}$$

It is usually beneficial in computer vision applications to apply a data reduction to the image data by extracting descriptive image features. In this paper, we extract Haar-like features in different styles and sizes. These features are calculated by summing up pixel intensity values in two image regions and then subtracting one of these sums from the other. These image regions are visualized in light grey and dark grey in Figure 6.

As mentioned, model parameters change the relative positions of model points. However, most model parameter influence only a small subset of model points. For instance, the parameter that represents rising eye brows has no influence on model points at the chin. Therefore, we extract image features only in the neighborhood of influenced model points. Features are extracted at the model point positions and at positions along the model point motion defined by the model parameter. To integrate all image-bands, the image features are calculated from every image-band, see Figure 6. Afterwards, they are assembled in an image feature vector.

We integrate model trees to create a mapping of these feature values to the model point's displacement. During

 ISBN: 978-1-61208-117-5

execution time, we calculate the multi-band image representation from the probe image as demonstrated in Figure 3. We extract feature values from all image bands as shown in Figure 6. The trained model tree infers the model parameter update from these features and the model parameters are adapted, thus fitting the face model onto the image.

IV. Experimental Evaluation

This section conducts a twofold evaluation of our approach. First, the computation of the image-bands is inspected. Second, the accuracy of learned displacement expert is measured in unconstrained environments.

A. Classification of Facial Components

This experiment inspects the pixel classification accuracy. We collect a large number of images from the internet to cover a wide variety of people. The reason why we decided against using images of a standard database is, that we require the image content to vary in a large extend to train robust classifiers. We decided against using the "Labeled Faces in the Wild" database, because the images in this database are difficult to annotate due to their small size. The facial components are manually annotated in the images and the complete data set is split to images for training and testing. Adjusted pixel features are calculated from the training images and the classifiers are trained on them. Table I illustrates the accuracy of each classifier. The first row presents the number of correctly classified pixels (true positives and true negatives) divided by the total number of pixels within the face bounding box. The second row presents the number of pixels correctly classified as depicting a facial component is divided by the ground truth number of pixels depicting that facial component.

B. Fitting Accuracy on Multi-band Images

This section conducts an evaluation of our fitting approach. We collect random images from the Internet and annotate them to both, generate pixel classifiers and train displacement experts. For evaluation, we make use of the "Labeled Faces in the Wild" database, which provides images of people of public live that have been publish in the media. Therefore, these images have not been taken with a computer vision application in mind and include many challenges that have to be faced in real-world conditions. Due to the size of the database, only images of persons starting with the letter "A" are considered as a representative subset. To measure the accuracy of learned displacement experts, we induce errors $\Delta\mathbf{p}_I^{error}$ in manually specified models $\mathbf{p}_I^{\star}$ to create erroneous model parameterizations $\mathbf{p}^{error} = \mathbf{p}_I^{\star} + \Delta\mathbf{p}_I^{error}$. We apply the fitting algorithm to them to compute a parameter update $\Delta\mathbf{p}$ that is intended to compensate the artificially induced error. The difference $\Delta\mathbf{p} - \Delta\mathbf{p}_I^{error}$ between the induced error $\Delta\mathbf{p}_I^{error}$ and the suggested parameter update $\Delta\mathbf{p}$ serves as a measurement of accuracy. The maximum error induced is 1.0.

skin	lip	brow
91.5%	96.2%	92.6%
90.1%	95.1%	89.8%

Table I
THE EXTRACTION OF THE ADDITIONAL IMAGE BANDS IS ROBUST EVEN ON IMAGES THAT ARE TAKEN IN REAL-WORLD ENVIRONMENT. THIS SUPPORTS THE SUBSEQUENT TASK OF MODEL-FITTING.

image bands	original	skin	lip	brow	retina
error rate	100.0%	78.1%	64.2%	60.7%	60.2%

Table II
PROVIDING ADDITIONAL IMAGE BANDS REDUCES THE FITTING ERROR.

C. Image Bands

In Table II the impact of provided image bands on the fitting accuracy is inspected. As a baseline, a displacement expert is trained and applied on the original image data only. Then, additional image bands are provided and the accuracy of these displacement experts is compared to the baseline displacement expert. The first row of the table shows which image band has been added to train the displacement expert. Single displacement experts are represented by table columns. Please note, that for every displacement expert, additional image bands are provided and therefore, the number of image bands provided increases with the table column index. For instance, the displacement expert represented by the third column, has been trained on the original image data, the skin color image band and the lip color image band and its error is 64.1% of the baseline error. The error is computed as the average of all single parameter error values for all probe images. Since error values are reduced with increasing table column index, providing additional image bands increases the fitting accuracy.

D. Absolute Fitting Accuracy

To obtain an absolute error measurement also, we visualize the cumulative error distribution of Δp_i for all images and a selection of parameters after fitting in Figure 7. We present only a subset of the complete parameter vector to prevent results from being difficult to read. We inspect five shape parameter and four facial expression parameters. This measures the fraction of models that have at most a specific error in a specific parameter after fitting. For instance, 60% of all models are fitted with an error of 0.2 or less in the jaw drop parameter. The displacement expert evaluated here is provided the complete set of image bands. The parameters are selected from the order proposed by the face model and intuitively reflect the most important changes. As can be seen, the displacement expert significantly compensates the induced error. In some few cases, however, the initial error is even increased by the fitting algorithm, represented by the

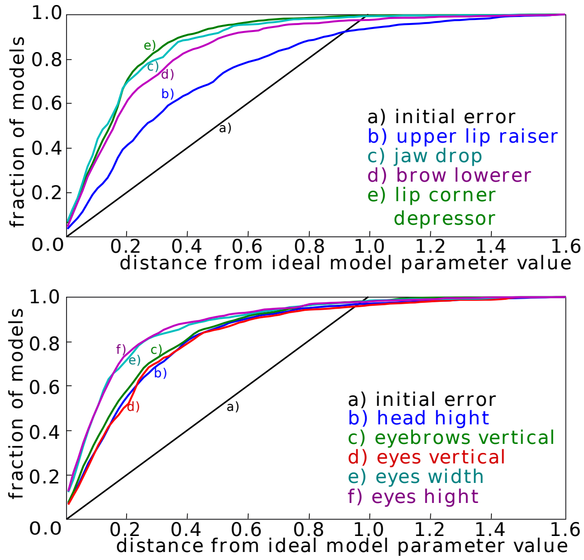


Figure 7. The fitting algorithm compensates the induced error to a large extend. For clarity of presentation, only selected parameters are evaluated.

models with errors larger than 1.0. This is due to occlusion by glasses, beards or long hair that covers the eye brows.

V. Conclusion and Future Work

We presented an algorithm for face model fitting in real-world environments, based on preprocessing the raw image data to generate multi-band images that highlight the position of various facial components. To demonstrate the robustness of the complete approach, images from the ”Labeled Faces in the Wild” database are taken for evaluation. Future work will consider applying the proposed techniques for face model tracking by calculating the multi-band images for subsequent images.

References

[1] C. Mayer and M. Wimmer, and B. Radig Adjusted Pixel Features for Facial Component Classification *Image and Vision Computing Journal, 2009*

[2] *J. Ahlberg. Candide-3 – an updated parameterized face. Technical Report LiTH-ISY-R-2326, Linköping University, Sweden, 2001.*

[3] *S.L. Phung, A. Bouzerdoum, and D.Chai. Skin segmentation using color pixel classification: analysis and comparison.* IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1), 2005.*

[4] *Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K. Jain. Face detection in color images.* IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.*

[5] *M.Beigzahed and M.Vafadoost. Detection of face and facial features in digital images and video frames. In* Proceedings of the 4th IEEE Cairo International Biomedical Engineering Conference*, 2008.*

[6] *M.T. Sadeghi, J.V. Kittler, and K. Messer. Modelling and segmentation of lip area in face images.* Vision, Image and Signal Processing*, 2002.*

[7] *V. S. Sadeghi and K. Yaghmaie. Vowel recognition using neural networks.* International Journal of Computer Science and Network Security*, 2006.*

[8] *D. Cristinacce and T. F. Cootes. Feature detection and tracking with constrained local models. In* 17th British Machine Vision Conference*, 2006.*

[9] *D. Cristinacce and T. F. Cootes. Boosted regression active shape models. In* Proceedings of the British Machine Vision Conference*, 2007.*

[10] *V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In* Siggraph, Computer Graphics Proceedings*, 1999*

[11] *T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H. Burkhardt and Bernd Neumann, editors,* 5th European Conference on Computer Vision*, 1998*

[12] *K. I. Chang, K. W. Bowyer, and P. J. Flynn. Face recognition using 2d and 3d facial data. In* ACM Workshop on Multimodal User Authentication*, 2003.*

[13] *P. J. Phillips, H. Moon, S. A. Rizvi, and Patrick J. Rauss. The feret evaluation methodology for face-recognition algorithms.* IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.*

[14] *O. Jesorsky, K. J. Kirchberg, and R. Frischholz. Robust face detection using the hausdorff distance. In* Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, 2001*

[15] *T. F. Cootes and C. J. Taylor. On representing edge structure for model matching.* Computer Vision and Pattern Recognition*, 2001.*

[16] *Mikkel Bille Stegmann and R. Larsen. Multi-band modelling of appearance.* Image and Vision Computing Journal*, 2003.*

[17] *F. Kahmaran and M. Gokmen. Illumination invariant face alignment using multi-band active appearance models. In* Pattern Recognition and Machine Intelligence*, 2005.*

[18] *Gary B. Huang, M. Ramesh, T. Berg, and E. Learned-miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments, 2008.*

[19] *H. Wu, X. Liu, and G. Doretto. Face alignment using boosted ranking models. In* Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2008.*

[20] *P. Tresadern, H. Bhaskar, S. Adeshina, C. Taylor, and T.F. Cootes. Combining local and global shape models for deformable object matching. In* Proceedings of the British Machine Vision Conference*, 2009.*

[21] *S. Romdhani.* Face Image Analysis using a Multiple Feature Fitting Strategy. *PhD thesis, University of Basel, Computer Science Department, Basel, CH, January 2005.*

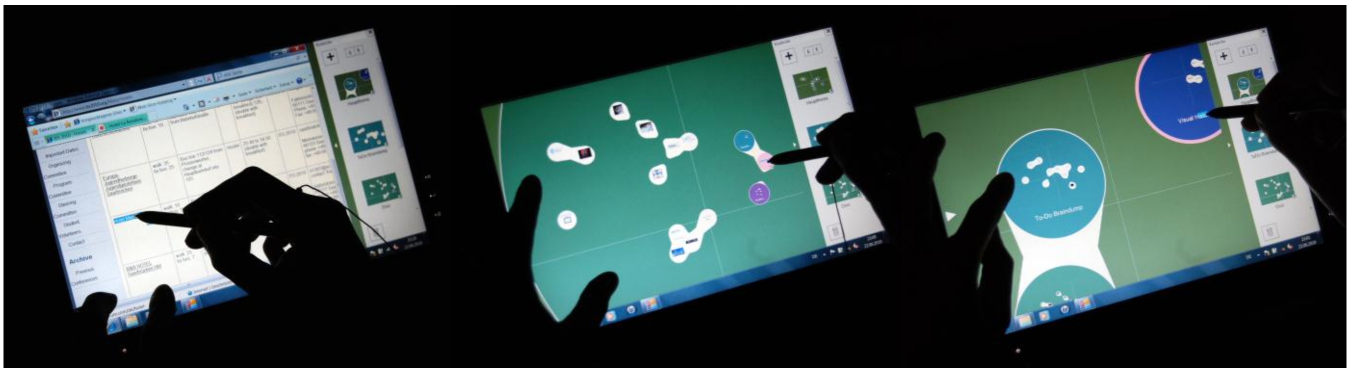
[22] *P. Viola and M. J. Jones. Robust real-time face detection.* International Journal of Computer Vision*, 57(2):137–154, 2004.*

 ISBN: 978-1-61208-117-5

An Interface for Visual Information-Gathering During Web Browsing Sessions:

BrainDump - A Versatile Visual Workspace for Memorizing and Organizing Information

Marius Brade, Joerg Heseler, Rainer Groh
Chair of Media Design
Technische Universität Dresden
Dresden, Germany
marius.brade@tu-dresden.de, joerg.heseler@gmx.de, rainer.groh@tu-dresden.de



***Abstract*— We propose a new method for visualizing the strength of associations based on a fluid metaphor and cell structure. This enables users to place gathered information visually in relation, while solving complex information tasks. Both, vague and precise relations can be visualized. We applied our approach to a scenario of information retrieval during web browsing sessions. In this paper we describe our novel visual information-gathering system called *BrainDump*. To support the user's sensemaking process, this system provides the possibility to change the initial associations, follow links back to the source of information, annotate content and hierarchically group collected items. A preliminary user test was designed and conducted.**

Keywords-graphical user interface; visual information gathering; visual sensemaking; personal web information system

I. INTRODUCTION

Complex web information tasks involving answers to more than one question to satisfy a user's goal are called *information-gathering tasks*. Users executing this kind of task are called knowledge workers and are defined by Sellen et al. as people "whose paid work involves significant time: gathering, finding, analyzing, creating, producing or archiving information" [36].

Working on complex information-gathering tasks, knowledge workers have to understand unfamiliar contexts – a process which has been coined *sensemaking* [12][35]. This complicated, iterative process occurs in everyday life and requires a high cognitive load. For instance, when a researcher faces an unfamiliar field of research, he or she has to gather information from an unfamiliar field and make sense of that information.

Considerable research has investigated how people organize information making use of spatial memory [24][27][31]. Spatial layouts are often used to manage transient or temporary information in current visual interfaces. Examples include web bookmarks [34], spatial hypertexts [38] and desktop icons [2]. In most systems, a cluster represented by a spatial aggregation of elements, exists only in the user's mind. However, internal mental representations perform poorly on making sense of complex and rich object relations or associations. In contrast, well designed external representations can support this sensemaking process [35].

Especially at the beginning of this process, the user anticipates associations and potential relationships. While making sense, there are a lot of changes in the user's mind, concerning the structure of collected information. Organizing information with the help of spatial layouts is far more flexible and lightweight than using explicit grouping mechanisms such as bookmark-folders [1][3]. This makes spatial layouts an excellent technique to organize information in the transient, temporary states of a sensemaking process.

To understand the relations between the contents found in a complex web information-gathering task, knowledge workers use bookmarks, tabs, browsing history, different browser-windows, virtual documents, folders, offline documents, and sketches. The challenge is to keep in mind where specific information is located. This problem has been called *information fragmentation* [22][34]. Users might lose the *big picture* of their task or the context of the currently

 ISBN: 978-1-61208-117-5

pursued question, during a long session on the web. Since most information-gathering tasks require multiple web sessions to satisfy a goal [24], annoyance and frustration with current systems is increased [28]. Loss of context happens when users are getting sidetracked, forgetting their original questions involved in the information-gathering task or when the title of a bookmark cannot be remembered. The results of a study by Kaasten et al. [37] show that thumbnails of visited websites support the recognition considerably. In this study, color, distinctive images, layout, and pieces of text on the webpage are features helping the user to recognize the exact page. Plain use of titles, like in browser-tabs and browser-windows cannot take this into account, and therefore may not be enough to remember the websites behind them. Hence, a part of the task already achieved has to be repeated in order to reacquire the lost information.

We are addressing the problems discussed above with our new approach. A knowledge worker can use a highly flexible visual map to represent and refine his current understanding of a task using the human strength in spatial memory. A metaphor based on fluids and cell structure allows the user to visually memorize anticipations about relations with the option of using textual annotations. We assume that resuming an interrupted task, or continue an earlier finished research is highly accelerated by utilizing a map that helps the user in information retrieval and to remember his last status of understanding. The concept aims towards offering the user a high degree of recognition for collected elements.

This paper consists of four parts: the first part shows related work, the second part explains our design decisions while the third part describes the interface of our prototype named BrainDump from a user's perspective. The fourth part briefly describes our first short preliminary user study.

II. Related Work

In this section, related work of different fields concerning the BrainDump approach is presented.

A. Spatial Layout and Aggregation

Spatial layout has been a basic feature in window based systems, like the early Macintosh. It has long been used for document management. There are several extensions to this 2D layout, for example ZUIs (Zoomable User Interfaces). ZUIs aim on solving the problem of having more information, than fits on the screen by providing panning and zooming operations. The first system to explore this approach was Pad [33]. Bederson defines ZUIs as "systems that support the spatial organization of and navigation among multiple documents or visual objects [4]."

Spatial aggregation plays an important role in management of information using spatial layout. As a part of an original experimental workshop [9][10] we collected ideas how to visualize relations between objects based on the properties of fluids. For example, a person can drag parts of fluid substances over a surface and form them to bigger or smaller groups (see Figure 1).

Figure 1. Snapshot of our experiments with fluids [9]

An interface for manipulating spatial aggregations of objects is Bubble Clusters [41]. A spatially efficient bubble shape is drawn around objects, leveraging the natural expectation of users concerning the behavior of bubbles.

This approach in line with a visual gimmick named Realtime Blobs, found on the website of the designer Ivanov [20], inspired us to use rotund shapes in our system.

B. Techniques for Mapping the Mind

Several techniques to create different kinds of external representations exist [14]. One of the most commonly used is the mind mapping technique. Eppler defines a mind map as "a multicolored and imagecentred, radial diagram that represents semantic or other connections between portions of learned material hierarchically" [14]. The typical application context is personal note taking and reviewing. In addition to Concept Maps, Conceptual Diagrams, Semantic Maps [18], and Topic Maps [32], the major advantage is that these techniques mostly provide a concise overview. Inspired by these approaches, our vision is to give the user a permanent and reliable overview of the context of his information-gathering tasks while working on them.

C. Systems Supporting Web Tasks

An example for the 3D arrangement of Bookmarks is Data Mountain [34]. In this approach the user can place documents on an inclined plane textured with landmarks. It follows an interesting approach that takes advantage of human spatial memory skills.

For sensemaking and manual collection via drag and drop of web-scraps like pieces of text, pictures, and URLs, Scratch Pad [7] is a conceptual approach. The simple way to collect items from websites via drag and drop is extended in our approach by displaying a visual metaphor for the collected items.

III. Design Decisions

Our contribution is a system, where the user defines a visual map of his or her comprehension of relations between content, established while working on web information-gathering tasks. Manually collected information can be spatially arranged, visually related, and annotated.

 ISBN: 978-1-61208-117-5

A. Visual Metaphor

For displaying associations and relationships between collected objects, we use a visual metaphor. As described in Section II, it is based on the idea of fluids, Bubble Clusters, and Realtime Blobs. The authors of Bubble Clusters proved that users can quickly create, merge, and split groups by moving objects on the screen. In addition to the existing approaches, we display visual real-time feedback of the strength of relations, while a user drags two objects next to each other. Thus, different intensities of a relationship can be defined as shown in Figure 2: objects labeled 1, 2 and 3 are dragged at a position, where they have a visually loose connection (left) or a rather visually strong connection (right). Due to the real-time visual feedback while dragging objects next to others, it seems as if the user is dealing with a fluid substance.

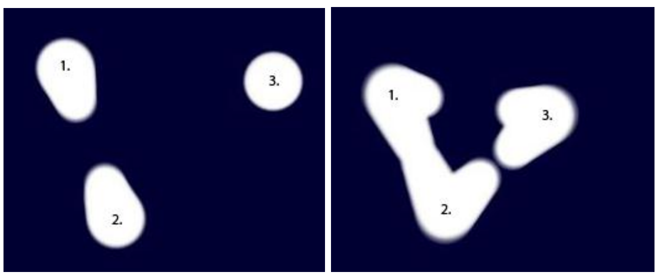


Figure 2. Metaphor for displaying relationships of different intensities. Left: weakly related objects. Right: strongly related objects

B. Visualization Algorithm

Objects are represented as white filled circles. The visual relation depending on the distance between two objects is represented by several white filled circles, drawn on a non-visible straight line between the center points of these objects (see Figure 3). This visual relation is displayed by an algorithm based on our observation of Realtime Blobs [20].

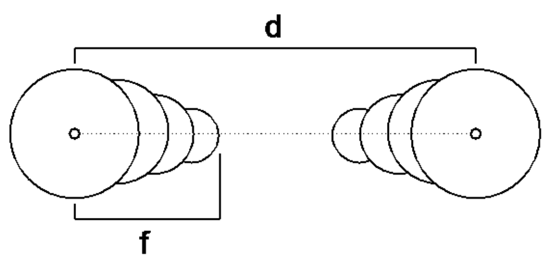


Figure 3. Schema of the drawing-algorithm: distance *d* between two objects and the circles expansion distance *f*

If the distance *d* between two objects falls below a prior defined limit, the particular circles for the relation are drawn by the system. They become linear smaller until one third of the original objects diameter is reached. If the user drags an object half the distance of the prior defined limit of *d* the circles reach the center of the other object. Until this point the circle expansion distance *f* increases - and therefore the number of circles drawn. As one object is dragged closer to another object, *f* decreases again, until it reaches zero. The objects appear to attract each other (see Figure 2). This algorithm achieves a reasonable balance between aesthetics and performance.

C. Hierarchical data structure

With our current concept we are combining the flexibility of spatial layouts with the benefits of hierarchical grouping. An object in our system can be an *item* or a *topic*. Items are manually collected parts from websites (see Table 1) with a white circle in the background - imitating a cell with a core (see Figure 4). They are created by the user to represent a hyperlink to that specific website. The users' collected items can be grouped into topics. These topics are slightly bigger than items, with a colored background (see Figure 4) and can be put into other topics to build a hierarchy. The data structure provided by the BrainDump system is a tree structure with one single root node. Every node is a topic and every leaf is an item. The user can decide to work in the root topic alone or to create other topics for hierarchical grouping.

For navigation and organization in our system, we are using the technique of a ZUI with animated transitions and a multi-level layout [5][15]. By accessing the different hierarchical structured topics the predefined zoom levels are helping the user in navigation tasks. Studies on how people learned from zoomable spaces imply that users working with ZUIs may be more engaged and may remember the spatial structure of the content better [8][16].

D. Representation and Handling of Gathered Items

One of the design guidelines for ZUIs by Bederson is only to use ZUIs, if there is a small visual representation of the data available [4]. Thus, our concept lets the user choose his or her personal small representation for a collected website displayed in the center of an item. In addition, the title of the website is displayed under the representation and can be edited by the user (see Figure 4).

Table 1 shows which parts of a website can be manually collected using what kind of interaction.

TABLE I. ADDING OBJECTS TO THE SYSTEM

Web browser source	**Representation in the system**	**Interaction**	
		for selection	***for collecting***
URL	Thumbnail of the actual entire website	Drag	Drop
Picture	Picture	Drag	Drop
Video[a]	Video (playing while hovering cursor)	Drag	Drop
Pieces of text	Thumbnail of user selected text with direct surroundings	Marking, dragging	Drop
User-defined clipping of a website[a]	Thumbnail showing the respective part	Lasso/ rectangle selecting, drag	Drop
search queries	outline around object, with favicon of the search engine used before finding the respective object	added automatic ally	added automatic ally

a. not included in the current implementation of the prototype

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-117-5

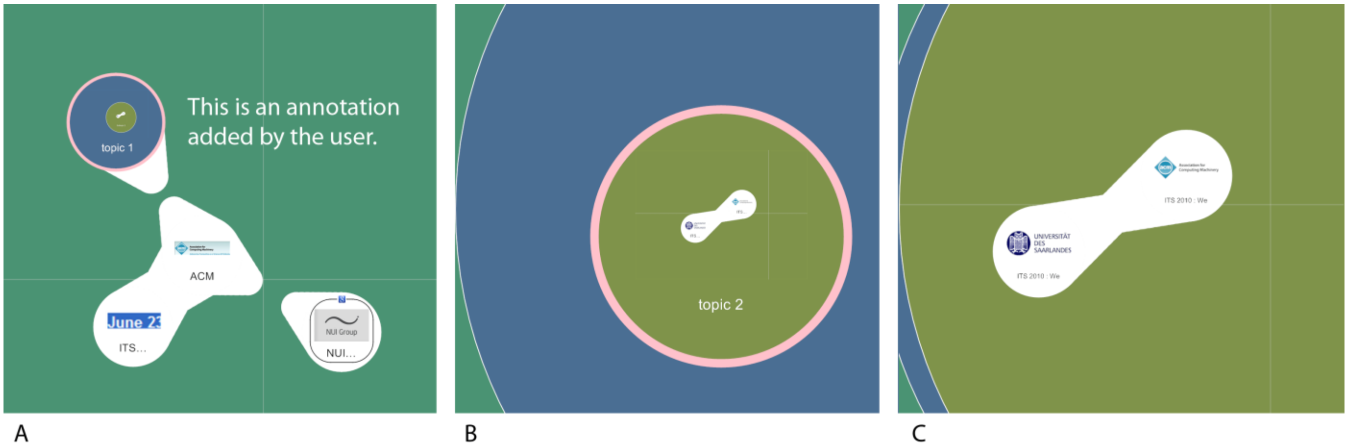


Figure 4. Look and feel of BrainDump prototype interface with hierarchically clustered topics (colored) and items (white). *A*: Screen-capture of the green main topic with three items and one blue topic. *B*: Screen-capture of the zoomed-in view on the blue *topic 1* with no items and one topic. *C*: Zoomed-in view on the chartreuse *topic 2* with two items. The rings on the left side of the screens *B* and *C* can be clicked to zoom-out and get up the hierarchy.

The only element which is added automatically to the system is a search query the user submitted to find a collected item. This query will appear at the respective website representation in the system (see Figure 4, *A* and Figure 7 - black outline), making it easy to recall successful queries later. The website's favicon - a small graphic that is associated with a website - is displayed on top of the outline.

To recognize a collected website later, we are expecting, that choosing his or her own representative for this site helps the user, since learning techniques rely on associations between easy-to-remember constructs and the data that has to be remembered [14]. Despite being carefully chosen, the items in a user's collection are sorted and their relevance is questioned by the user. Since any changes in the layout are only made by the user himself, a consistent layout is provided. This can help the user to build up an internal map over time and make it easy to recognize objects. Using a consistent layout is another design guideline by Bederson [4] which is fulfilled by our concept.

In addition to gathering items, the user can include annotations in his collection (see Figure 4, *A* - white text on green background). This is achieved by a right-click of the mouse on the background at the position where text should be inserted. The font-size is always constant in relation to the pixel dimensions of the display. Thus, the user can place different sizes of annotations into the system by changing the actual zoom-level before adding an annotation.

E. *Summary of Design Decisions*

To give an overview of our design decisions, we list them in form of requirements for Information-Gathering Systems:

- manual visual definition of relations with optional textual description
- definable strength of relations between objects
- free spatial sorting and arranging of objects
- hierarchical structuring of items with topics
- automatic saving of item related search queries
- manually collected representations of websites, according to interesting content for the user (e.g., pictures, text, videos)
- making annotations to the collection
- local saving of websites related to collected items to prevent the loss of outdated websites
- local saving of links (URLs) referring to the respective website

F. *Implementation of the prototype*

As proof of concept, we implemented a first prototype, a windows application, running on Windows XP and higher versions. The current version is based on the .NET-Framework and uses OLE (Object Linking and Embedding) for embedding and linking to documents and other objects. Therefore, the prototype can be used in combination with Microsoft Internet Explorer 7 or higher. The collected data is stored in an XML-file. For realizing the ZUI, we used the Piccolo.NET framework [6].

IV. THE BRAINDUMP INTERFACE

The BrainDump interface consists of two parts: The *Dropping Area* and the *Management Mode*. We will describe them in the first two subsections. The third subsection describes navigation in BrainDump.

A. *Dropping Area*

This part of the interface appears at the right side of the browser in a separate window (see Figure 5). The dropping area can be equipped with links referring to topics existing in the user's collection. We call these links *Topic Portals* (see Figure 5, right column). The user can drag pictures, pieces of text, and URLs onto these links. While hovering over a Topic Portal its size increases (see Figure 5, B). The user can place his dragged content and define the strength of the visual relationship to other gathered items. He can also put items into topics existing in the Topic Portal. These portals are shortcuts for the user to optimize his conduct of information-gathering tasks. Portals avoid changing the application window for placing or moving an object. Moreover, the user can have multiple topics easily accessible while browsing.

B. Management Mode

The Management Mode offers a fullscreen view of the user's collection. Here he can manage and organize all collected items and topics (see Figure 6) - labeling of topics included (see labels in Figure 4. A and B). Furthermore, the user can manage the Dropping Area and utilize Topic Portals as shortcuts for moving objects to topics in different hierarchies. He can create a Topic Portal by dragging a chosen topic from the canvas onto the Dropping Area. Topic Portals can be deleted without any consequences for the actual topic inside the system. As a result, the user can optimize his dropping area for the context of his current information-gathering tasks.

A **B**

Figure 5. Dropping area on the right side of the browser, *A*: Marking a piece of text and *B*: Dragging the piece of text into a topic portal

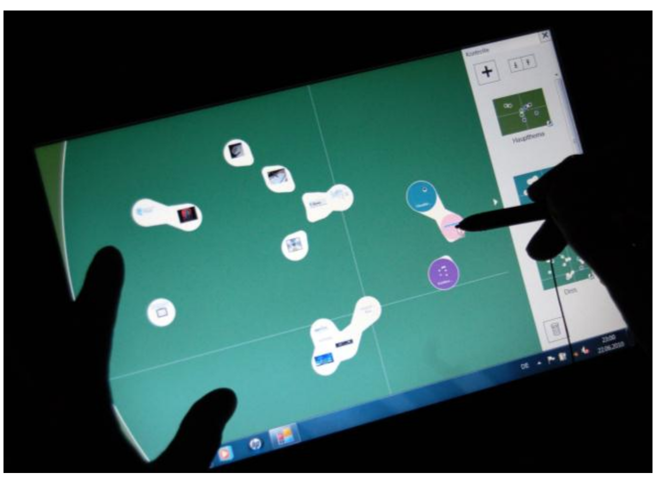


Figure 6. The Management Mode with the Dropping Area on the right

To create a new empty topic the user can press the "+" button in the upper left of the dropping area. To create a new topic with an object already in it, an item or topic can be dropped on this button. This instantly creates a new topic in the one currently viewed. Thus, a new node is inserted into the internal tree of the systems data hierarchy, as described in Section III.C.

C. Navigation in BrainDump

The user can zoom in on a chosen topic by double clicking it. By doing so an interactive ring on the left side of the screen appears (see Figure 4, B). More rings appear if the user navigates deeper into the hierarchy (see Figure 4, C). These rings enable zooming out of a topic and thus navigating up in the hierarchy. Additionally, these rings provide an overview at which hierarchy level the currently viewed topic is located. Backtracking can be done step by step up the hierarchy or directly jumping to an upper topic. The transitions are animated to support user orientation. An example is shown in Figure 4: First the user views the green main topic with three items and one topic in it (*A*). Then he navigates inside the blue topic named *topic 1* (*B*). Being inside this topic, a green ring for backtracking to the main topic appears on the left side of the screen (*B*). Similarily, an additional blue ring appears when the user navigates into the chartreuse colored *topic 2* (*C*).

Depending on the zoom scale, the text in an item may not be legible. To address this problem, an alternative, summarized representation is chosen by the system. When a text line is small on the screen the user may only want to see its beginning. As the item is magnified, this may be augmented by a short summary or outline. At some point, the entire text line is revealed. This technique is called semantic zooming [33]. As shown in Figure 7, we have implemented semantic zooming for displaying a reduced *(A)* and full *(B)* display of search queries and titles of items.

A B

Figure 7. *A*: minimum and *B*: maximum of displayed information. Items consist of: representation of the respective website (here a picture), title of the website and search query displayed through favicon and black outline.

In doing so, we follow the suggestions by Kaasten et al. [23], who found out that right truncation of website titles allows for best recognition.

V. PRELIMINARY USER TEST

Since our prototype is in an early development stage, we chose to conduct a formative evaluation, a process of ongoing feedback to improve and optimize our concept. Thus we conducted a preliminary user test with nine participants to elicit first directions concerning three questions:

1. Do users get lost while navigating in the ZUI?
2. How is the acceptance of the used metaphor?
3. Can users recognize collected visual information?

The average age of the test persons was 29, two of the nine participants were female, no one used the Internet Explorer during their work before.

A. Design

In order to obtain valid results, we made a checklist for ensuring exactly the same course of actions for each test user. A task was designed including a scenario, where the users had to plan a trip to an exhibition. The participants should imagine they want to attend the conference *games convention*. To prepare for the visit, they should get an overview through saving related information using BrainDump. The scenario was described in detail with three subtasks: getting started, getting information about exhibitors and getting information about the conference. For each subtask the participants were given detailed information, what they have to look for, e.g. "Because you are interested in browser games you want to meet the company gamigo AG".

This scenario consisted of two stages: training and evaluation. The training stage aimed at providing some practice in the use of information-gathering functions of the prototype and to support the test persons in understanding the purpose of BrainDump. The target of the evaluation stage was to identify usability problems referring to the three questions mentioned above. We provided precise subtasks for the test persons in the given scenario.

To get impartial results, we observed the participants during the evaluation and recorded their sessions, including mouse clicks, keystrokes and voice using the software *Morae 3* from *Techsmith* [30]. Apart from that, we wrote down comments and reactions of the participants (think aloud protocols).

A questionnaire was given to the test persons to investigate their subjective attitude towards the usage of the prototype. This questionnaire was divided into ten sections with 75 questions in total: the first section covered feedback about the evaluation itself, and if the scenario including subtasks was comprehensible. The second section served to estimate general usability problems by means of the System-Usability-Scale. Detailed usability questions were asked within sections three to seven, using a standardized heuristic. Section eight investigated aspects of visualization, interaction, and user satisfaction. In the ninth section, participants were asked to state missing features. Demographic questions together with earlier used software were asked in the last section. Throughout the questionnaire, a five point Likert scale was used to examine the participants' level of agreement to each statement.

B. Procedure

In total, the evaluation took about 35 minutes per test person. At first, the purpose of BrainDump was explained briefly. In order to not affect the results, no further aims of the evaluation and software were revealed. Exactly three features of BrainDump were shown to all participants to create equal starting knowledge: 1. How to collect and put items into a Topic Portal in the Dragging Area. 2. Making annotations in the system. 3. Creating a new topic.

With handing out the scenario descriptions, the recording started. The practical part took about 25 minutes to complete. Afterwards the questionnaire was given to the test persons.

C. Results

The purpose of BrainDump was understood completely. All of the participants would like to use the application more often in their daily work. They especially liked the fast way to visually change relations and associations between information.

Some shortcomings of the current implementation were identified: Four of the nine test users thought that the software does not provide sufficient information about which actions are currently permitted. An equal number of the participants found the font size used for the labels too small. Six persons were irritated by the automatically drawn relation between objects. Four persons expected the possibility of manipulating the drawn relations. Three participants expected by mistake that some of their topics were empty, because the content was not visible. Two of them deleted such a topic that was not empty. Five test persons expected to be able to select multiple objects at once. Four users where irritated by topic portals or rather did not recognize them as links referring to topics.

Further additional features were requested: A solution to cut off unwanted relations with a gesture, manipulation of relations between items and scaling of objects to highlight importance.

D. Discussion

The results can be interpreted by taking the three initial questions into account. Since the number of nine participants is hardly representative, no generalized conclusion can be drawn at this point. The results have to be treated carefully. Further, the participants had no possibility to get used to the system which increases irritation, especially when not being used to a ZUI. In addition, the performance of the prototype was slowed down by the capturing software Morae. This made dragging of items from websites into the system extremely slow - thus further irritating users.

1) Do the users get lost while navigating in the ZUI?

The deletion of non empty topics may be a result of the fact that users got lost while navigating. Bederson discovered that when objects are placed at many different levels - and therefore very small sizes - users won't remember that they exist [4]. To solve this problem, the concepts' spatial layout can be designed to indicate hidden objects to the user. Another possibility is to minimize the depth of the hierarchy to circumvent undersized objects. As observed, some people had problems understanding and using Topic Portals. A study by Hornbæk showed that an integration of overview and detail windows requires mental and motor effort [19] - since Topic Portals are very similar, this might indicate that this problem does not necessarily depend on being used to it. In a following study, it has to be investigated whether this

 ISBN: 978-1-61208-117-5

problem stems from graphical representation or the concept itself.

A solution for these problems might be to use a hybrid system - a combination of a ZUI with traditional approaches like facets or folders.

2) How is the acceptance of the used metaphor ?

Despite liking the visually changeable relations and being engaged with the systems use, some participants were irritated by the automatically drawn relations and missed direct manipulation of these.

One possible explanation for this irritation might be that the relation depends solely on distance - which was not expected by the user and might be a less critical factor when participants are used to the system. To minimize these irritations, the metaphor could be refined to a more natural, expected behavior and with regard to connecting items independently from distance. Direct manipulation of the relations has already been part of our concept, but not yet included in the current prototype.

3) Can users recognize collected visual information?

Despite observing the problems with font size, the duration of the evaluation was too short to simulate an interruption long enough for the users to be a handicap. Therefore, a study taking several days with a long break in between will be necessary.

VI. CONCLUSION AND FUTURE WORK

This paper presents a new approach to ease the process of memorizing and organizing information during complex tasks using the internet. This is achieved by a visual approach to connect objects to unique shaped groups and the possibility to relate objects in different visual intensities. The user can choose what part of a website he wants to be the representation of his bookmark. Manifold associations can be mapped on one simple and powerful metaphor. Since the user places objects in space and no one else is moving them except himself, he can build up a personal internal map over time. Indicated by considerable research [4], a consistent layout can help the user to recover objects.

As proof of concept, we implemented a prototype and conducted a first preliminary user test, showing that participants enjoy using such a system, but also indicating that several parts of the prototype and concept have to be refined. Based on the results, we will improve the presentation of items in the next version of our prototype to improve legibility. The approach of Strobelt et al. uses interesting algorithms and heuristics which we want to take into account for a refined concept [40]. Apart from displaying well designed representations, Klemmers' work on Design History [25] suggests that also the history of creation may be important to reconstruct meaning. We plan to integrate this idea in the current concept.

Since zooming in and out repeatedly is typically straining [4], this should be minimized. Therefore, we will consider using a hybrid system - a combination of a ZUI with traditional approaches to minimize depth of zoom levels. A further improvement in orientation could be to use landmarks, as seen in Data Mountain [34]. Inspired by this idea, we will enhance our concept with structured background pictures.

To minimize user irritation, the visual metaphor will be refined to a more natural behavior, possibly enhancing it to enable visual relation apart from using distance. We already made first experiments with new algorithms (see Figure 8) derived from experimental observations [9][10].

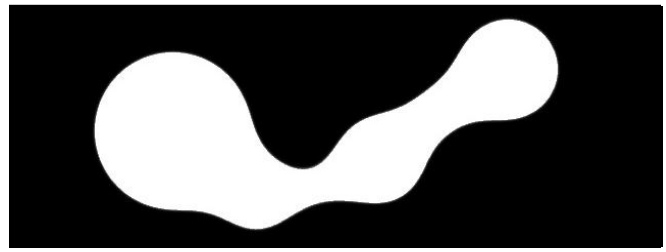


Figure 8. First Experiments with new algorithms

Besides, we plan to enable the use of documents other than websites. For example contact files, office files, e-mails, and information from the file system or social networks. Thereby, we are addressing the problem that users lose track of their project documents and relations, studied by Bergman et al. [7], the authors called this *project fragmentation*. Furthermore, we plan to conduct a qualified user study and compare other systems to our system.

ACKNOWLEDGMENT

On behalf of Marius Brade this research project (contract no. 080951799) is funded by the European Social Fond and the Free State of Saxony as well as SAP Research.

The authors would like to thank Matthias Baumgart for his support on user testing. Further, we would like to thank the test participants for their helpful and valuable feedback and Dietrich Kammer for his valuable proofreading.

REFERENCES

[1] Abrams, D., Baecker, R., and Chignell, M. Information archiving with bookmarks: personal web space construction and organization. In CHI ’98: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1998, 41–48

[2] Agarawala, A. and Balakrishnan, R. Keepin'it real: pushing the desktop metaphor with physics, piles and the pen. Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM, 2006, 1283--1292

[3] Bauer, D., Fastrez, P., and Hollan, J. Spatial Tools for Managing Personal Information Collections. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on System Sciences, IEEE, 2005, 104b.

[4] Bederson, B.B. The Promise of Zoomable User Interfaces. In Proc. CHI 2009, ACM Press, 2009, 4-9.

[5] Bederson, B.B. and Boltman, A. Does animation help users build mental maps of spatial information? In INFOVIS, 1999, 28–35.

[6] Bederson, B. B., Grosjean, J., and Meyer, J. (2004). Toolkit Design for Interactive Structured Graphics, IEEE Transactions on Software Engineering, 30 (8), pp. 535-546.

[7] Bergman, O., Beyth-Marom, R., and Nachmias, R. The project fragmentation problem in personal information management. Proceedings of the SIGCHI conference on Human Factors in computing systems, ACM, 2006, 274

[8] Boltman, A., Druin, A., Bederson, B., Hourcade, J.P., Stanton, D., O'Malley, C., Cobb, S., Benford, S., Fast, C., Kjellin, M., and Sundblad, Y. (2002). The nature of children's storytelling with and without technology. Proceedings of American Educational Research Association (AERA) Conference, New Orleans.

[9] Brade, M., Kammer, D., Keck, M., and Groh, R. Immersive Data Grasping using the eXplore Table. TEI'11, Funchal, Portugal. ACM Press, 2011 (In press)

[10] Brade, M., Kammer, D., Keck, M., and Groh, R. Immersive Data Grasping. Workshop at TU Dresden, 2010. http://www.youtube.com/watch?v=ygAq8JHGtvM

[11] Buzan T. Using Both Sides of the Brain. Dutton: New York, 1974.

[12] Dervin, B. An overview of Sense-Making research: Concepts, methods and results. Paper presented at the meeting of the International Communication Association, Dallas, TX (May 1983).

[13] Eklund, J. Sawers, J., and Zeiliger, R. Nestor naviga-tor: A tool for the collaborative construction of knowledge through constructive navigation. In R. Debreceny & A. Ellis (eds.) proceedings of Ausweb99, The Fifth Australian World Wide Web Conference. Southern Cross University Press., Lismore, 1999, 396–408

[14] Eppler, M. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. Information Visualization 2006 Volume 5, Palgrave Macmillan Ltd. 2006, 202 -210

[15] Furnas, G.W. and Bederson, B.B. Space-scale dia-grams: understanding multiscale interfaces. In CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1995, 234–241

[16] Good, L. and Bederson, B. B. 2002. Zoomable user interfaces as a medium for slide show presentations. Information Visualization 1, 1 (Mar. 2002), 35-49.

[17] Gotz, D. The scratchpad: sensemaking support for the web. In WWW '07: Proceedings of the 16th international conference on World Wide Web, ACM, New York, NY, USA, 2007, 1329–1330

[18] Hodel, M. Organisationales Lernen - dargestellt an der Erarbeitung und Implementierung eines durch Mind Mapping visualisierten Qualitätsleitbildes. Theoretische Überlegungen und praktische Erfahrungen aus einer Fallstudie, St. Gallen (Dissertation). 1995

[19] Hornbæk, K., Bederson, B. B., and Plaisant, C. 2002. Navigation patterns and usability of zoomable user interfaces with and without an overview. ACM Trans. Comput.-Hum. Interact. 9, 4 (Dec. 2002), 362-389.

[20] Ivanov, D. Realtime Blobs. Den Ivanov blog. http://www.cleoag.ru/labs/flash4/009/blobs.html 20.10.2010

[21] Jones, W., Phuwanartnurak, A.J., Gill, R. and Bruce, H. Don't Take My Folders Away! Organizing Personal Information to Get Things Done. In Proc. CHI 2005, ACM press, 2005, 1505-1508.

[22] Jones, W.P. and Teevan, J. Personal Information Management. University of Washington Press, Okt. 2007

[23] Kaasten, S., Greenberg, S., and Edwards, C. How people recognize previously seen web pages from titles, urls and thumbnails, 2002.

[24] Kirsh, D. The intelligent use of space. Artificial intelligence, volume 73, Elsevier 1995, 31--68.

[25] Klemmer, S.R., Thomsen, M., Phelps-Goodman, E., Lee, R., and Landay, J.A. Where do web sites come from?: capturing and interacting with design history. Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves. ACM, 2002, 8.

[26] MacKay, B. and Watters, C. Exploring multi-session web tasks. In CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, ACM, New York, NY, USA, 2008, 1187–1196

[27] Malone, T.W. How do people organize their desks?: Implications for the design of office information systems. ACM Transactions on Information Systems (TOIS), Volume 1, ACM, 1983, 99--112.

[28] Mark, G., Gudith, D., and Klocke, U. The Cost of Interrupted Work: More Speed and Stress. Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. ACM, 2008, 107--110.

[29] Mayer, M. and Bederson, B.B. Browsing Icons: A Task-Based Approach for a Visual Web History. In HCIL-200119, CS-TR-4308, UMIACS-TR-2001, vo-lume 85, 2001.

[30] Morae 3 - Kundenwünsche wirklich verstehen | TechSmith http://www.techsmith.de/morae.asp 20.10.2010

[31] Nakakoji, K., Yamamoto, Y., Takada, S., and Reeves, B.N. Two dimensional spatial positioning as a means of reflection in design. Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques. ACM, 2000, 154.

[32] Park, J. and Hunting, S. XML Topic Maps: Creating and Using Topic Maps for the Web. Addison-Wesley Longman Publishing Co., Inc. 2002. Boston, MA, USA

[33] Perlin, K. and Fox, D. Pad: an alternative approach to the computer interface. In SIGGRAPH '93: Proceed-ings of the 20th annual conference on Computer graphics and interactive techniques, ACM, New York, NY, USA, 1993, 57–64

[34] Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., and van Dantzich, M. Data Mountain: Using Spatial Memory for Document Management. Proc. of User Interface and Software Technology (UIST 98) ACM Press, 1998, pp. 153-162.Scr

[35] Russell, D., Stefik, M., Pirolli, P., and Card, S. The cost structure of sensemaking. Proceedings of ACM CHI '93. New York: ACM, 1993, 269-276

[36] Sellen, A.J., Murphy, R., and Shaw, K.L. How know-ledge workers use the web. In CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, NY, USA, 2002, 227–23

[37] Shaun Kaasten, Saul Greenberg, and Christopher Edwards. How people recognize previously seen web pages from titles, urls and thumbnails, 2002.

[38] Shipman III, F.M., Marshall, C.C. Spatial hypertext and the practice of information triage. Proceedings of the eighth ACM conference on Hypertext, April 06-11, 1997, Southampton, United Kingdom, 124-133.

[39] Shipman III, F.M., Marshall, C.C., and Moran, T.P. Finding and using implicit structure in human-organized spatial layouts. Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1995, 346–353.

[40] Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D. A., and Deussen, O. Document Cards : A Top Trumps Visualization for Documents. First publ. in: IEEE Transactions on Visualization and Computer Graphics 15, 2009, 6, pp. 1145-1152

[41] Watanabe, N. Washida, M., and Igarashi, T. Bubble clusters: an interface for manipulating spatial aggregation of graphical objects. Proceedings of the 20th annual ACM symposium on User interface software and technology, ACM Press, 2007, 173--182.

Navigation and Interaction in the Virtual Reconstruction of the Town of Otranto in the Middle Ages

Lucio T. De Paolis, Giovanni Aloisio
Department of Innovation Engineering
Salento University
Lecce, Italy
lucio.depaolis@unisalento.it
giovanni.aloisio@unisalento.it

Massimo Manco
Engineering Faculty
Salento University
Lecce, Italy
massimo.manco@unisalento.it

Abstract— **The main goal of the Human-Computer Interaction technology is to improve interactions between users and computers by making computers more usable and receptive to the user's needs. This paper focuses on an application of navigation and interaction in a virtual environment using the Wiimote and the Balance Board of Nintendo. The aim is to make the interaction easier for users without any experience of navigation in a virtual world and more efficient for trained users. The application has been developed for the navigation and interaction in the virtual environment of the MediaEvo project. The MediaEvo Project aims at developing a multi-channel and multi-sensory platform for edutainment in Cultural Heritage.**

Keywords-component; user interface, cultural virtual heritage, Nintendo WiiMote, Nintendo Balance Board

I. INTRODUCTION

Edutainment, a neologism created from the combination of the words education and entertainment, refers to any form of entertainment aimed at an educational role. The videogame is one of the most exciting and immediate media of the edutainment applications because the game enables a type of multisensory and immersive relationship of the user through its interactive interface; moreover, the cyberspace of the videogame is a privileged point of sharing and socializing among players.

Edutainment is an up-and-coming field that combines education with entertainment aspects; thus, it enhances the learning environment and makes it much more engaging and fun-filled.

One of the most important applications of edutainment is undoubtedly the reconstruction of 3D environments aimed at the study of cultural heritage; the use of Virtual Reality in this field makes it possible to examine the three-dimensional high-resolution environments reconstructed by using information retrieved from the archaeological and historical studies and to navigate in these in order to test new methodologies or to practically evaluate the assessment. Virtual Reality (VR) technology makes it also possible to create applications for edutainment purposes for the general public and to integrate different learning approaches.

The building of three-dimensional renderings is an efficient way of storing information, a means to communicate a large amount of visual information and a tool for constructing collaborative worlds with a combination of different media and methods. By recreating or simulating something concerning an ancient culture, virtual heritage applications are a bridge between people of the ancient culture and modern users.

One of the best uses of the virtual models is that of creating a mental tool to help students learn about things and explore ancient cultures and places that no longer exist or that might be too dangerous or too expensive to visit. In addition, it allows students to interact in a new way, using many possibilities for collaboration. A very effective way to use VR to teach students about ancient cultures is to make them enter the virtual environment as a shared social space and allow them to play as members of that society.

The development technologies of video games are today driven by strong and ever-increasing request, but there are very few investments related to teaching usage of such technologies, they are still restricted to the entertainment context. Several VR applications in Cultural Heritage have been developed, but only very few of these with an edutainment aim.

The Human-Computer Interaction (HCI) technology is concerned with methodologies and methods for designing new interfaces and interaction techniques, for evaluating and comparing interfaces and developing descriptive and predictive models and theories of interaction.

The HCIs improve interactions between users and computers by making computers more usable and receptive to the user's needs.

Researches in HCI field focus on the developing of new design methodologies and new hardware devices and on exploring new paradigms and theories for the interaction. The end point in the interface design would then lead to a

 ISBN: 978-1-61208-117-5

paradigm in which the interaction with computers becomes similar to the one between human beings.

II. Previous Works

The techniques for navigation within virtual environments have covered a broad kind of approaches ranging from directly manipulating the environment with gestures of the hands, to indirectly navigating using hand-held widgets, to identifying some body gestures and to recognizing speech commands. Perhaps the most prevalent style of navigation control for virtual environments is directly manipulating the environment with gestures or movements of part of the user's body.

Some developed systems are based on a head-directed navigation technique in which the orientation of the users head determines the direction and speed of navigation [1]. This technique has the advantage of requiring no additional hardware besides a head tracker, but has the disadvantage that casual head motions when viewing a scene can be misinterpreted as navigation commands. In addition, a severe drawback of this and other head-based techniques is that it is impossible to perform the common and desirable real-world operation of moving in one direction while looking in another.

Another direct body-based navigation technique is found in some systems that use sensors to measure the tilt of the user's spine or the orientation of the user's torso in order to determine the direction of the motion and to enable the decoupling of the user's head orientation from their direction of movement [2].

Another category of techniques for motion control is based on speech recognition. Speech allows a user to indicate parameters of navigation and can often be used in conjunction with gestures to provide rich, natural immersive navigation controls [3]. Speech controls should play a role in virtual environment navigation, but it is also critical to support an effective navigation based on speech-free techniques.

In the last few years, systems based on locomotion interfaces and on control navigation by walking in place for the navigation in a virtual environment have also been developed.

String Walker [4] is a locomotion interface that uses eight strings actuated by motor-pulley mechanisms mounted on a turntable in order to cancel the displacement of the walker. String Walker enables users to maintain their positions while walking in various directions in virtual environments because, when the shoes move, the strings pull them in the opposite direction and cancel the step. The position of the walker is fixed in the real world by this computer-controlled tension of the strings that can pull the shoes in any direction, so the walker can perform a variety of gaits, including side-walking or backward walking

The CirculaFloor [5] locomotion interface uses a group of movable floors that employ a holonomic mechanism in order to achieve omni-directional motion. The circulation of the floors enables users to walk in arbitrary directions in a virtual environment while their positions are maintained. The CirculaFloor creates an infinite omni-directional surface using a set of movable tiles that provide a sufficient area for walking and a precision tracing of the foot position is not required. This method has the potential to create an uneven surface by mounting an up-and-down mechanism on each tile.

Powered Shoes [6] employs roller skates actuated by motors and flexible shafts and supports omni-directional walking, but the walker cannot perform a variety of gaits. Powered Shoes is a revolutionary advance for entertainment and simulation applications, because it provides the proprioceptive feedback of walking.

III. The MediaEvo Project

The MediaEvo Project aims to develop a multi-channel and multi-sensory platform in Cultural Heritage and to test new data processing technologies for the realization of a digital didactic game oriented to the knowledge of medieval history and society [7].

The game is intended as a means to experience a loyal representation of the possible scenarios (environments, characters and social roles) in the historic-geographical context of Otranto during Frederick Age (XIII century).

We chose Otranto as an example town; Otranto is located in the easternmost tip of the Italian peninsula, in Puglia, in the so-called Italy's heel,. Due to its geographical position, Otranto was like a bridge between East and West.

Otranto was a Byzantine and a Gothic centre, later ruled by the Normans, Swabians, the Anjou and the Aragonese. After a long siege, on 14 August 1480 the town was caught and the inhabitants were massacred by the Turkish army. This mix of history can be seen in the enigmatic mosaic of the Cathedral, a Romanesque church built during the Norman domination in the 10th century on the axis that joined Rome to Byzantium. The mosaic, done by the monk Pantaleone in the 12th century, covers almost the entire floor of the Cathedral, for over 16 metres; its size is nothing compared to the complexity of images and references that mixes Biblical narration from the Old and New Testaments with some pagan elements and others of Eastern derivation.

The implementation of an edutainment platform is strongly influenced by the definition of the scenery that is the world in which the framework is placed with the related learning objects and learning path, the characters, the scene's objects, the logic, hence, the rules of the game, the audio content, the texts and anything related to its use.

The framework will have features of strategy games, in which the decision capabilities of a user have a big impact

 ISBN: 978-1-61208-117-5

on the result, which in our case is the achievement of a learning target. Nevertheless, the strategy and tactics are in general opposed by unforeseeable factors (provided by the game), connected with the edutainment modules, in order to provide a higher level of participation, which is expressed in terms of the ease with which it is learnt. The idea is to provide a competition between the players, during their learning.

The system, on the basis of a well-defined learning target and eventually based on the knowledge of the user, will continuously propose a learning path (learning path composed by a sequence of learning objects), in order to allow the achievement of particular learning results.

A Digital Terrain Model (DTM) that has been produced using ESRI ArcGIS, containing all historical information like sea level, rivers, etc. It has been saved in .dif format and imported in the game engine.

For building and street modelling, we first used AutoCAD, 3ds Max, Cinema 4D. Characters and animation are made using 3ds Max.

Once defined a list of modular elementary residential units, according to the local medieval unit system, we composed the urban landscape in which monuments, infrastructures and situations are located.

For the building of the virtual environment we used the Torque Constructor editor of GarageGames in order to create 3D architectural contents for the Torque 3D engine. The choice of the Torque Constructor was prompted by technical considerations regarding the ability of software to perform a direct mapping of the files ".map", the compatibility level with the Torque Game Engine chosen to develop the game, the immediacy and the usability of internal tools.

The Torque Constructor has proved to be an efficient tool for the direct implementation of 3D graphics models. In particular, it has many geometrical tools for the graphic processing of the reality context and different controls to select the top of the structure or individual brush model.

All units made in the Torque Constructor have been imported into the Torque Game Engine. The initial testing step revealed several problems of navigability of the objects. These problems were related to the incompatibility between the domains of collision associated with the objects imported into the three-dimensional environment and the avatar.

Tests carried out have helped identify and solve these problems by setting the values associated to the collision domains and to the proportions between objects and avatars. At present, all units are properly imported and successfully navigated.

In the context of computer graphics for cultural heritage, a stable algorithm has been implemented to import CAD objects into the Torque Game Engine platform and to ensure navigation into each graphic structure. This technique together with an efficient system for exporting textures and paintings will be used to realize graphic complex environments for the 2D/3D reconstruction in cultural heritage [8].

Figure 1 shows some areas of the virtual reconstruction of the town of Otranto in the Middle Ages.

Figure 1. Parts of the virtual reconstruction of the town of Otranto.

IV. THE NINTENDO WII

Wii is the last console produced by Nintendo; it was released in October 2006 and, according to official data of 2010, has surpassed 70 million units sold. The reasons for this success can be undoubtedly found in the new approach that the gaming console gives the user in terms of interaction that effectively makes it usable and enjoyable by a large part of users. The secret of this usability is the innovative interaction system; the Wiimote (word obtained as a combination of "Wii" and "Remote") replaces the traditional gamepad controller type (with cross directional stick and several buttons) with a common object: the remote

 ISBN: 978-1-61208-117-5

control.

The Wiimote is provided with an infrared camera that can sense the infrared LED of a special bar (called "Sensor Bar") and it can interpret, by means of a built-in accelerometer, the movements of translation, rotation and tilt.

The Wiimote has been equipped with a series of accessories that increase its potential, such as the Balance Board, that, by means of four pressure sensors at each corner, is able to interpret the movements of the body in order to control the actions of the user in a videogame.

Figure 2 shows the interaction modalities of Wiimote and Balance Board.

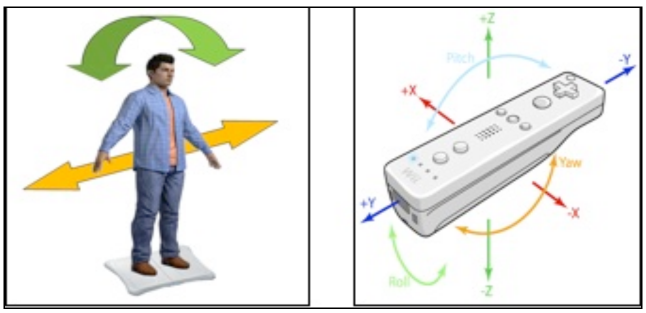


Figure 2. Interaction modalities of Wiimote and Balance Board.

Since the frequency of communication between the Wii console and the Wiimote/Balance Board are those of the standard Bluetooth, these devices can be used as tools to interact with any computer equipped with the same technology. Appropriate libraries have been realized in order to allow the interfacing between these devices and a computer.

V. THE DEVELOPED APPLICATION

This paper presents an application of navigation and interaction in a virtual environment using the Wiimote and the Balance Board of Nintendo. The aim is to make the interaction easier for users without any experience of navigation in a virtual world and more efficient for trained users; for this reason we need to use some intuitive input devices oriented to its purpose and that can increase the sense of immersion.

Because we walk on our feet, controlling walking in Virtual Reality could be felt as more natural when done with the feet than with other modes of input. For this reason we used the Nintendo Balance Board as input device for navigation that offers a new and accessible way to gain input. It is a low-cost interface that transmits via Bluetooth the sensor data to the computer and enables the calculation of the direction the user is leaning to.

In addition, in order to implement the control of different views and to change the point of view of the user, in our application we use the Nintendo Wiimote.

A software layer that allows to use the Balance Board and the Wiimote as input devices for any application that runs on a computer has been realized. The aim is to allow to receive signals and commands from the Wiimote and the Balance Board and to translate these into commands for the computer in order to emulate the keyboard and the mouse.

Figure 3 shows the use of Wiimote and Balance Board in the MediaEvo game.

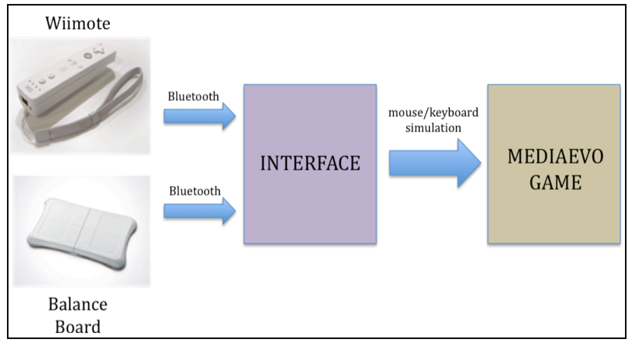


Figure 3. Use of Wiimote and Balance Board in the MediaEvo game.

The application, created to provide a new system of interaction in the virtual world of the MediaEvo project, can be coupled to any application of navigation in a virtual world.

To run the application, it is first necessary to configure the keys able to emulate any type of movement, to set the sensitivity of the Balance Board and then to connect the device; the information on the data received from the device are displayed in real time.

The interface that visualize the data received from the Wiimote and Balance Board is divided into two main sections: the left panels contain the control with all the data received via Bluetooth from the devices, whereas in the right side it is possible to set the associations among the command given to the device and the equivalent command simulated from the computer, the levels of sensitivity and threshold beyond which the interactions occur.

For these operations the software uses two open-source libraries in C# and the WiimoteLib InputSimulator; the WiimoteLib is a library for interfacing the Nintendo Wiimote and other devices (such as the Balance Board) in an environment .NET [9]. The purpose of this library within the application is to simulate the use of a mouse and a keyboard starting from the properly interpreted and translated inputs received from the Wiimote and Balance Board.

Regarding the interaction by means of the Wiimote, the aim is to simulate the mouse using two modalities of interaction.

"Mode 1" uses the movement on the X and Y axes of the accelerometer to move the mouse (and, in the 3D environment, the user's point of view) on the longitudinal and latitudinal axes; the value provided by the accelerometer is compared with the sensitivity set during configuration.

"Mode 2", that is the default mode, allows to move the mouse (and, then, the user's point of view) using the direction arrows of the Wiimote.

Figure 4 shows the configuration interface of the Wiimote and Balance Board devices.

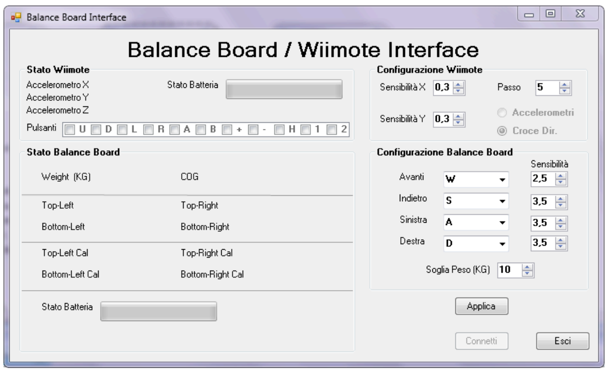


Figure 4. Configuration interface of the Wiimote and Balance Board.

The modalities of interaction provided by the application involve the use of the Wiimote and Balance Board simultaneously. In particular, the user is able to move the avatar in the virtual environment by tipping the scales in the direction where he wants to obtain the move; an imbalance in forward or reverse leads a movement forward or backward of the virtual character, while the lateral imbalance corresponds to the so-called "strafe" in video games, where the movement is made on the horizontal axis while maintaining a fixed pointing direction of the gaze.

The Wiimote, however, is used to impart the look direction of the character.

Figure 5 shows a user during the navigation in the MediaEvo virtual environment using the Wiimote and the Balance Board.

VI. CONCLUSIONS AND FUTURE WORK

This paper presents an application able to communicate with the Nintendo Wiimote and Balance Board and developed in order to provide a new modality of navigation and interaction in the virtual environment of the MediaEvo project. The aim of the MediaEvo Project is the development of a multi-channel and multi-sensory platform for the edutainment in Cultural Heritage.

Possible future developments could include the conversion of the application in external library, by adding specific methods and attributes to be directly integrated into other applications, and the porting of the developed application in a multi-platform language in order to be used in different development environments.

Figure 5. Navigation in the MediaEvo virtual environment.

REFERENCES

[1] A. Fuhrmann , D. Schmalstieg, and M. Gervautz, “Strolling through Cyberspace with Your Hands in Your Pockets: Head Directed Navigation in Virtual Environments”, the 4th Eurographics Workshop on Virtual Environments, Springer-Verlag, 1998, pp. 216-227.

[2] P. R. Cohen, M. Johnston, D. McGee, S. Oviatt, J. Pittman, I. Smith, L. Chen, and J. Clow, “QuickSet: Multimodal interaction for distributed applications”, the Fifth International Multimedia Conference (Multimedia ’97), ACM Press, 1997, pp. 31-40.

[3] D.A. Bowman, S. Coquillart, B. Froehlich, M. Hirose, and Y. Kitamura, “3D User Interfaces: Theory and Practice”, Boston, MA, Addison-Wesley, Pearson Education, 2005, pp. 342-344.

[4] H. Iwata, H. Yano, and M. Tomiyoshi, “String Walker”, International Conference on Computer Graphics and Interactive Techniques ACM SIGGRAPH 2007, August 2007, San Diego, California.

[5] H. Iwata, H. Yano, M., H. Fukushima, and H. Noma, “CirculaFloor”, IEEE Computer Graphics and Applications, Jan-Feb. 2005, Vol. 25 pp. 64 – 67.

[6] H. Iwata, H. Yano, M., and H. Tomioka, “Powered Shoes”, International Conference on Computer Graphics and Interactive Techniques ACM SIGGRAPH 2006, Boston, Massachusetts.

[7] L.T. De Paolis, M.G. Celentano, P. Vecchio, L. Oliva, and G. Aloisio, "Otranto in the Middle Ages: a Virtual Cultural Heritage Application", 10th VAST International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST2009), September 22-25, 2009, Malta.

[8] L.T. De Paolis, M.G. Celentano, P. Vecchio, L. Oliva, and G. Aloisio, " A Multi-Channel and Multi-Sensorial Platform for the Edutainment in Cultural Heritage", IADIS International Conference Web Virtual Reality and Three-Dimensional Worlds 2010, July 27-29, 2010, Freiburg, Germany.

[9] WiimoteLib - Managed Library for Nintendo Wii Remote, November 2008. Available: http://www.brianpeek.com.

Using Different Gestural-Input Methods for Personal and Public Touchscreen Devices

Jiyoung Kang
Graduate School of Culture Technology
KAIST
Seoul, Korea
e-mail: lov2jung@gmail.com

Jung-Hee Ryu
Graduate School of Culture Technology
KAIST
Seoul, Korea
e-mail: ryu@kgsm.kaist.ac.kr

***Abstract*— User interfaces for personal communication devices or public devices are a flourishing research area. This article begins with a brief history of the current user interfaces of personal communication devices and public devices. Key factors in introducing different types of interfaces for different types of devices are presented, including their experiment methodology. Important factors to consider are identified and elaborated, such as focus of attention, text-related symbols versus simple linear symbols, novice versus expert performance, stressful versus stressless process, and the speed-accuracy trade-off.**

Keywords- User interface; personal device; public device; text-related symbol; simple linear symbol

I. INTRODUCTION

Today, most input systems for user interfaces of personal or public devices ask users to input commands using a mouse or a keyboard with lots of toolbar buttons or menu items, or long addresses for selection. This clumsy type of user interface is inconvenient for small-screen devices such as PDAs, as it has no room to accommodate so many menu items and toolbars. Also, it is very time-consuming to type long addresses.

As such, user interfaces of different devices have changed in the past 50 years. To this day, many handy mobile devices use touchscreens. Computers and public kiosks are also starting to introduce touchscreen interfaces. Due to this, traditional input systems such as keyboards and mouse devices are being replaced with touch-based input systems. Ironically, this touch-based input system mimics the interaction of users with pen and paper, to which most persons are used from an early age. This system enables the users to interact with devices in efficient and natural ways [1]. Furthermore, even computer-unskilled persons can use their fingers efficiently with such devices. This makes touch-based interfaces suitable for a wide range of users and situations for which limited traditional manners of interaction could pose serious problems.

For this developed interface, a user-centered design approach is suggested. Although many devices are starting to use touch-based interfaces, there are no suitable interfaces for different purposes. For small-screen devices such as mobile touch phones and PDAs, which depend fully on pen- or finger-based user interfaces, traditional menu-selection or button-clicking interfaces or interfaces that require typing of long website addresses are inconvenient and useless with respect to fast and natural inputting. Even public kiosks in libraries or museums that use touch-based interfaces are encountering the same problems.

To avoid this problem, two types of gestural-input methods are suggested for devices with different purposes. Gesture based interfaces provide a new way for us to interact with devices, but also require us to make new decisions about which gestures we decide are usable and appropriate [2]. These decisions are based on the social and public settings where these devices are used on a daily basis.

II. BACKGROUND

Several methods are currently commonly used on mobile devices with touch-based input systems. Two of these methods are gestural-input systems. They use alphabet-related text inputs and simple linear symbol inputs, respectively. In the alphabet-related text input method, a user inputs the first letter of a website name or program. For example, a user can input 'g' to connect to the Google website or 'w' to open a Microsoft Word program. A user can also go to the next page of the website by merely drawing a line from the left to the right. This is an advanced input model that saves time by making it unnecessary to type the full website addresses or to click many icons to open a program. This convenient input system is featured in some handy programs such as Sensiva. This program was originally designed for mouse movement.

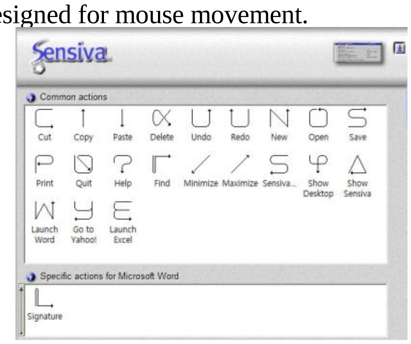


Figure 1. Sensiva

While the alphabet-related text input system is popular with some users, a more advanced input model is focused on in this paper, the simple linear symbol input method, which

 ISBN: 978-1-61208-117-5

is famous for small devices such as touch mobile phones (e.g., iPhone, Galaxy S). In the simple linear symbol input system, a user simply draws a single line in different ways to control the input.

These two sets of symbol inputs have their own strengths and weaknesses. For the alphabet-related text input user, if the user understands the logic behind the relationship between the initials and the symbols, s/he will find it very easy to infer other symbols that have not yet been studied. In some cases, however, it is more time-consuming and difficult to draw complicated symbols than simple linear symbols. On the other hand, even if it takes a shorter time to draw simple linear symbols, such input system would be more difficult for a user who could not infer any without adequate practice.

There have been similar researches for gestural-input systems, but they didn't concentrate to touchscreen devices or user's purpose. Gestural input is an interesting starting point when investigating the social factors of multimodal interface acceptance [3].

III. RELATED RESEARCH

A. *Sketch-based User Interface*

The sketch-based user interface was designed to improve clumsy and inconvenient user interfaces. In this interface, a user inputs composite graphic objects using the sketch-based user interface. With the sketches of a few constituent primitive shapes of the user-intended graphic object, the candidate graphic objects in the shape database are guessed and displayed in a ranked list according to their partial structural similarity to what the user has drawn. The user can then choose the right shape from the list and replace the sketch strokes with the exact graphic object with proper parameters, such as position, size, and angle. This user interface is natural for graphic inputs and is especially suitable for schematic designs [4].

B. *Calligraphic Interface*

CaliEdit is an application that is under development as a year-long undergraduate project at IST. It combines calligraphic interfaces with a traditional text editor. It allows users to manipulate text by directly drawing over it symbols that represent the most common editing tasks. It recognizes several common symbols that users draw on paper to represent their desired changes to the text, and effects those changes. The task analysis phase, with the help of questionnaires, identifies the most common symbols for the most common test correction tasks. The editor itself uses the CALI shape and gesture recognizer, ported to PalmOS, to help recognize those symbols [5].

C. *Unipad*

Unipad is a stylus-based text entry technique. It combines single-stroke text inputs with language-based acceleration techniques, including word completion, suffix completion, and frequent word prompting. In a study with 10 participants, entry rates averaged 11.6 wpm with 0.9% errors after two hours of practice. In follow-on sessions to establish the expert potential, four users entered "The quick brown fox" phrase repeatedly for four blocks of 15 minutes each. Average rates on the last block ranged from 17.1 to 35.1 wpm, and peak rates reached 48 wpm [6].

D. *Touchscreens for Public Kiosks*

To cater to a wide range of experiences with general public systems, it is important that little or no user experience is assumed for inputting in the system. Touchscreens provide a way of presenting keys or touch areas that can be changed for individual outputs [7].

Several years ago, the British public was less familiar with the concept of touchscreens and was less confident in using them than keyboards. Today, British users are more familiar with touchscreens. In the mentioned PD Web study (Maguire, 1997), 38 users were asked to rate how easy they found the use of touchscreens [8]. The answers follow.

TABLE 1 EASE OF USE OF TOUCHSCREENS

Ease of Use of Buttons and Commands				
Very easy	Easy	Medium	Difficult	Very difficult
23	11	2	2	2

As the results show, the large majority found touchscreens either 'easy' or 'very easy' to use. Touchscreens are therefore a flexible solution to inputting via kiosks.

IV. CUSTOMIZED SYMBOLS

In this experiment, symbols were created using Sensiva. The first set of symbols was for Test 1, and consisted of initials of programs or websites. Sensiva only works with one continuous line, so first letters of programs and websites were chosen and designed for to be drawn in one line. The second set of symbols was for Test 2, and consisted of simple linear symbols that could be drawn very easily and quickly. Those input symbols had their own outputs. Test 1 and Test 2 had their own method of symbols. Some of the symbols existed in the program, but most of the symbols were newly created for the tests.

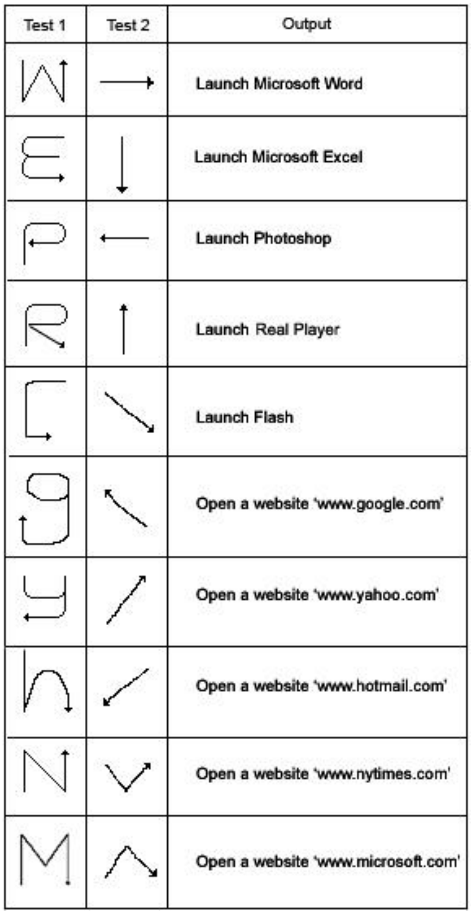


Figure 2. Symbols for the Tests

Ten different outputs were selected through the research for the tests. The ten sites that users most commonly used from their own systems (in English) were chosen. For each output, all the Test 1 symbols related to their initial letters were customized. For the Test 2 symbols, the simplest linear symbols were chosen for comparison. For Test 1, the participants easily guessed the other symbols after they understood the simple logic behind the meaning of the symbols. This is because all the symbols were related to their programs and sites' first letters. The Test 2 symbols, however, had no relationship to the outputs.

V. MEASUREMENT OF EFFICIENCY

To decide which set of symbols was suitable to different devices, a comparative study of the two methods was performed. This study consisted of an in-depth comparison of the two input methods to choose the most suitable interface for different purposes. There were 100 participants in the experiments, 10 for each age group. 52 percent of them were clerks who are familiar with computers. 25 percent were students. The last 33 percent of participants were housewives and self-employed. Among them 43 percent answered they own their touch screen devices such as laptops or mobile phones and 25 percent answered they are familiar with touch pad or screens and rests have rare experiences to the touch screen devices.

A. Choice of Platform

For the smooth use of the two sketch-based interfaces, a touchpad that was actually working was needed for the input system. A laptop computer with a built-in touchpad was used (Apple Macbook). To make the inputs actually work, the program Sensiva (Symbol Commander Software) was used, as it could customize the symbols for the experiments. To measure the time of use of the symbols, a digital camera took moving pictures of the process.

B. Stress Measurement

The study focused on the amount of stress that occurred when the users were using the two sets of symbols. According to a study by researchers from the New York Presbyterian Hospital/Weill Cornell Medical Center, work-related stress causes only a temporary rise in blood pressure [9]. Thus, to measure the amount of stress, the instant change in blood pressure during the experiment was regarded as the amount of stress from the process. An OMRON automatic digital sphygmomanometer was used for the measurement. First, the normal heartbeat of each participant was measured before the test. During the test, the second set of blood pressure changes was compared with the first set.

C. Time Measurement

The study also focused on the amount of time that was spent in using the symbols. From the recorded video, the time consumed while using the symbols was measured. This process had two steps. One was the analogy on the brain process time, and the other was the time that was spent for drawing the symbols. To distinguish between these two steps, the total time was divided into the recognition time and the drawing time.

1) Recognition Time

This is time in which the participants first read the given questions and started to think of what the correct symbols for them were. The participants needed a certain amount of time to think before they started to draw the symbols. This time can vary according to each person's ability to memorize or recall symbols.

2) Drawing Time

After the participants realized the correct symbols, they needed time to draw them on the touchpad. This time can vary according to each participant's ability to handle the touchpad.

3) Accuracy

To measure the accuracy of both sets of symbols, the correct answers were counted among the 10 given test subjects for both sets of symbols during the whole study period. If the participants drew similar symbols but the output did not work, it was considered a wrong answer. Only cases when the output worked correctly were counted as right answers.

4) Period of Study

To survey the changes in the test results, the participants were made to undergo three tests in four weeks. The first test

(Trial a) was given to each group of participants who memorized for 10 minutes the alphabet-related text input symbols. The second test (Trial b) was given after three days of studying the symbols for at least 10 minutes each day. The third test (Trial c) was given a week after the first test, and the participants were asked to study the symbols everyday for at least 10 minutes a day. Simple linear input symbols (Test 2) were tested in exactly the same way as was the first set.

After one month, with no obligation to study the symbols, participants tested one more time how much they remember the symbols.

VI. RESULTS

Test 1 (alphabet-related text input symbols) and Test 2 (simple linear input symbols) showed typical results for each section (i.e., stress, speed, accuracy, and study efficiency). These results showed which method is efficient for different devices. Average 1 means the average of Test 1, and Average 2 means the average of Test 2.

A. *Heartbeat Changes*

The beats per minute (bpm) in each test changed with time, especially in Test 2 (simple linear symbols), in which the average bpm decreased at a high speed. Test 1 also had a low diminution rate throughout the test. Finally, for Trial c (one week of study), Test 2 had a much lower bpm rate than Test 1.

For both sets of symbols, the users' heartbeat rates were close to their normal bpm (before the tests). In Trial a (the 10-minute study), Test 2's bpm had a +9.38 score against the normal bpm, slightly higher than Test 1's +8.30 score. This means that right after the 10-minute study, the users experienced much stress in the tests for both sets of symbols. The simple linear symbols that required more memorization time and effort were slightly more stressful.

After three days and seven days, however, the bpm scores dropped gradually. The Test 2 scores were obviously much lower in Trial 2 and Trial 3. This means the participants found it more comfortable to use the simple linear symbols after much practice and memorization.

On the other hand, when the participants used the alphabetical symbols, their bpm did not change. This is because the initial symbols are easy to guess, so users have no problem using them.

TABLE 2 CHANGES IN THE BPM IN TEST 1, TEST 2

Bpm	Normal	Trial a	Trial b	Trial c
Average 1 (Test 1)	79.18	87.48	87.3	86.16
Average 2 (Test 2)	76.66	86.04	83.9	80.88

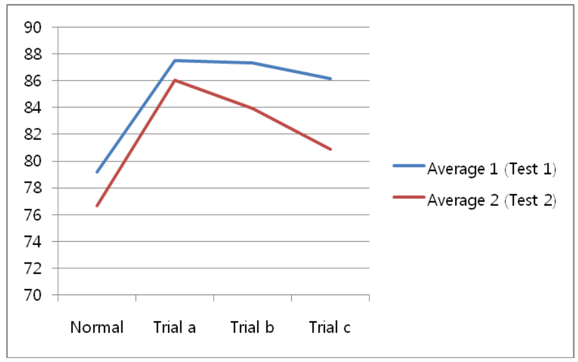


Figure 4. Comparison of the bpm Changes in Tests 1 and 2

The results of Tests 1 and 2 showed that once the participants became used to the symbols, they became less stressed than in the first trial (the 10-minute study). Also, the users were more comfortable with the simple linear symbols than the alphabetical symbols if they already knew the symbols well and have become used to them.

B. *Bpm Variation*

The decline of the bpm to the normal bpm also showed a noticeable decline in bpm variations. Test 1 had a flat average decline, but Test 2 had a noticeably steep falling line. Due to this, in Trial c, Test 2 had less bpm variations than Test 1. The former results also showed that in Test 2, when the participants were using the simple linear symbols, they were less stressed than in Test 1. They became more stressed when they analogized the symbols than when they just drew simple symbols unconsciously after studying them.

TABLE 3 VARIATIONS IN THE BPM

Variation (%)	Trial a	Trial b	Trial c
Average 1	10.64	10.40	8.94
Average 2	12.55	9.70	5.67

C. *Recognition Time*

While the participants were solving the given questions, the time they consumed was divided into two. The first period, defined as the "recognition time," was when the participants read the problem and memorized the symbols. The results of the two tests showed that the recognition time gradually decreased due to the time of study.

TABLE 4 AMOUNT OF RECOGNITION TIME

Recognition (sec)	Trial a	Trial b	Trial c
Average 1	2.25	2.13	1.98
Average 2	3.72	3.29	2.39

1) *Drawing Time*

The second period was the time when the symbols were drawn. The results also showed that the more the participants studied the symbols, the more became familiar with them. The participants were gradually able to draw the symbols more easily and quickly. The Test 2 symbols, which were

 ISBN: 978-1-61208-117-5

less complicated, took up a shorter time than the Test 1 symbols.

TABLE 5 AMOUNT OF TIME FOR DRAWING

Drawing (sec)	Trial a	Trial b	Trial c
Average 1	2.28	2.13	2.01
Average 2	1.56	1.36	1.25

2) The Total Time

To sum up, the recognition time and the drawing time were added to come up with the total time. It was found that the total time had the same aspect as the recognition time and the drawing time. The length of both time periods gradually decreased in the two tests. This means that the participants adjusted to both methods when they were studying for days, and that this helped reduce their study time.

One thing that is exceptional is that the total time in Test 2 caught up with the total time in Test 1 after one week of study (Trial c). In Trial a, Test 2 took a longer time than Test 1; but in Trial c, Test 2 took less time than Test 1. This means that once the participants became used to the simple linear symbols, they learned them faster than the alphabetical symbols.

TABLE 6 THE TOTAL TIME TAKEN FOR THE TESTS

Total Time (sec)	Trial a	Trial b	Trial c
Average 1	4.528	4.262	3.984
Average 2	5.288	4.654	3.513

D. Correct Answer Changes

The participants' answers to the 10 questions in each test improved with time. They got more correct answers when they studied the symbols for a week than for 10 minutes. In Test 1, the average number of correct answers (CA) was much higher than in Test 2 in the first trial (Trial a).

After a week of study, however, Test 2 saw a marked 90% increase in CA, which almost caught up with Test 1. This implies that if the participants studied the simple linear symbols much longer, they would have gotten more correct answers than they did with the alphabet-related symbols.

TABLE 7 CHANGES IN NUMBERS OF CORRECT ANSWERS

Correct Answers (%)	Trial a	Trial b	Trial c
Average 1	88.20	91.40	93.20
Average 2	60.60	76.60	90.60

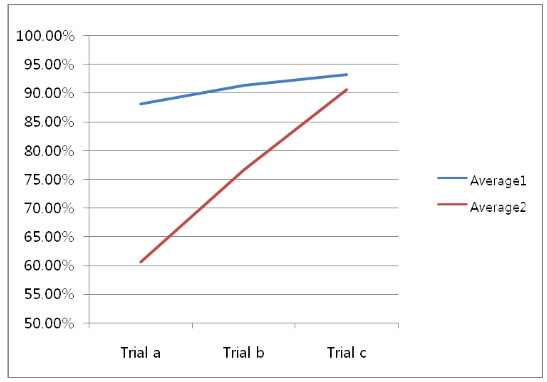


Figure 6. Changes in Numbers of Correct Answers

E. Experts vs. Novices

The participants' 43 percents owns their personal touch screen devices like laptops and mobile devices and they answered they are very familiar to those devices. And rests are barely use touch screens devices or have little experiences. Due to this fact, we divided participants to two groups. 43 percents who have many experiences to experts, and the others to novices. Interestingly the results show experts was noticeably superior for every experiments at the first trial but soon novices get to the similar point after one week. This shows that novices who were not familiar to the touch devices can be experts within few weeks after sufficient experiences.

TABLE 8 BPM COMPARISON OF EXPERTS AND NOVICS

BPM	Trial a	Trial b	Trial c
Experts	85.4	84.2	83.7
Novices	87.4	85.1	84.5

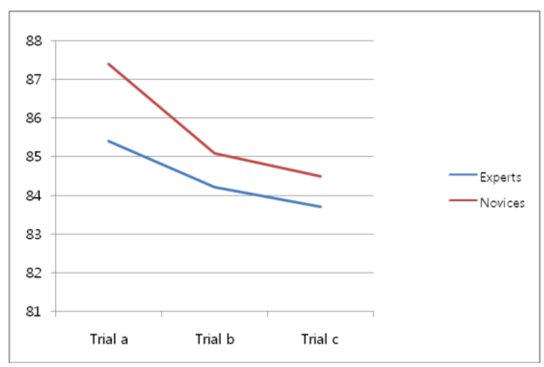


Figure 7. Bpm Comparison of Experts and Novices

TABLE 9 TIME COMPARISON OF EXPERTS AND NOVICES

Total Time (sec)	Trial a	Trial b	Trial c
Experts	4.125	4.024	3.872
Novices	5.52	4.484	3.925

 ISBN: 978-1-61208-117-5

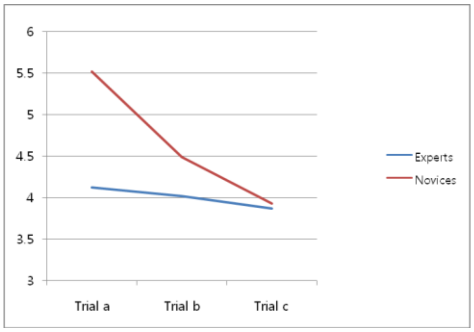


Figure 8. Time Comparison of Experts and Novices

TABLE 10 CORRECT ANSWER COMPARISON OF EXPERTS AND NOVICES

Correct Answer (%)	Trial a	Trial b	Trial c
Experts	85.00	88.50	92.70
Novices	72.00	74.60	89.30

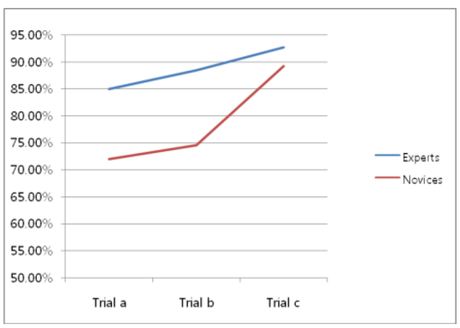


Figure 9. Correct Answer Comparison of Experts and Novices

F. Final Test

After the three experiments, participants have one month break from the study. Then we give the last test. The result show that even the simple linear symbols became superior at the third trial, it comes to the start again without continuous study of symbols. Without steady use of those symbols most participants barely remember the use of linear symbols. The result show the averages of after one month correct answer become almost the same level to the first trial. On the other hand, experts group has significant result that they recognize linear symbols better then the novices. Through this, we can suggest that experienced touchscreen devices users can memorize both symbols more easily and remember longer.

TABLE 11 CHANGES IN NUMBERS OF CORRECT ANSWERS AFTER ONE MONTH

Correct Answers (%)	Trial a	One month
Average 1	88.20	87.60
Average 2	60.60	65.78

TABLE 12 NUMBERS OF CORRECT ANSWERS FOR EXPERTS AND NOVICES AFTER ONE MONTH

Correct Answers (%)	Experts	Novices
Average 1	92.38	84.20
Average 2	79.27	68.32

VII. DICUSSION

The results showed the changes in the stress by heartbeat, speed, and accuracy with the time of study. These three elements all improved in time. Test 2 (simple linear symbols) especially saw a marked improvement in all the three tests. The stresses and the total time consumed remarkably decreased. The accuracy also increased step by step. The alphabetical symbols seemed more efficient in the first trial (after a 10-minute study) because the participants were able to guess the symbols without memorizing them. The more they learn the Test 2 symbols, however, the more efficient they became with such symbols than with the alphabetical symbols. Through their repeated study of the symbols, they were able to memorize the Test 2 symbols, after which the said symbols became easier to use. The experts who are familiar to the touchscreen devices get more correct answers, take shorter time and get less stress at first. However, novices get nearly equivalent scores after one week of study. At the last experiment, novices remembered both symbols less than experts.

VIII. DICUSSION

Two methods of improving gestural-input symbol interfaces for different devices were presented in this paper. As the results showed, the two types of symbols have their own strengths and weaknesses. The alphabet-related symbols, which are initials of programs or site names, are easy to learn and guess in the case of new users who are not familiar with them. In Trial 1, even though the participants studied the symbols for only 10 minutes, they got a high rate of accuracy. After continuously studying the simple linear symbols, however, the participants became faster and more efficient with them. They became less stressed after they became used to the simple symbols. Participants who are experts to the touch screen devices get much higher scores for accuracy and need shorter time and little stress. Also, they remember symbols longer than novices.

Due to this, different gestural input methods for different devices are presented. We provide a new way to interact with touch screen devices upon the users' usability and appropriation.

For personal and small devices such as mobile phones, PDAs, and PMPs (Portable Media Players), simple linear symbols can be more practical and efficient. Users are always carrying these devices and spend much time using them. These are very intimate devices. This means that users can spare sufficient time to study the symbols these devices use. Once they become used to the simple symbols, they become very efficient with them.

 ISBN: 978-1-61208-117-5

On the other hand, for public devices such as information kiosks in transportation stations or museum guide devices, alphabet-related symbols will be more efficient. Those devices have to be easy to use for both novices and experts. Once the users know the initial logic, they can easily and readily use any device.

At present, it is assumed that all users have certain abilities to study the symbols. In future works, this work may be extended by dividing users into specific ages and sexes for a usability test, to see how their abilities will change. A re-customization of the symbols is also planned for their optimum usability.

REFERENCES

[1] Masui, Toshiyuki, "An Efficient Text Input Method for Pen-based Computers," In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 328-335, April 18-23, 1998, Los Angeles, California, USA.

[2] Rico, J., Brewster, S. Gestures all around us, "User differences in social acceptability perceptions of gesture based interfaces," In Proc. MobileHCI'09, ACM Press, pp. 1-2, 2009.

[3] Julie Rico, "Evaluating the social acceptability of multimodal mobile interactions," CHI 2010, April 10-15, 2010, Atlanta, Georgia, USA.

[4] MacKenzie, I. S.; Chen, J.; and Oniszczak, "A. Unipad: Single-stroke Text Entry with Language-based Acceleration," In Proceedings of the 4th Nordic Conference on Human-Computer Systems, 2006.

[5] Pedro Miguel,Coelho Gonçalo,França Faria,Av. Rovisco Pais,Daniel Gonçalves and Joaquim A Jorge, "Using Sketches to Edit Electronic Documents." Scientific Literature Digital Library and Search. 2007.USA.

[6] Liu Wenyin, Xiangyu Jin and Zhengxing Sun. "Sketch-Based User Interface for Inputting Graphic Objects on Small Screen Devices," Graphics Recognition Algorithms and Applications Lecture Notes in Computer Science, Volume 2390/2002, pp. 67-80, 2002.

[7] May, A. J, "Development of Genetic User Requirements for Interactive Public Information Systems," Master of Science thesis in Ergonomics, Faculty of Science, University College, London, 1993.

[8] Maguire, M. C, "A Review of User Interface Design Guidelines for Public Information Kiosk Systems," International Journal of Human-Computer Studies, Elms Grove, Loughborough, Leics, LE11 1RG 1RG, UK, 1999.

[9] Mann, Samuel J, "Job Stress and Blood Pressure: A Critical Appraisal of Reported Studies," Current Hypertension Reviews. Vol. 2, pp. 127-138, 2006.

 ISBN: 978-1-61208-117-5

From Individual to Collective Personas

Modeling Realistic Groups and Communities of Users (and not Only Realistic Individual Users)

Alain Giboin
INRIA Sophia Antipolis - Méditerranée
Team-Project Edelweiss
Sophia Antipolis, France
Alain.Giboin@inria.fr

***Abstract*— Personas are user models that are represented as specific, realistic humans. Initially focused on the modeling of individual users, the Persona method (see, e.g., Cooper) is gradually changing towards the inclusion of collectives of users (groups, communities, etc.). In other words, a *"Collective Personas"* trend is emerging. This paper reports a literature review reflecting this emerging trend. It synthesizes some issues and avenues related to collective personas development.**

Keywords- personas, collective personas, interaction design, social interaction design, user modeling, group modeling

I. INTRODUCTION

Personas are user models that are represented as specific, realistic humans. Initially focused on the modeling of individual users, the Persona method (see, e.g., Cooper [2][3], Pruitt and Grudin [15]), also called method of "individual personas" [14], is gradually changing towards the inclusion of collectives of users (groups, communities, etc.). In other words, a *"Collective Personas"* trend is emerging, as evidenced by a number of works attempting to extend the Individual Persona method to collectives. In this paper, after a short reminder on Individual Personas, we report a chronological and comparative literature review of the works we consider to belong to the Collective Personas trend. We conclude the review by synthesizing some issues and avenues related to the development of collective personas.

II. INDIVIDUAL PERSONAS: A SHORT REMINDER

Our presentation of Individual Personas will be mainly based on the Cooper's founding Persona method ([2][3]). For a review of existing Individual Persona methods, see [5].

A. Individual Persona Definition and Construction

Personas are "user models that are represented as specific, individual humans" [3]. In other words, personas are fictional *personifications* [3] which represent *realistic* individual persons [14]. This realism is heightened by identifying the persona by a name (e.g., Mike) and a photo. Personas are derived from significant *behavior patterns* (i.e., sets of behavioral variables) elicited from interviews with and observations of users (and sometimes customers) of the future product. *Behavioral variables* are axes or ranges (e.g., "necessity-only" vs. "entertainment") across which product use is segmented. For each pattern identified, information about related users' characteristics (e.g., goals, attitudes, activity flow) is synthesized in order to reveal personas. To translate the knowledge about users/personas into a user-oriented design solution, personas are role-played as the characters of *scenarios* of product use.

B. The Collective Aspects in Individual Personas

The typical case for individual personas, Cooper and Reimann [3] claimed, is "to be completely unrelated to each other and often from completely different geographic locations and social groups." However, Cooper and Reimann acknowledged, it sometimes "makes sense" for personas "to be part of the same family or corporation and to have interpersonal or social relationships with each other." This acknowledgment shows that the collective aspects are not completely absent from the Individual Personas method. Nevertheless, the emphasis is not on the collectives to whom individuals belong, but on the relationships between the individuals, so that the notion of Collective Persona is not made explicit, as opposed to the works described below.

III. COLLECTIVE PERSONAS: A CHRONOLOGY

The notion of a Collective Persona appeared explicitly in six works that we will present chronologically in this section. For each work, we will report its motivation, its genre (researcher's work, practitioner's work, marketer's work), its contents (collective persona definition, persona construction method, etc.), and its applications (when all this information is provided by the authors of the works).

A. 2004: Group Personas$_1$

To our knowledge, the first explicit work on collective personas is Kuniavsky's work [11], a practitioner's work. It describes how Kuniavsky and his students-practitioners constructed the *Group Persona* method while respectively training, and being trained to, the original Cooper's individual persona technique. Students were told to apply the technique to the design of some wearable/portable technology for people to use in an amusement park. In the course of elaborating individual personas and their related scenarios, Kuniavsky and his students realized that the technology will be simultaneously used by two or more people forming a coherent group. They consequently got the idea of modeling such groups of people as a *group persona*, rather than as individual users, and of modeling their goals as group goals (defined as a negotiated combination of individual goals).

Kuniavsky and his students developed a 4-step method:

1) *Making rough outlines of the clusters of people* to focus on, and giving them distinctive names (e.g., "Young Parents, Young Kids;" "College-age Friends").

2) *Defining axes along which situating the group* (e.g., for the "Young Parents, Young Kids" persona: NUMBER OF PEOPLE IN GROUP: 5; [KIND OF] PEOPLE IN GROUP: 2 adults, 2 kids ages 3-10, grandparent;etc.).

3) *Iteratively creating personas:* (a) Roughly sketching the persona; (b) Brainstorming the persona details; (c) Editing the persona description (cutting the irrelevant details); (d) Writing preliminary scenarios; (e) Tuning the personas according to the scenarios.

4) *Finalizing the personas:* Finalizing the fleshing-out of group personas (e.g., The Ancona Family) and related scenarios felt to be typical examples of groups who visited the park and how they would behave.

Kuniavsky suggested applying the group persona method to "groupware" such as entertainment, education, and collaboration software.

B. 2006: Organizational Personas$_1$

A second explicit work dealing with collective personas was achieved at "Cooper" (the corporation founded by Alan Cooper). This is a practitioner's work on the so-called *Organizational Personas*. Such personas are created when the problem the practitioners are working on "is in the domain of a complex, multi-user, business system" [4]. Organizational personas "are fictional organizations that represent certain key characteristics of the (visited) companies." They "highlight the patterns and objectives of the kind of organization that requires this type of complex system." Practitioners from Cooper position their user personas in "relevant roles" within their organizational personas, to model the functioning of the system they are examining.

C. 2006: Communitas & 2009: Collective Personas

A third explicit work on collective personas is our own work, a researcher's work. This work was initiated in 2006 in order to extend the Cooper's persona technique to the design of intranets as collective tools [8]. This work was interrupted in 2007 (because it was not a priority for our team), and resumed in 2009, in a different context: the design of social semantic collaborative tools for assisting corporate intelligence tasks [9][6].

Communitas. The idea of addressing Collective Personas is born under the USABLEINTRANET project. The idea came from the intersection of two research goals: (1) to elaborate individual personas representing intranet users; (2) to elaborate collective scenarios of intranet use. Knowing that in the scenario-based design literature, a distinction was made between individual and collective scenarios, we thought we might as well distinguish individual and collective personas, so that we could develop intranets better adapted to the collectives who use them, and not only to individual users.

We reviewed the literature, and discovered only the two works on Group Personas$_1$ and Organizational Personas$_1$. We tried to find the most appropriate name to refer to the notion of a "collective persona." We found a Latin name *à la* "Persona," namely *Communitas*. We selected this word because it referred to: (a) a notion increasingly employed at the time by researchers interested in collective system design, the notion of a "community" (community of practice, community of interest, etc.); (b) definitions that we wanted to convey, such as "an unstructured community in which people are equal" (Wikipedia).

The work had stopped at the time to setting the following research agenda: (1) to continue the literature review on collective personas, (2) to further analyze Kuniavsky's "Group Persona" concept and method, (3) to further analyze the collective aspects appearing in individual personas, (4) to feed the collective personas method by elements of the scenario method, (5) to explicit the links between individual and collective personas.

Collective Personas. When we resumed our work in 2009 under the French ANR project ISICIL, we decided to abandon the term "Communitas," because it seemed too restrictive and too community-connoted. We choose a more neutral and generic term, that of *Collective Personas*. We updated the literature review on collective personas, and extended it to collective scenarios. To type the collectives and their characteristics, we relied on the literature on collective models—models describing standard collectives (groups, teams, departments, companies) and non standard collectives (online communities, social networks).

We elaborated a 4-step method: *(1) Identifying "directing contexts"* (i.e., the contexts orienting the collective and individual personas' or actors' activities to be supported by the system), and the actors carrying the directing contexts. *(2a) Identifying the critical individual and collective actors* and their critical characteristics (goals, roles, tools, etc.) through interviews and observations; concerning the collective actors, the dimensions considered were among others: formal ↔ informal, sustainable ↔ ephemeral, intra-organizational ↔ extra-organizational. *(2b) Identifying the relationships that users maintain*. *(3) Identifying the current and future scenarios involving the interacting actors*. *(4) Translating scenarios into story-boards*.

This method has been used in particular for analyzing practices and needs of the individuals and collectives performing corporate intelligence activities within an Environment and Energy Management Agency.

D. 2008: Group Personas$_2$

A fourth explicit work on collective personas is also a researcher's work [1]. The focus of this work was the visualization of *Group Personas* through Web 2.0 collaborative software. This software "analyzes data from social networks and measures the feelings, perceptions, and activities of the group members and displays a summary of these measures." Group Persona Visualization is intended "to inspire communication and collaboration among groups in which status information is often fragmented across a wide variety of Web locations."

 ISBN: 978-1-61208-117-5

E. 2010: Organizational Personas$_2$ & Persona Ecosystems

A fifth explicit work on collective personas is a marketer's work [10]. It introduces the very interesting notion of a *Persona Ecosystem*, which is an "additionally refined organizational persona." An *Organizational Persona* is "an archetypal model of a segment of customer or partner companies." It "provides characteristics of a typical organization within a given segment, and the goals of that organization as a whole." A *Persona Ecosystem* is "an additionally refined organizational persona that uses business role personas to model typical, relevant, title roles within an organization, their relationships to each other, their tasks that relate to [the] company's concerns, and their roles in a buying decision." This level of detail provides "information about how role relationships impact decisions." "Buyers," "product users," and "influencers" or "stakeholders" are typical role persona groupings.

Note that Microsoft Research perspective on personas resonates with the idea of persona role since it implements an approach to systems design tailored by users' roles. For example, Miller and Williams [13] elaborated the so-called "MSF Agile Persona Template" with a *Role* slot "to place the user group in which the persona belongs." In this example, however, the notion of a "group" doesn't resonate with the notion of an "organization." It refers to a role type, e.g. "normal user" or "administrator."

F. 2010: Collaboration Personas

The sixth and final explicit work to date about collective personas is a researcher's work by Matthews, Whittaker, Moran, and Yang [12]. The tools targeted by Matthews et al. are intended to support Collective Intelligence in organizations.

Collaboration personas are empirically derived descriptions of hypothetical groups of people with specific qualities, goals, and needs. They are grounded in multiple empirical studies and CSCW research literature. Collaboration personas are a super-structure that includes individual personas and describes their actions in the context of collaboration. Collaboration personas are inherently about interactions, and thus should describe the actions, decisions, and norms that have led to stable interactions.

Matthews et al. propose a framework—also named *Collaboration Personas*—for *(a)* helping users choose and adopt appropriate workplace collaboration tools and *(b)* design new collaborative tools that better meet the needs of typical groups. The framework describes distinct types of collaborations common to global companies used for large-scale efforts (communities of interest/practice and task forces) and small-scale efforts (traditional team-focused uses of collaborative tools). The main element of the framework is a set of dimensions or variables (also called "collaboration qualities") along which groups can be situated: (a) *a purposeful dimension* (SHARED ELEMENT: shared interest ↔ shared objective), (b) *a temporal dimension* (GROUP LIFESPAN: short-lived ↔ long-lived), and (c) *five compositional dimensions* (SIZE: small ↔ large; SUB-GROUPS: no subgroups ↔ many subgroups; PERSONNEL: stable ↔ dynamic; MEMBERS DEPENDENCY: interdependent ↔ independent; GROUP MANAGEMENT: self-managed ↔ designated leader).

Collaborative groups have dynamics, roles for individuals (e.g. regular member, super-team leader), and collaboration phases (e.g., Starting, Planning, Executing, and Reporting) that should be considered. Collaboration personas are supposed to make these various aspects visible and understandable.

Matthews et al. reported how collaboration personas might be applied to a core collaborative intelligence problem: supporting communities of practice with tools such as Lotus Activities (a tool helping teams organize their work, and tap their professional networks to help execute), Lotus Quickr (a "Team space" Web tool), and Lotus Communities (a Web tool helping people with similar interests or job responsibilities share information).

IV. COLLECTIVE PERSONAS: A COMPARISON

The works on collective personas can be compared along the dimensions which characterize the "collective persona" notion/artifact, and the method for building collective personas as described in the reviewed works. The comparison is just outlined here.

A. Comparing the "Collective Persona" Notions/Artefacts

Persona name: The terms used to name collective personas focus either on the collective as an entity (e.g., "{group|organizational|collective} persona", "communitas") or on the activity performed by the collective (e.g., "collaboration persona").

Persona definition: Collective personas are defined either as personas belonging to a collective (see "organizational personas$_2$" and "persona ecosystem", "group personas$_2$") or as an entities composed of individuals (see "communitas", "collaboration personas").

Persona variables: The variables used to define collective personas are individual personas variables transposed to collective personas (e.g., goals), or variables coming from existing models of collectives, etc.

Collectives considered: Some works focus on specific collectives (e.g. families or groups of friends in group personas$_1$) while others cover a wider range of collectives (see collective personas and collaboration personas).

Realism of personas: The level of realism of the collective personas as entities varies among the presented works. However this level has not reached the realism level of individual personas.

B. Comparing the "Collective Personas" Methods

Origin of the method: The methods originate from a practical need (e.g. group personas$_1$) or from a theoretical motivation, or a mixture of both (e.g., collective personas).

Aim of the method: Most of the methods aim at informing or supporting the design of explicitly collective tools (groupware, multi-user business system, intranets, Web 2.0 tools), or of individual tools that can be used within groups.

Availability and accessibility of the method: Some works proposed a method while others do not; when a method is available, it is not always accessible. When accessible, the level of detail of the method description differs.

Degree of achievement of the method: The methods presented can be mainly considered as initial versions needing to be improved.

Place of the scenarios in the method: Most of the time, scenarios have a greater place in the methods; the main reason is that scenarios make explicit the interactions between the individual personas acting within a collective.

V. Collective Personas: Issues and Avenues

To conclude this literature review illustrating the emergence of the "Collective Personas" trend, we would like to synthesize some issues and avenues related to the development of collective personas.

Issues. We have seen that when trying to design collective tools, designers are faced with a double need: *(1)* to characterize the collectives for whom these tools are intended and *(2)* to get a method for doing such a characterization or modeling. This double need raises several *issues* (some of them have been identified by the authors quoted in this paper), e.g.:

- Do we need to elaborate collective personas or is it enough to develop the collective dimensions of individual personas (e.g., interpersonal or social relationships, roles, etc.), and/or to develop the collective scenarios related to individual personas acting collectively? This issue may arise in cases where the collective is extremely short-lived.
- What are the features/dimensions allowing to characterize collective personas realistically? Which features/dimensions can we transfer from individual personas to collective personas? For example, what kind of image can represent a collective: a collection of individual photos or a group photo?
- How can we account for the links between individual and collective personas? Can we attach the same individual persona to a user playing a role in collectives referring to two different collective personas?
- What kinds of scenarios can we attach to collective personas?

Avenues. The issues listed above, and many others, clearly need to be further explored. For this exploration, it would be beneficial to rely on modeling techniques coming from communities close to the HCI community, especially the User Modeling and Groupware/CSCW communities. For example, we could rely on: (a) *Models of "collectives:"* communities, social networks, groups, teams, families; (b) *Collective scenarios techniques*, an extension of individual scenarios technique; (c) *Group modeling techniques*, an extension to groups of the classical user modeling techniques dedicated to individuals. These techniques aim to develop "models of groups, collaboration and communities [which] collect and structure the rich information describing interactions between users" [7]. We are currently exploring these avenues.

Acknowledgment

This work was partially supported by grants from INRIA COLOR Action UsableIntranet and ANR Project ISICIL. The author thanks Sophie de Bonis, Sandra Belfils, and Anne-Laure Négri for stimulating discussions about persona-based design, and anonymous reviewers for their constructive comments on an earlier version of this paper.

References

[1] A. Armstrong, A. Bartlett, A., M. Taylor, and A. Burman. *Group Persona Visualization*, IBM Hursley, 2008. Last access December 2010 from: services.alphaworks.ibm.com/groupvis/

[2] A. Cooper. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How To Restore the Sanity*. Indianapolis, IN: Macmillan, 1999.

[3] A. Cooper and R.M. Reimann. *About Face 2.0: The Essentials of Interaction Design*, John Wiley and Sons, 2003.

[4] Cooper Corp. Modelling [Phase of the Interaction Design Process]. Retrieved November 2006 from (no longer accessible): www.cooper.com/content/why_cooper/modeling.asp

[5] I.R. Floyd, M.C. Jones, and M.B. Twidale. Resolving Incommensurable Debates: A Preliminary Identification of Persona Kinds, Attributes, and Characteristics. *Artifact, 2(1)*: 12-26, 2008.

[6] F. Gandon et al. ISICIL: Information Semantic Integration through Communities of Intelligence Online, *Proc. International Workshop on Web Intelligence and Virtual Enterprise (WIVE),* at 10th IFIP Working Conference on Virtual Enterprises (PRO-VE'09), Thessaloniki, Greece, 2009, pp. 1-25.

[7] E. Gaudioso, A. Soller, and J. Vassileva (Eds). User Modeling to Support Groups, Communities and Collaboration. *User Modeling and User-Adapted Interaction* 16, Special Issue, 2006.

[8] A. Giboin. Des "personas" aux "communitas" ou comment représenter des collectifs concrets particuliers pour aider à la conception de systèmes destinés à des collectifs, INRIA Sophia Antipolis, INRIA Color Action "UsableIntranet", Working Document, December 2006.

[9] A. Giboin, M. Leitzelman, E. Soulier, F. Bugeaud, V. Le Meur, F. Herledan, and A. Merle. *Analyse des pratiques et des exigences pour les utilisateurs d'ISICIL*, ANR Project ISICIL, ANR-08-CORD-011-05, Deliverable ISICIL-ANR-RA01, 2009.

[10] Goal Centric Corp. *What Is a Persona Ecosystem?* 2010. Last access December 2010 from: www.goalcentric.com/contentdisplay.asp?id=1007

[11] M. Kuniavsky. *Extending a Technique: Group Personas*, 2004, Last Access December 2010 from: www.boxesandarrows.com/view/extending_a_technique_group_personas

[12] T. Matthews, S. Whittaker, T. Moran, and M. Yang. Collaboration Personas: A Framework for Understanding and Designing Collaborative Workplace Tools. In *Workshop Collective Intelligence in Organizations: Toward a Research Agenda* at Computer Supported Cooperative Work (CSCW), 2010, pp. 1-8.

[13] G. Miller and L. Williams. Personas: Moving Beyond Role-based requirements Engineering. Microsoft and North Carolina State University, 2006.

[14] D.A. Norman. Ad hoc Personas and Empathetic Focus. In J. Pruitt & T. Adlin (Eds.). *The Persona Lifecycle: Keeping People in Mind during Product Design* (pp. 154-157). Amsterdam, Netherlands: Morgan Kaufmann Press; Elsevier, 2006.

[15] J. Pruitt and J. Grudin. Personas: Practice and Theory. Proceedings of the 2003 conference on Designing for user experiences. New York: ACM Press, 2003, pp 1-15.

Enhanced Stability of Three-Users Multirate Distributed Haptic Cooperation via Coordination to Average Peer Position

Ramtin Rakhsha
Department of Mechanical Engineering
University of Victoria
Victoria, BC, V8W 3P6, Canada
rrakhsha@uvic.ca

Daniela Constantinescu
Department of Mechanical Engineering
University of Victoria
Victoria, BC, V8W 3P6, Canada
danielac@me.uvic.ca

***Abstract*—Distributed networked haptic cooperation may become unstable when the number of interacting users increases because the effective coordination gain for the shared virtual object increases. The average position coordination strategy maintains the coordination gain of the shared virtual object constant regardless of the number of cooperating participants. Therefore, the average position strategy is expected to increase the stability region of networked haptic cooperation among multiple users. This paper confirms through analysis and experiments that AP coordination maintains the three-users haptic cooperation stable for larger coordination gains than traditional virtual coupling coordination. The stability analysis is performed in a multirate control framework. Multirate control is deployed to support high sampling rate of the peer force feedback loops in the presence of a low network update rate. The experiments report a one degree of freedom manipulation of a virtual cube by three cooperating users.**

Keywords-Networked haptic cooperation; distributed control; multirate control; coordination to averaged position.

I. Introduction

The rapid growth of the computer networking infrastructure has inspired haptics researchers to develop methods to add the sense of touch to applications like defence [6], cooperative industrial design [10], surgical teletraining [18], telerehabilitation [22], emotion recognition [3] and multi-user on-line computer games [12]. Such applications demand that multiple networked users be enabled to cooperate in a shared virtual environment, i.e., be enabled to simultaneously interact with the virtual environment and with each other. Simultaneous haptic interaction in a shared virtual environment can be supported through client-server communications and centralized control of the interaction, or through peer-to-peer communications and distributed haptic control. The client-server connectivity maintains better consistency among the interacting users and is typically deployed to support cooperation among more than two users [4], [5], [16]. Nonetheless, it incurs double communication delay and can render only much lower contact stiffness even for small network delay [9]. Peer-to-peer communications and distributed control of the interaction are preferable for realistic force feedback in rigid shared virtual environments [9].

Existing architectures for distributed control of networked haptic cooperation [9], [19], [20] typically coordinate each peer's local copy of the shared virtual object (SVO) to all other SVO copies. The coordination of each SVO copy to the other SVO copies may lead to instability because the impedances of the many coordinating controllers compound and may exceed the Z-width [8] of the users' haptic interfaces, i.e., the maximum impedance that the interfaces can stably display to their respective users. However, little research addresses this difficulty. Recent work in [14], [15] introduces a passivity-based framework for optimizing the distributed connectivity and for computing virtual coupling [1], [7] parameters that guarantee stable cooperation among a fixed number of operators. The work in [14], [15] hinges on a passive mechanical integrator [13] whose extension to rigid body interaction is unclear.

In prior work [17], we have introduced a distributed haptic control architecture whose SVO coordination gain at all peer sites is independent of the number of participants involved in the cooperation. In that architecture, the local SVO copy at each user has been coordinated to a SVO representative whose motion has been computed by averaging the motion of all other SVO copies. In the preliminary study [17], the average position (AP) coordination strategy has been contrasted through simulations to virtual coupling coordination. In this paper, we present analytical and experimental support that AP coordination maintains three-users networked haptic cooperation stable for larger coordination gains than traditional virtual coupling coordination. The paper considers operators connected across a network with limited bandwidth and constant and relatively small communication delay [9]. It deploys multirate control in order to support a high sampling rate of the peer force feedback loops in the presence of a low network update rate. The paper uses lifting [2], [9] to derive the state transition matrix of the multirate three-users haptic cooperation system with AP coordination and to carry out an eigenvalue-based analysis of its stability. Experimental one degree of freedom (DOF) cooperative manipulations of a virtual cube by three cooperating users are presented to validate the analytical results. Although the analysis and experiments presented in the paper involve one DOF networked haptic cooperation, the AP coordination strategy is readily applicable to cooperative rigid body manipulations. This is because the AP coordination is independent of the SVO simulation algorithms and because virtual coupling control is suitable for rigid body coordination [8].

Section II overviews the traditional virtual coupling and the AP coordination strategies for three-users networked haptic cooperation. Section III presents the steps involved in developing the multirate state space model of three-users haptic cooperation across a network with low update rate and small and constant network delay, and uses the multirate state transition matrix to compare the stability regions of the two coordination techniques. Section IV validates the analytical results through experiments in which three networked users manipulate a shared virtual cube together. Section V presents the conclusions drawn from this work and the directions for future work.

II. Coordination of Three-Users Distributed Networked Haptic Cooperation

In distributed networked haptic cooperation, the users interact through manipulating a SVO together. The control of such cooperation is achieved through coordinating all peers' local copies of the SVO. Virtual coupling control [1], [7] has typically been selected to coordinate the SVO copies [9], [19], [20]. The extension of virtual coupling coordination to three-users networked haptic cooperation is schematically depicted in Fig. 1. This figure illustrates that the virtual environment at each peer includes a copy of the SVO and an avatar of the peer's haptic interface. The SVO damping is assigned to each local copy, and the SVO mass is equally divided among all copies. The avatar inherits the dynamics of the haptic interface. The dynamics of two-users distributed networked haptic cooperation with virtual coupling coordination are presented in [21], and their extension to three-users interaction inserts a virtual coupler between each pair of peers, as shown in Fig. 1. One problem with this control architecture is that the effective coordination gain for each SVO copy becomes larger as the number of users increases and may exceed for Z-width of the haptic interface and thus, destabilize the cooperation. To address this shortcoming, we have proposed the AP coordination strategy [17].

In the AP scheme, the SVO copy of each peer is coordinated to the average position of all other SVO copies. The AP coordination strategy is illustrated schematically in Fig. 2 for the local SVO copy of Peer 1 and for three-users networked haptic cooperation. In this figure, notation is used as follows: m_{HD1} and b_{HD1} are the mass and damping of the haptic device of Peer 1; m_{O1} and b_{O1} are the mass and damping of the SVO copy of Peer 1's; K_{C1} and B_{C1} are the stiffness and damping of the contact between Peer 1 and its local SVO copy, respectively; K_{T} and B_{T} are the stiffness and damping gains of the virtual coupler which coordinates the distributed SVO copies. As in traditional virtual coupling coordination, the SVO mass m_{O} is equally distributed among the SVO copies and the SVO damping b_{O} is assigned to each SVO copy, i.e., $m_{\text{O1}} = \frac{m_{\text{O}_1}}{3}$ and $b_{\text{O1}} = b_{\text{O}}$.

According to Fig. 2, the dynamics of three-users distributed networked haptic cooperation with AP coordination are:

- for the haptic device of Peer i:

$$m_{\text{HD}i}\ddot{x}_{\text{HD}i} + b_{\text{HD}i}\dot{x}_{\text{HD}i} = F_{\text{h}i} - F_{\text{C}i}, \tag{1}$$

  where $F_{\text{h}i}$ is the force applied by Peer i to its haptic interface and $F_{\text{C}i}$ is the contact force between Peer i and its SVO copy;
- for Peer i's copy of the SVO:

$$m_{\text{O}i}\ddot{x}_{\text{O}i} + b_{\text{O}i}\dot{x}_{\text{O}i} = F_{\text{C}i} - F_{\text{T}i}, \tag{2}$$

  where $F_{\text{T}i}$ is the coordination force applied by the AP coordination controller to Peer i's copy of the SVO.

In Equations (1) and (2), the contact and coordination forces are computed using:

$$F_{\text{C}i} = K_{\text{C}i}(x_{\text{HD}i} - x_{\text{O}i}) + B_{\text{C}i}(\dot{x}_{\text{HD}i} - \dot{x}_{\text{O}i}) \tag{3}$$

and:

$$F_{\text{T}i} = K_{\text{T}}(x_{\text{O}i} - x_{\text{O}i_d}) + B_{\text{T}}(\dot{x}_{\text{O}i} - \dot{x}_{\text{O}i_d}). \tag{4}$$

In Equation (4), $x_{\text{O}i_d}$ and $\dot{x}_{\text{O}i_d}$ are the desired position and velocity of the SVO copy of Peer i and, according to the AP strategy, are computed through averaging the motion commands coming from the other two peers:

$$x_{\text{O}i_d} = \frac{\sum_{j=1,j\neq i}^{3} x_{\text{O}j_n}}{2} \quad \text{and} \quad \dot{x}_{\text{O}i_d} = \frac{\sum_{j=1,j\neq i}^{3} \dot{x}_{\text{O}j_n}}{2} \tag{5}$$

Lastly, the motion commands are the delayed position and velocity of the sending peer:

$$x_{\text{O}i_n}(t) = x_{\text{O}i}(t - T_d) \quad \text{and} \quad \dot{x}_{\text{O}i_n}(t) = \dot{x}_{\text{O}i}(t - T_d), \tag{6}$$

and T_d is the network delay. In this paper, the network delay is assumed constant and equal in all communication channels, and an integer multiple of the network packet update interval T_n.

III. Stability Analysis

Because of network bandwidth limitations, the packet update rate is generally lower than the typical 1 KHz haptic rate required for realistic force feedback, especially in rigid virtual environments. Therefore, three-users networked haptic cooperation is a system with two sampling intervals: the (fast) sampling interval of the local force feedback loops at the peer users, T_c, typically equal to 0.001 s; and the (slow) sampling interval of the network updates, T_n, typically equal to 0.008 s [9]. Its stability can be investigated using eigenvalue analysis of its multirate state transition matrix.

 ISBN: 978-1-61208-117-5

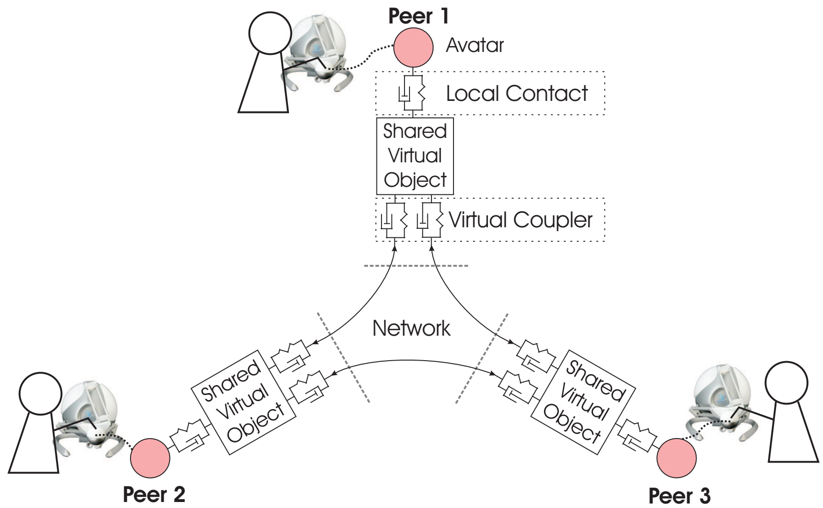


Figure 1. Traditional virtual coupling coordination of distributed haptic cooperation among three users.

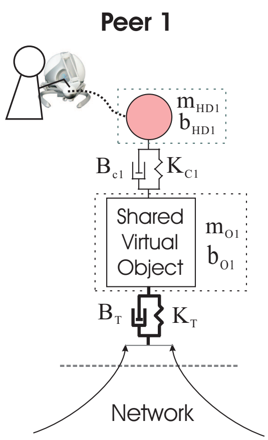


Figure 2. AP coordination of haptic cooperation among three users as applied at Peer 1.

In turn, this matrix can be computed through deriving the state space model of the multirate haptic cooperation system using the lifting approach in [2] and [9]. The derivations are briefly presented in the following sections. They assume a communication delay in each direction equal to one network update interval, i.e., $T_d = T_n = 0.008$ s.

A. Open-loop Continuous-time State-space Representation

The continuous-time state-space representation of the open-loop three-users networked haptic cooperation system is obtained from the dynamics of the users, of the haptic interfaces and of the SVO copies by grouping the system inputs and outputs into fast and slow sub-vectors, hereafter denoted with c and n indices respectively. Specifically, the system inputs comprise the contact forces, updated at the fast haptic rate (Equation (3)), and the SVO coordination forces, including both fast and slow updated components (Equation (4)):

$$\mathbf{u}^T = (\mathbf{u}_c^T \quad \mathbf{u}_n^T)^T \tag{7}$$

where:

$$\mathbf{u}_c^T = (F_{\mathrm{C1}} \quad F_{\mathrm{T1}_c} \quad F_{\mathrm{C2}} \quad F_{\mathrm{T2}_c} \quad F_{\mathrm{C3}} \quad F_{\mathrm{T3}_c})^T \tag{8}$$

$$\mathbf{u}_n^T = (F_{\mathrm{T1}_n} \quad F_{\mathrm{T2}_n} \quad F_{\mathrm{T3}_n})^T \tag{9}$$

$$F_{\mathrm{T}i_c} = K_{\mathrm{T}} x_{\mathrm{O}i} + K_{\mathrm{T}} x_{\mathrm{O}i} \tag{10}$$

and:

$$F_{\mathrm{T}i_n} = -K_{\mathrm{T}} x_{\mathrm{O}i_d} - K_{\mathrm{T}} x_{\mathrm{O}i_d}. \tag{11}$$

The state vector comprises the states of all haptic interfaces and SVO copies:

$$\mathbf{x}^T = (\mathbf{x}_{\text{peer1}} \quad \mathbf{x}_{\text{peer2}} \quad \mathbf{x}_{\text{peer3}})^T \tag{12}$$

where:

$$\mathbf{x}_{\text{peeri}}^T = (x_{\mathrm{HD}i} \quad \dot{x}_{\mathrm{HD}i} \quad x_{\mathrm{O}i} \quad \dot{x}_{\mathrm{O}i})^T; \quad i = 1,2,3. \tag{13}$$

The output vector is:

$$\mathbf{y}^T = (\mathbf{y}_{\mathrm{c}}{}^T \quad \mathbf{y}_{\mathrm{n}}{}^T)^T, \tag{14}$$

where:

$$\mathbf{y}_{\mathrm{c}}{}^T = \mathbf{x}^T, \tag{15}$$

$$\mathbf{y}_n^T = (\mathbf{y}_{\text{peer1}_n} \quad \mathbf{y}_{\text{peer2}_n} \quad \mathbf{y}_{\text{peer3}_n})^T \tag{16}$$

and

$$\mathbf{y}_{\text{peeri}_n}^T = (x_{\mathrm{O}i_d} \quad \dot{x}_{\mathrm{O}i_d})^T; \quad i = 1,2,3. \tag{17}$$

Hence, the continuous-time state-space model of open-loop three-user networked haptic cooperation with AP coordination is:

$$\begin{aligned} \dot{\mathbf{x}}_{12\times1} &= \mathbf{A}_{12\times12}\mathbf{x}_{12\times1} + \mathbf{B}_{12\times9}\mathbf{u}_{9\times1} \\ \mathbf{y}_{18\times1} &= \mathbf{C}_{18\times12}\mathbf{x}_{12\times1} \end{aligned} \tag{18}$$

 ISBN: 978-1-61208-117-5

B. Discrete-Time State-Space Representation

Following the approach [2] and assuming that the network sampling interval is an integer multiple of the sampling interval of the peers' force control loops and that the force feedback and network update sampling times are synchronized, the discrete-time state-space representation of the open-loop system can be written in the form:

$$\begin{aligned} \mathbf{x}_{\mathrm{D}_{96\times1}}[k+1] &= \mathbf{A}_{\mathrm{D}_{96\times96}}\ \mathbf{x}_{\mathrm{D}_{96\times1}}[k] + \mathbf{B}_{\mathrm{D}_{96\times51}}\ \mathbf{u}_{\mathrm{D}_{51\times1}}[k] \\ \mathbf{y}_{\mathrm{D}_{102\times1}}[k] &= \hat{\mathbf{C}}_{\mathrm{D}_{102\times96}}\ \mathbf{x}_{\mathrm{D}_{96\times1}}[k] + \hat{\mathbf{D}}_{\mathrm{D}_{102\times51}}\ \mathbf{u}_{\mathrm{D}_{51\times1}}[k] \end{aligned} \quad (19)$$

where k is the k-th network update interval and more details about the derivations of the system matrices A_{D}, B_{D}, $\hat{C}_{\mathrm{D}}$ and $\hat{D}_{\mathrm{D}}$ can be found in [9]. Furthermore, by augmenting the state vector with the delayed inputs [9], computational and communication delays are incorporated into the discrete-time open-loop model in Equation (19).

C. Stability Analysis

For three-users networked haptic cooperation, the feedback matrix F_{D} comprises the contact and SVO coordination forces and is computed using the approach [2]. Thereafter, the stability of the multirate closed-loop system can be derived through eigenvalue analysis of the closed-loop state transition matrix A_{D}^{cl}, calculated via:

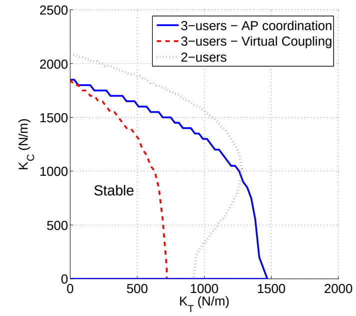
$$A_{\mathrm{D}}^{cl} = A_{\mathrm{D}_{aug}} + B_{\mathrm{D}_{aug}}\ F_{\mathrm{D}}\ (I - D_{\mathrm{D}_{aug}} F_{\mathrm{D}})^{-1} C_{\mathrm{D}_{aug}} \quad (20)$$

where $A_{\mathrm{D}_{aug}}$, $B_{\mathrm{D}_{aug}}$, $C_{\mathrm{D}_{aug}}$ and $D_{\mathrm{D}_{aug}}$ are the state transition matrices obtained after augmentation with computational and communication delays. Specifically, the three-users networked haptic cooperation system is stable if and only if all eigenvalues of A_{D}^{cl} are inside the unit circle:

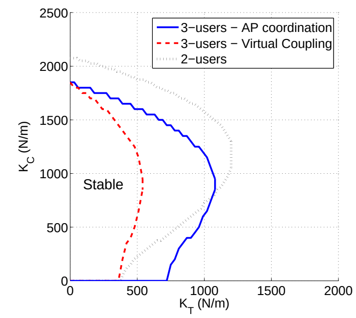
$$\left\| eig(A_{\mathrm{D}}^{cl}) \right\| < 1 \quad (21)$$

The stability regions for cooperation with AP and with virtual coupling coordination are presented: (i) in Fig. 3 for the case of Proportional coordination, when damping is injected as SVO damping ($B_{\mathrm{T}} = B_{\mathrm{C}i} = 0$ Ns/m); and (ii) in Fig. 4 for the case of Proportional-Derivative coordination, when damping is injected as coordination and local contact damping ($B_{\mathrm{T}} = B_{\mathrm{C}i} = 2$ Ns/m). The numerical computations are carried out considering that the haptic devices has mass $m_{\mathrm{HD}_i} = 0.1$ kg and physical damping $b_{\mathrm{HD}_i} = 5.0$ Ns/m and for a SVO with mass $m_{\mathrm{O}} = 0.6$ kg and damping $b_{\mathrm{O}} = 5.0$ Ns/m. The sampling interval of the force feedback loops at all peers is $T_{\mathrm{c}} = 0.001$ s and the network update interval is $T_{\mathrm{n}} = 0.008$ s. The computational delay is equal to one sampling interval of the force feedback loop, i.e., $T_{VE} = T_c = 0.001$ s. The communication delay is equal to one network update interval, i.e., $T_d = T_n = 0.008$ s.

Fig. 3 illustrates that the Proportional AP coordination maintains the cooperation stable for K_T gain twice as large as the K_T gain of Proportional virtual coupling coordination. The increase in the K_T gain afforded by the AP strategy

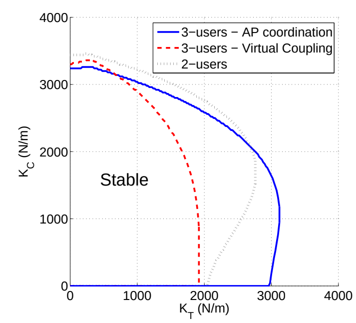


(a) Computational delay.

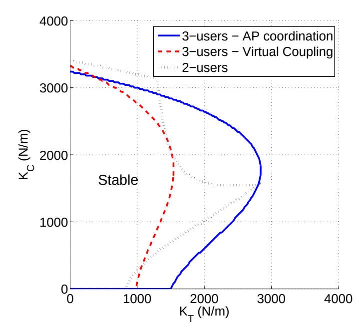


(b) Computational and communication delay.

Figure 3. Stability region for three-users haptic cooperation with Proportional SVO coordination.



(a) Computational delay.



(b) Computational and communication delay.

Figure 4. Stability region for three-users haptic cooperation with Proportional-Derivative SVO coordination.

ISBN: 978-1-61208-117-5

is smaller compared to the virtual coupling strategy when Proportional-Derivative coordination is deployed. This result confirms the negative impact of delayed damping on stability. Both figures also attest to the negative impact on system stability of the communication (network) delay.

IV. EXPERIMENTS

This section contrasts the AP coordination strategy to traditional virtual coupling coordination through controlled three-users cooperative manipulations. The controlled experimental cooperations are performed on a platform that enables point interaction between the networked users and the shared virtual environment. The platform comprises three FALCON NOVINT haptic devices, each connected to a respective computer and providing 3DOF force feedback to users. All computers run Windows XP on an Intel Core 2 Duo CPU at 2.67 Ghz with 2 GB RAM. The computers are in the same laboratory and communicate over the network via the UDP protocol. Given the proximity of the three computers, the network delay is negligible and a Wide Area Emulator (WANem) running on a separate computer implements the network environment. The shared virtual environment is generated as a C++ console application. It comprises a shared virtual cube in a rigid enclosure that constrains the cube to move horizontally along a single direction. To ensure the "same" users during successive experiments, the user-applied forces are replaced by forces commanded to the actuators through software. Since the haptic devices are impedance-type interfaces, the forces commanded through software eliminate the adaptive damping associated with the user manipulations of the haptic devices and represent a worst-case scenario for stability [11].

In the experiments, Peer 1 is located on the right side of the cube and pushes the cube to the left with constant force $F_{\text{h1}} = 4$ N. Peer 2 and Peer 3 are on the left side of the cube and push the cube to the right with constant and equal forces $F_{\text{h2}} = F_{\text{h3}} = 2$ N. The experiment starts with the users not in contact with the virtual cube. Fig. 5 depicts the snapshot of initial experimental conditions displayed at Peer 1. The virtual cube has mass $m_{\text{O}} = 0.3$ kg and hence, the local cube copies at the three peers have mass at each peer $m_{\text{O}i} = 0.1$ kg. The sampling interval of the force control loop is $T_{\text{c}} = 0.001$ s, and the network update interval is $T_{\text{n}} = 0.008$ s. The network delay is considered negligible. The stiffness and damping of the local contacts are $K_{\text{C}} = 3900$ N/m and $B_{\text{C}} = 0.05$ Ns/m, respectively. The SVO coordination gains are $K_{\text{T}} = 2000$ N/m and $B_{\text{T}} = 1.0$ Ns/m.

Experimental three-users networked haptic cooperation results are shown in Fig. 6a for AP coordination of the SVO, and in Fig. 6b for traditional virtual coupling coordination. These results illustrate that the cooperation is stable for AP coordination, and unstable for virtual coupling coordination. Thus, they validate the ability of the AP coordination strategy to maintain the networked cooperative manipulation stable for stiffer coordination gains.

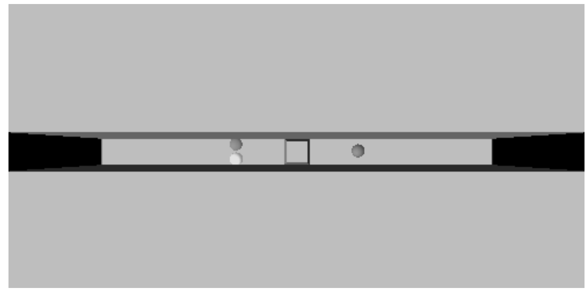
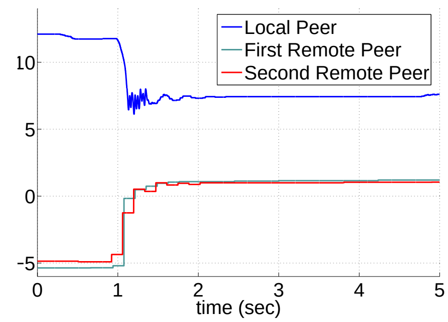
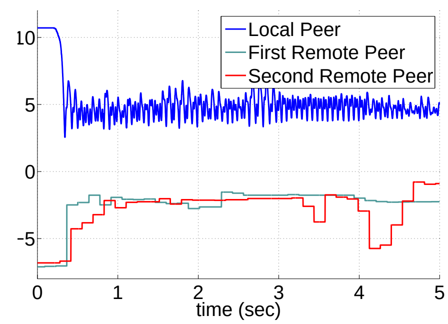


Figure 5. Snapshot of initial experimental conditions displayed to Peer 1.



(a) AP coordination.



(b) Virtual coupling coordination.

Figure 6. Controlled experimental three-users cooperation.

V. CONCLUSION

This paper has investigated the stability of the average position (AP) coordination strategy for three-users networked haptic cooperation. The AP strategy maintains the SVO coordination gain constant regardless of the number of cooperating participants. For three-users networked haptic cooperation, the paper has validated through analysis and experiments that AP coordination maintains the interaction stable for larger coordination gains. The stability analysis has been based on the eigenvalues of the state transition matrix of the three-users multirate haptic feedback system. The experiments have illustrated a one DOF cooperative manipulation of a virtual cube. Regardless of the one DOF analysis and experiments, the AP coordination strategy is readily applicable to rigid body cooperative haptic manipulations.

Upcoming work investigates the transparency of AP coordination and techniques to improve it in the presence of significant network delay, as well as its stability for cooperation across networks with variable delay and packet loss.

ACKNOWLEDGMENT

This work has been supported through an NSERC Discovery Grant.

REFERENCES

[1] R. J. Adams and B. Hannaford. Stable haptic interaction with virtual environments. *IEEE T Robot Autom*, 5(3):465–474, 1999.

[2] M. Araki and K. Yamamoto. Multivariable multirate sampled-data systems: state-space description, transfer characteristics, and nyquist criterion. *IEEE T Autom Contr*, 31(2):145–154, 1986.

[3] J.N. Bailenson, N. Yee, S. Brave, D. Merget, and D. Koslow. Virtual interpersonal touch: expressing and recognizing emotions through haptic devices. *Hum.-Comput. Interact.*, 22(3):325–353, 2007.

[4] G. Bianchini, M. Orlandesi, and D. Prattichizzo. An LMI framework for analysis and design of multi-dimensional haptic systems. In *IEEE Conf Decision Contr*, pages 4564–4569, Cancun, MEXICO, 2008.

[5] G. Bianchini, M. Orlandesi, and D. Prattichizzo. Stability analysis and design of multi-dimensional haptic systems. In *IEEE Symp Haptic Interf Virt Envir Teleop Syst*, pages 177–184, Reno, NV, 2008.

[6] M. Capps, P. PMcDowell, and M. Zyda. A Future for Entertainment-Defense Research Collaboration. *IEEE Computer Graphics Applic*, 21(1):37–43, 2001.

[7] J. E. Colgate and G. G. Schenkel. Passivity of a class of sampled-data systems: application to haptic interfaces. *J Robotic Syst*, 14(1):37–47, 1997.

[8] J.E. Colgate and M.C. Stanley. Issues in the haptic display of tool use. In *IEEE/RSJ Int Conf Intell Robot Syst*, pages 140–145, Pittsburgh, PA, 1995.

[9] M. Fotoohi, S. Sirouspour, and D. Capson. Stability and performance analysis of centralized and distributed multi-rate control architectures for multi-user haptic interaction. *Int J Robot Res*, 26(9):977–994, 2007.

[10] M. Glencross, C. Jay, J. Feasel, L. Kohli, M. Whitton, and R. Hubbold. Effective cooperative haptic interaction over the Internet. In *IEEE Virt Real Conf*, pages 115–122, Charlotte, NC, 2007.

[11] B. Hannaford and R. Anderson. Experimental and simulation studies of hard contact force reflecting teleoperation. In *IEEE Int Conf Robot Autom*, pages 24–29, Scottsdale, AZ, 1988.

[12] Y.B Kim, S.H. Han, S.J. Kim, E.J. Kim, and C.G. Song. Multi-player virtual ping-pong game. In *Int Conf Artif Real Telexistence*, pages 269–273, 2007.

[13] D.J. Lee and K. Huang. On passive non-iterative variable-step numerical integration of mechanical systems for haptic rendering. In *ASME Dyn Syst Contr Div Conf*, pages 1147–1154, 2008.

[14] D.J. Lee and K. Huang. Passive-set-position-modulation framework for interactive robotic systems. *IEEE T Robot*, 26(2):354–369, 2010.

[15] D.J. Lee and K. Huang. Peer-to-peer control architecture for multiuser haptic collaboration over undirected delayed packet-switching network. In *IEEE Int Conf Robot Autom*, pages 1333 –1338, Anchorage, AK, 2010.

[16] S. Lee and JW. Kim. Transparency Analysis and Delay Compensation Scheme for Haptic-based Networked Virtual Environments. *Comput Commun*, 32(5):992–999, 2009.

[17] Z. Li and D. Constantinescu. Networked Haptic Cooperation among Multiple Users via Virtual Object Coordinatino to Averaged Position of Peer Copies. In *IEEE Int Workshop Haptic Audio Vis Env Games*, pages 170–175, Lecco, Italy, 2009.

[18] P. Mitra and G. Niemeyer. Haptic Simulation of Manipulator Collisions Using Dynamic Proxies. *Presence: Teleop Virt Envir*, 16(4):367–384, 2007.

[19] G. Sankaranarayanan and B. Hannaford. Experimental comparison of internet haptic collaboration with time-delay compensation techniques. In *IEEE Int Conf Robot Autom*, volume 1, pages 206–211, Pasadena, CA, 2008.

[20] G. Sankaranarayanan and B. Hannaford. Experimental Internet Haptic Collaboration using Virtual Coupling Scheme. In *IEEE Symp Haptic Interf Virt Envir Teleop Syst*, pages 259–266, Waltham, MA, 2008.

[21] G. Sankaranarayanan and B. Hannaford. Experimental Internet Haptic Collaboration using Virtual Coupling Scheme. In *IEEE Symp Haptic Interf Virt Envir Teleop Syst*, pages 259–266, Waltham, MA, 2008.

[22] H. Sugarman, E. Dayan, A. Weisel-Eichler, and J. Tiran. The Jerusalem Telerehabilitation System, a New, Low-Cost, Haptic Rehabilitation Approach. *CyberPsychology Behav*, 9(2):178–182, 2006.

Design of a Wearable Direct-driven Optimized Hand Exoskeleton Device

Jamshed Iqbal
Advanced Robotics Department,
Italian Institute of Technology (IIT), Genova, Italy
University of Genova, Italy
jamshed.iqbal@iit.it

Nikos G. Tsagarakis, Darwin G. Caldwell
Advanced Robotics Department,
Italian Institute of Technology, (IIT), Genova, Italy
nikos.tsagarakis@iit.it, darwin.caldwell@iit.it

***Abstract*—This work introduces a hand exoskeleton that allows full range of motion and can exert bi-directional forces on the finger phalanges. The link lengths and structure of the proposed mechanism have been emerged as a result of kinematics-based optimization criteria. It is an under-actuated mechanism allowing 4 DOF/finger with one active degree. The selection of the actuator has been based on human hand capabilities to accomplish common daily life activities. An initial un-actuated finger prototype has been developed to analyze the mechanism functionality and to confirm the optimization results. Results have demonstrated that the device covers the complete operational range of motion of a human hand.**

Keywords-Hand exoskeleton System, Wearable robotics, Haptic device, Optimized robotic mechanism.

I. INTRODUCTION

Thanks to revolutionary trends in multidisciplinary areas of mechatronics and computing, realistic exoskeleton devices do exist. They are devices aiming to transmit kinaesthetic feedback at the level of the finger in order to emulate the constrains imposed by grasping of virtually or remotely manipulated objects. Exoskeleton based systems combine machine power and human intelligence to enhance human operator's capabilities as well as machine intelligence. Therefore, such systems have greatly improved to acquire the performance level which has not been possible otherwise. They are always designed as an external mechanical link structure and find potential applications in various areas including haptics, Virtual Reality (VR) and rehabilitation.

As the field of haptics evolved, the application of force feedback techniques in VR has become more demanding. Haptic interfaces have proven to be extremely useful during the interaction of users within virtual environment. Exoskeleton robotic devices make use of the haptic sense to enhance the presence impression by reproducing the contact forces on the wearer hand. Thus the use of such system improves the interaction with the virtual environment and increases the dexterity of the operator with the virtual objects. Such interfaces can be classified into grounded and portable devices. Grounded devices can only simulate fixed objects because they limit the range of the operators' freedom of motion. For pseudo-natural interactions, portable haptic devices are used because they provide more flexibility in terms of operator's freedom. However they can simulate fixed forces only when employed with a grounded device.

Considering the domain of rehabilitation, theraphy procedures are usually required to regain the normal hand strength and capabilities. In past, such procedures were executed manually by physiotherapists. Occasionally simple passive assistive devices have been employed to aid in rehabilitation. However technological revolution has evolved robotic hand exoskeletons that may be used in rehabilitation to improve the medical outcome.

This paper is structured as follows: Section II explains the related work while Section III highlights brief design details including goals and device mechanism. Section IV deals with the design optimization of the proposed device while Section V introduces design requirements, Section VI presents the details of preliminary and final prototype, and finally, Section VII comments on the conclusion.

II. RELATED WORK

The development of a hand exoskeleton device is a very challenging endeavor, which has been targeted by many research professionals. A multi-phalanx hand exoskeleton consisting of four fingers that was able to exert forces on each phalanx of each finger was developed at PERCRO lab [1]. Few years later, researchers at Keio University realized a three fingers non-isomorphic device actuated by passive clutches [2]. Springer and Nicola at the University of Wisconsin have presented a 1-finger prototype utilizing a planar four-bar linkage. They analyzed the haptic effect perceived by the user [3]. A 2-finger hand exoskeleton intended for VR grasping simulation, having 3 Degree Of Freedom(DOF)/finger and 4 for the thumb has been developed by Stergiopoulos [4]. Lelieveld *et al.* proposed a 4 DOF portable wearable haptic interface with active and passive multi-point feedback for the index finger in master-slave configuration [5]. Another hand exoskeleton developed at PERCRO intended for haptic interaction in virtual environment has 3 DOF/finger and can exert controlled forces on the finger tips [6]. A under-actuated 2-

 ISBN: 978-1-61208-117-5

finger hand exoskeleton has been conceived by researchers at IIT. It consists of an optimized Revolute-Revolute-Revolute (RRR) mechanism and can provide force levels (45N) beyond any existing system. The main optimization criteria are Global Isotropy Index (GII) and Perpendicular Impact Force (PIF) factors. The proposed system can be used for tele-operation, VR, Human-Robot-Interaction (HRI) [7].

The present Hand EXOskeleton SYStem (HEXOSYS) is an effort to combine good salient features present individually in the existing hand exoskeleton designs: Portability, Human hand compatibility, Optimization, Direct-driven, Back-drivability, Full range of motion, Light mass, Low complexity, Provision of bi-directional force and so on. At the time of writing this paper, there is not any existing hand exoskeleton system reported in the literature that encompasses all of these mentioned features in a single device. This fact essentially puts the novelty in the proposed system. It has been designed as a general purpose system. That is why instead of concentrating on the requirements of a particular application, the device specifications have been derived from a series of experiments with the human hand, thus keeping the system beneficial for wide range of applications. Potential applications of this system include tele/virtual presence, tele/virtual manipulation and rehabilitation. Trying to match the human hand force capabilities as well as natural workspace resulted in a human hand compatible device.

III. Design Goals & Device Mechanism

The human hand can be counted as one of the most difficult systems to emulate with a mechatronic device. This difficulty is mainly due to two reasons. First, is the unavailability of ample space for components placement and second is high number of DOF (4 or even more in case of thumb). These pose a great challenge in terms of design requirements. Some obvious desired characteristics of such a system are summarized below. Detailed desired requirements of such a system are reported in [7].

(i) Low mass/inertia.
(ii) Unconstrained range of motion.
(iii) Minimum complexity.
(iv) Comfort.

It is clear that the above requirements cannot be satisfied by solutions employing large number of actuation units trying to power most of the finger phalanges. Our approach to design the HEXOSYS is to use a less complicated finger exoskeleton mechanism. The conceptual mechanism proposed in this work is shown in Figure 1. The finger exoskeleton system is a two links planar under-actuated mechanism which is attached to the user finger at a single point. A single actuation unit residing at the proximal joint of the exoskeleton is used to power the device.

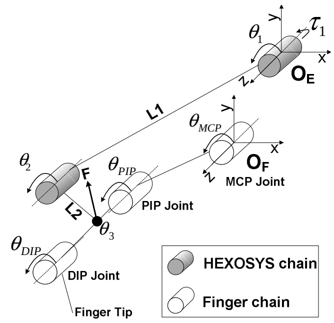


Fig. 1. Kinematic model of the proposed device

Table I lists HEXOSYS design goals and primary specifications while the DH parameters of the device are presented in Table II. Based on these, (1) states the overall transformation from end-effector to the base of exoskeleton.

TABLE I
HEXOSYS Specifications

MECHANICS	✓ Direct drive
	✓ Optimized link structure
	✓ Underactuated Revolute-Revolute (RR) mechanism
	✓ 4 DoF/finger (1 active)
	✓ Low complexity
ERGONOMICS	✓ Full range of motion
	✓ Light mass expected
	✓ Low volume
	✓ Variable hand sizes
	✓ Portable
	✓ Palm-free
	✓ Easy removal and donning
PERFORMANCE	✓ Bi-directional forces of upto 8N
	✓ Position and force feedback
	✓ Up to 5 fingers

TABLE II
DH Parameters of The Proposed Device

i	α_{i-1}	a_{i-1}	d_i	θ_i
1	0	0	0	θ_1
2	0	L_1	0	θ_2
3	0	L_2	0	θ_3

$$ {}^{0}_{3}T = \begin{bmatrix} C_{123} & -S_{123} & 0 & L_1C_1 + L_2C_{12} \\ S_{123} & C_{123} & 0 & L_1S_1 + L_2S_{12} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1) $$

where
$S_{123} = \sin(\theta_1 + \theta_2 + \theta_3)$
$C_{123} = cos(\theta_1 + \theta_2 + \theta_3)$

The mapping from velocities in joint space to Cartesian space (Jacobian matrix) is given by

$$ J = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 1 & 1 & 1 \end{bmatrix} \quad (2) $$

 ISBN: 978-1-61208-117-5

where

$a_{11} = -L_1S_1 - L_2S_{12}$
$a_{12} = -L_2S_{12}$
$a_{13} = 0$
$a_{21} = L_1C_1 + L_2C_{12}$
$a_{22} = L_2C_{12}$
$a_{23} = 0$

IV. Devise Optimization

The kinematic performance of the proposed device is essentially a strong function of proper choice of link lengths and the link shape. This motivated us to carry out an optimization procedure during design of the HEXOSYS to find out the optimum values of exoskeleton link lengths The primary optimization criteria is kinematics, finger-exoskeleton WorkSpaces (WS) matching and worst case collision avoidance. Since the HEXOSYS is attached to the finger at the mid of the middle phalange, this point is considered for matching WS.

To increase the collision-free reachable WS of HEXOSYS, the first link (L1) shown in Figure 1 has been sub-divided into three segments (L1-1, L1-2 and L1-3) as illustrated in Figure 2.

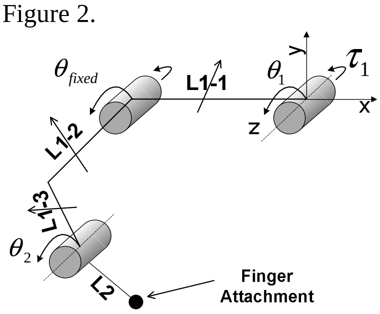


Fig. 2. First link structure of HEXOSYS ensuring collision-free workspace

These segment lengths are adjustable depending upon finger and hand size. Similarly the angle between first two sub-segments (θ_{fixed}) is also variable but is fixed for a certain finger and hand size. The objective of optimization algorithm is to find the optimum lengths of these three segments and angle (θ_{fixed}). The optimization algorithm takes the hand size and finger as inputs and gives these optimized values as an output.

HEXOSYS link lengths have been iterated through reasonable range. Each set of link lengths is then subjected to traverse through the complete finger WS for analysis. Using inverse kinematics, the set of link length is then analyzed to see how many points inside the finger WS, the exoskeleton can reach without collision. For collision detection, equidistant points (0.5cm apart) on the HEXOSYS links and the rectangular envelopes surrounding the finger centre of mass have been determined. An HEXOSYS link length set is considered as collision-free if all the points on the links reside outside the rectangular envelopes. Finally the collision-free WS is stored for comparison with the next iterated link lengths set. Figure 3 shows the overall optimization strategy of HEXOSYS.

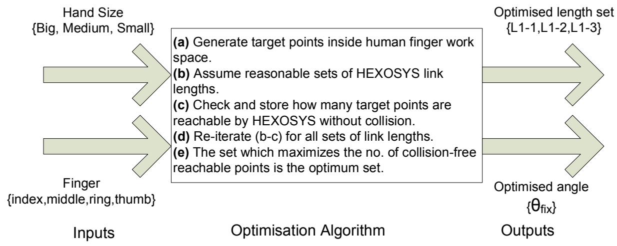


Fig. 3. Optimization strategy of HEXOSYS

For a medium hand size, in case of index finger, the optimized segments have been found to be L1-1=8cm, L1-2=2cm, L1-3=2cm and θ_{fixed}=55.4°. The corresponding WS is illustrated in Figure 4.

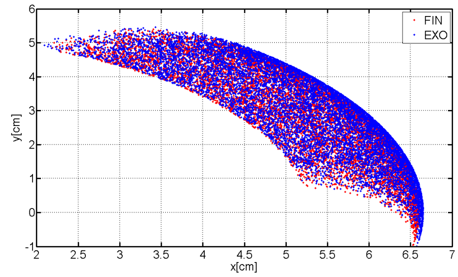


Fig. 4. Plot showing the workspaces of finger (red) and exoskeleton (blue) on optimized link lengths

V. Design Requirements

To collect necessary data for selecting the actuator, an analysis of the most common daily life activities of the hand has been carried. Included in this analysis are the experiments to measure average or maximum exerted force levels and the required range of motion. The data collected from these experiments can then be mapped to lower level actuator requirements. Commercially available data glove, a load cell and fingertip force sensors have been used to collect data from small, medium and big hand sizes in various activities. Detailed design requirements have been mentioned in [8]. Figures 5-6 show results of force profiles corresponding to interacting with a small and comparatively big object. The results of activities requiring average force levels demonstrate that we usually need 10-15N to accomplish many of our daily life activities. Another experiment intended to measure the maximum force levels exerted by human finger revealed that the maximum levels can go up to 45N.

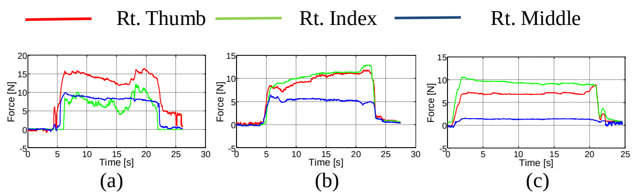
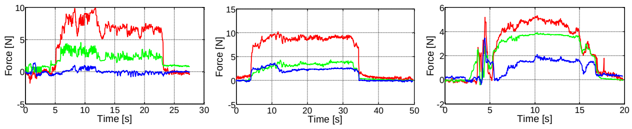


Fig. 5.Force profiles of taking a big object (cup) in case of (a) Small (b) Medium (c) Big hand sizes

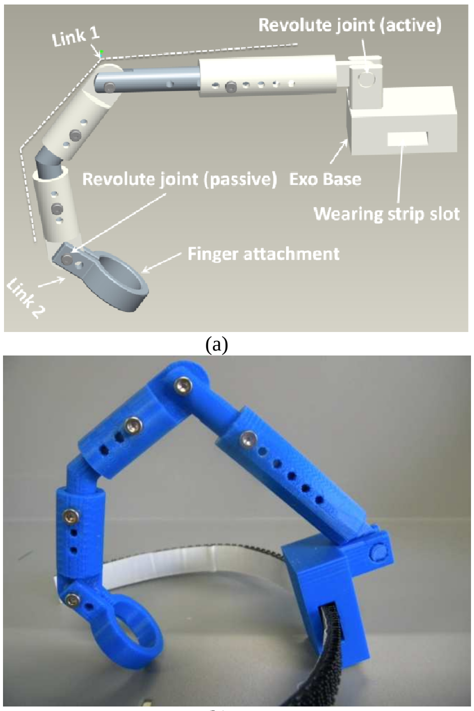


(a) (b) (c)

Fig. 6. Force profiles of interacting a small object (writing with pencil) in case of (a) Small (b) Medium (c) Big hand sizes

VI. DEVICE PROTOTYPE

Initially, an un-actuated finger prototype made up of ABS-plastic has been fabricated using a high-tech in-house 3D printer. The purpose was to verify the optimization results and analyze the behavior of the selected mechanism (RR). Both revolute joints can be moved passively without imposing any unrealistic constraints. The prototype can be easily fastened and removed from the hand using a single velcro clip. The CAD design and snapshot of this initial prototype has been illustrated in Figure 7.



(a)

(b)

Fig. 7. The un-actuated rapid finger prototype (a) CAD design (b) Prototype

With the initial prototype on the hand, the collision-free reachable workspace has been observed by moving the finger from complete flexion to complete extension. Figure 8(a-d) gives illustration of one complete cycle. The mechanism together with the optimized link lengths and shape provide full range of motion without any constraints as evident in Figure 8. This confirms the optimization results presented in Section III.

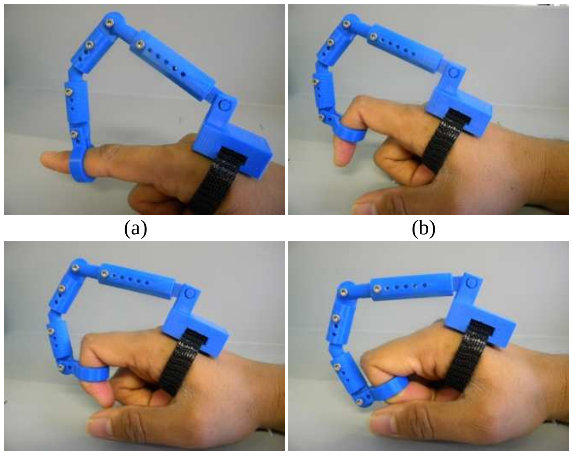

(a) (b) (c) (d)

Fig. 8. Complete flexion-extension cycle depicting the full collision-free workspace coverage of an index finger

The final prototype consists mainly of an actuator (Portescap 16G88-220P) with its accessories (R16 Gear and MR2 encoder) and a pair of bevel gears (1:1). The use of bevel gears, by changing the orientation of motor axis permits the extension for neighboring hand fingers. The prototype is planned to provide position as well as force feedback. Figure 9 shows CAD view of the final prototype. The prototype provides flexion, extension as well as passive abduction.

VII. CONCLUSION

We have proposed design of a portable, direct-driven and optimized hand exoskeleton system that has the capability to provide the force levels necessary to accomplish common daily life activities. The system design has been emerged as a consequence of optimization studies to ensure that the device has complete range of motion as that of human hand. Moreover, the data collected from the series of experiments on human hands paved the way to choose actuators. Initially an un-actuated rapid prototype has been fabricated to confirm the optimization results. The final prototype captured using CAD tools is being sent for fabrication and is expected to be light weight and less volumetric.

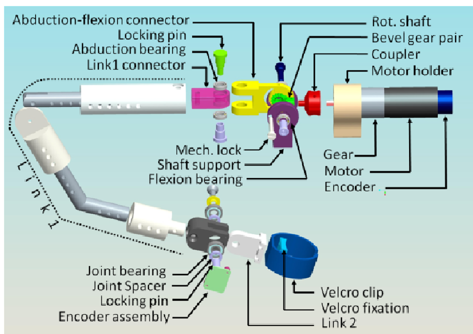


Fig. 9. CAD Design of final prototype (Exploded view)

 ISBN: 978-1-61208-117-5

REFERENCES

[1] Bergamasco M., "Design of hand force feedback systems for glovelike advanced interfaces", IEEE International Workshop on Robot and Human Communication, pp. 286–293, 1992.

[2] Koyama T., Yamano I., Takemura K., and Maeno T., "Multi-fingered exoskeleton haptic device using passive force feedback for dexterous teleoperation", IEEE/RSJ International Conference on Intelligent Robots and System (IROS) , vol. 3, pp. 2905–2910, 2002.

[3] Springer S. and Ferrier N. J., “Design and control of a force-reflecting haptic interface for teleoperational grasping”, Journal of Mechanical Design, vol. 124, pp. 277–283, 2002.

[4] Stergiopoulos P., Fuchs P., and Laurgeau C., “Design of a 2-finger hand exoskeleton for VR grasping simulation,” in Eurohaptics, Ireland, pp. 80–93, 2003.

[5] Lelieveld M.J., Maeno T., and Tomiyama T., “Design and development of two concepts for a 4 DOF portable haptic interface with active and passive multi-point force feedback for the index finger”, ASME International Design Engineering Technical Conference & Computers and Information in Engineering Conference, 2006.

[6] Fontana M., Dettori A., Salsedo F., and Bergamasco M., “Mechanical design of a novel hand exoskeleton for accurate force displaying”, IEEE International Conference on Robotics and Automation (ICRA), pp. 1704–1709, 2009.

[7] Iqbal J., Tsagarakis N.G., Fiorilla A.E., and Caldwell D.G., “Design requirements of a hand exoskeleton robotic device”, 14th IASTED International Conference on Robotics and Applications (RA), Massachusetts US, pp. 44-51, 2009.

[8] Iqbal J., Tsagarakis N.G., Fiorilla A.E., and Caldwell D.G., “A portable rehabilitation device for the hand”, 32nd annual IEEE international conference of Engineering in Medicine and Biology Society (EMBS), Argentina, pp. 3694-3697, 2010.

Road-based Adaptation of In-Car-Infotainment Systems

Sandro Rodriguez Garzon
Daimler Center for Automotive IT Innovations
Berlin, Germany
sandro.rodriguez.garzon@dcaiti.com

Kristof Schütt
Daimler Center for Automotive IT Innovations
Berlin, Germany
kschuett@cs.tu-berlin.de

***Abstract*—This paper introduces a prototype of a highly adaptive in-car-infotainment system. The prototype processes historical user interactions in order to discover regular sequences of interaction events within similar environments. The discovered environments in form of road segments and the activities will be used to adapt the human-machine interface (HMI) in case the car approaches an environment that is likely to contain a known activity. Temporal event patterns as well as grouping criteria need to be prespecified to control the detection of characteristical activities and to define the way two road segments are declared to be similar. Furthermore, a brief technical introduction of the automotive specific process of location information preprocessing is given, which is used by the prototype to interpret its environment or to group road segments. The use cases of a button that changes its size depending on the probability of being pressed at a certain road as well as a HMI that automatically switches to a certain radio station based on historical user initiated radio station changes at similar road segments will be discussed in detail.**

***Keywords*-context awareness, personalization, adaptation, intelligent user interface**

I. Introduction

Over the last years more and more mobile devices such as mobile phones or in-car-infotainment systems became location-aware. In particular, the high distribution of smart phones with GPS sensor spurred the use of location-based services. Several applications like searching for point-of-interests within a certain range or notifications about friends in close-by environments became popular and underlined the user demand for applications that provide location-based content.

So far, the dynamic content is generated in case the user approaches a certain environment and either prompts the system to show the content or gets notified if some predefined objects occur in the current environment. In both cases the user needs to manually define the kind of information that might be helpful in the current situation. In the former variant the user has to select a topic while in the latter variant the user has to prespecify a notion of interestingness for certain information. Both approaches use static user interests and location information for processing. But how to deal with users that have different interests depending on the location?

In order to present content based on interests that are influenced by the location the user would need to provide all individual associations between interests and locations beforehand. The expense for the user remains arguable as long as all interests are associable with user defined abstract locations like home or office. But it is getting difficult and impractical if the interests are connected to concrete locations characterized by latitude and longitude values. An interim solution would be to discover abstract locations of each user by investigating his moving behavior as showed in [1]. The abstract locations might be proposed to the user to facilitate the input of the mentioned interest and location relations.

A more pleasant approach is to discover abstract locations and the corresponding interests based on historical data of service use without the need to prompt the user to enter the interest-location relations. Such an approach would enhance several applications like restaurant search or GUI extensions like the generation of favorite button lists by considering the users interests at a certain location. An application for the restaurant search as a smart content filtering service can be realized by incorporating additional location parameter to the recommender system [2]. Complex personalization use cases like individualized favorite buttons that execute certain sequences of preferred user interactions to reduce click costs need a more sophisticated notion of interest.

This paper introduces an in-car-infotainment system prototype capable of personalizing itself based on regular user tasks at certain locations. The introduction is followed by a short survey of papers related to the fields of context-aware computing and location-based personalization. In the Section "Examples" two personalization examples are introduced that are used throughout this paper. A detailed description of all high-level components and their interactions with each other is given in Section "Architecture". The Section "Situation Recognizer" contains a step-by-step explanation of the event sequence detection and grouping process. A rather technical description of location preprocessing is given in Section "Map Matching". This paper ends with a summarizing conclusion and a discussion of refinements and future extensions.

II. Related Work

Over the last years a lot of research was done concerning the detection of regular tasks and its visualization. In [3]

 ISBN: 978-1-61208-117-5

Shen uses graph mining algorithms to extract regular work flows of desktop use as frequent subgraphs of a graph consisting of resources and the information flows in between. In [4] Brdiczka follows a different approach by considering the time in between user interactions in addition to the type of action as described by Shen. Neither of them deals with an observation of the environment at the moment the task was executed. An interesting approach of adapting a system based on its use at certain locations is described in [5]. Coutand uses case-based reasoning to detect relations between system utilization and location to decide about future mobile phone properties in certain situations. In this case there is no need to extract complex sequences of interactions to build sophisticated task representations because a mobile phone property modification is solely related to the task of approaching a location. Thus, a task is interpretable as the action of entering a certain context.

Several context-aware frameworks [6], [7] were introduced in order to extract abstract represenions of the context based on the aggregation of raw sensor data. Especially, the discovery of significant locations out of multiple sequences of GPS data to predict future movements of users or to enhance mobile phones became a popular field of research [8], [1], [9]. A promising approach of Liao in [10] deals with the task of mainly relating significant locations to regular user interactions. Liao discovers significant locations and infers corresponding activities using Relational Markov Networks. The process of detecting abstract tasks like dining is supported by location dependent information like a close-by restaurant obtained from geographical databases. In contrast, our approach does not need to infer high-level abstractions of tasks because personalization relevant activities are extracted by means of prespecified temporal event patterns. Inspired by [11], we are solely interested in actions originating from the use of an in-car-infotainment system. In [11], Rogers presents the application of user-centric personalizations within the automotive environment that is based on the analysis of the user actions "selecting preferred routes" and "spoken feedback". In case of our examples the user actions are "pressing a certain button" and "changing the radio station".

III. Examples

In the following sections, a simple and a rather complex personalization use case within the automotive environment will be discussed in detail. The simple use case deals with a button list of the HMI that changes its visualization based on its historical use. More precisely, the size of the buttons will depend on the probability of being pressed in a certain situation. Considering a situation to be defined as the road the car is currently driving on, the buttons should dynamically change their size according to their use at a certain road. This use case is declared to be simple because the actions that should be investigated namely the button click as well as the personalization triggering situation are single points in time.

```
<route>
  <waypoint>
    <position>
      <lat>48.422017</lat>
      <lon>9.536767</lon>
    </position>
    <time>1275296924299</time>
    <direction>189.8</direction>
    <height>808.7</height>
    <speed>15.3</speed>
  </waypoint>
  ...
</route>
```

Listing 1. Section of a context log

The more complex use case deals with a HMI that analyzes user initiated radio station changes in order to propose automatic radio stations changes depending on discovered regularities. A valid regularity would be discovered if the HMI user switches to a certain radio station every time he drives through a certain village. The proposal made by the HMI should first be visualized to the HMI user and later be executed if the user agreed to the automatic execution. Therefore, the personalization is executed indirectly in contrast to the simple use case where the resizing of a button is applied immediately. The complexity of the use case is caused by the action that consists of several temporal order events, which form a significant radio station change. In this case a radio station change is declared as significant if no other radio station change followed within 5 minutes.

IV. Architecture

The in-car-infotainment prototype consists of three main components: *Context Simulator*, *HMI* and *Situation Recognition*. Additionally, context log files based on XML are used in the Context Simulator to simulate environmental aspects like the position or the direction of the car. These log files are produced by recording location relevant information either at a real world in-car-infotainment system or by a mobile phone equipped with a GPS and compass sensor. Listing 1 shows a section of an example context log file.

A rapid prototyping approach is taken to implement an exemplary human machine interface based on Action Script. Within the client-server nomenclature the HMI is called the client and the Context Simulator can be seen as the server. Such an assignment became necessary because Action Script does not allow to implement listening sockets. During initialization the HMI tries to connect to the Context Simulator and waits for incoming context data. The HMI comprises many components like a navigation system and a radio in order to provide a wide range of possibilities for personalization use cases. Figure 1 shows a screenshot of

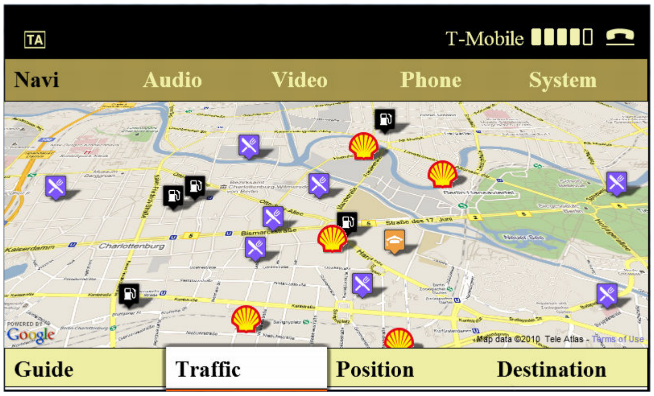


Figure 1. Screenshot of the HMI

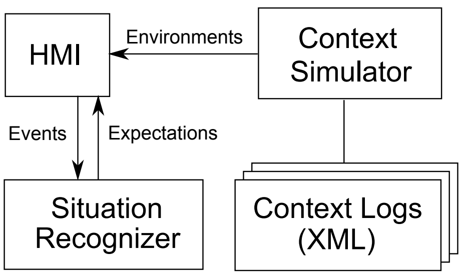


Figure 2. Overview architecture

the HMI in the navigation system modus.

In order to personalize the HMI based on its historical use we need to log all user interactions and the corresponding context. This is realized by sending all interaction data and indirect user interactions like location change to a separate component called Situation Recognizer. Since the Situation Recognizer does not differentiate between a direct or indirect interaction both events are summarized under the term *interaction event*. As stated above, the HMI acts as a client and connects to the Situation Recognizer at startup time. An overview about the whole architecture is given in Figure 2.

Each interaction event comprises of a header and a body section. The header section contains event attributes that are common to all events regardless of their type while the body section contains type dependent attributes. Listing 2 shows an example event that needs to be fired by the HMI to signalize a radio station change.

The `nr` attribute contains a successive number within a certain session while the `level` attribute denotes the abstraction layer of an event. So far, we identified 3 levels: *Raw events*, *logical events* and *instance events*. Raw events can represent a touch or a press of a button of the central command unit(CCU) while logical events can signalize a recognized gesture or a click (e.g. constructed of the raw events touch down and touch up). The raw events will be thrown globally at any time, whereas the instance events will only be thrown if a unique instance of an HMI element is affected. An example can be a click onto a certain button of the global application line(e.g. Navi). In this case the corresponding HMI element id is stored in the event attribute `sourceId`.

The body contains event attributes that belong to a certain type of instance event. Consider an interaction event that originates from a certain button of an application line signalizing a button click. In this case the button label might be included in the body to permit the Situation Recognizer to differentiate between clicks of different buttons that are part of the same button list.

The Situation Recognizer processes all interaction events and notifies the HMI about approaching a situation that will likely be the environment for a known interaction event sequence. In other words, the Situation Recognizer discovers regular interaction sequences and informs the HMI about future interaction event values and its environments. Therefore, the notification must contain all relevant situation information as well as future values of certain environmental aspects and its probabilities. These information will be used by the HMI to display a human-readable form of a discovered frequent interaction sequence as well as to execute personalizations. The interpretation of the notification message comprising use case dependent data is supported by a description of the message format specified beforehand.

Considering the complex example introduced in Section III the notification must contain the situation information comprising of the road name and the future radio station. Listing 3 shows the corresponding notification. The message will be interpreted by the HMI in the following way: The current road is called ”Main Street” and a user initiated switch to the radio station with name ”KissFM” is likely to appear now. The notion of the word ”likely” is discussed in Section ”Situation Recognizer”. The notification of the simple use case must contain the label of the button and the corresponding probability of execution.

V. Situation Recognizer

As stated above, the Situation Recognizer deals with the task of processing user interactions to discover frequent interaction sequences. These frequent user interaction sequences will be used to inform the HMI about the future steps of the user. To analyze sequences of events and its environments it is necessary to split the search process into 3 successive steps: *Action Discovery*, *Action Grouping* and *Situation Discovery*. Thereby, each step is supported by a use case description to keep control of the whole process by e.g., reducing the search space or by defining a fading function for an irrelevance factor for each user action. The use case descriptions must be defined by an expert together with the HMI developer.

```
<event nr="2156" level="I" type="StationChanged" sourceId="Radio" timestamp="1277112493971">
  <StationChanged>
    <frequency>http://www.kissfm.de</frequency>
    <stationName>KissFM</stationName>
  </StationChanged>
</event>
```

Listing 2. Example of an interaction event

```
<proposition uc="RadioChange">
  <action>
    <parameter>
      <stationFrequency>
        http://www.kissfm.de
      </stationFrequency>
      <stationName>
        KissFM
      </stationName>
    </paramenter>
    <environment>
        <street>Main Street</street>
    </environement>
  </action>
</proposition>
```

Listing 3. Example of a situation notification

A. Action Discovery

The first step namely Action Discovery uses a temporal event pattern of the use case description to extract a certain event sequence out of the stream of interaction events. For this purpose the complex event processing engine Esper [12] is selected, which analyzes event streams based on pattern descriptions written in Esper's event processing language(EPL). The resulting concrete sequence of events is called *action*.

In case of our complex use case, the temporal event pattern to extract significant radio station changes looks as follows:

$$StationChange \rightarrow Timer(5min, not(StationChange))$$

The temporal event pattern of the simple use case is omitted because it comprises only of the button click event. The resulting actions contain either a concrete radio station changed event or a concrete button click together with the corresponding context in form of a road segment id.

B. Action Grouping

The Action Grouping subprocess collects all actions and groups them by prespecified criteria. Each group should contain event sequences that occurred within similar environments. The notion of similarity is defined within the use case description. Therefore, each use case may contain different criteria to compare event sequences. A group is called significant if the amount of actions exceeds a preconfigured limit. Only significant groups are considered in the next step.

In the complex example use case, the actions should be grouped by similar radio stations and nearby road segments. Thus, each resulting group is characterizable by a single radio station and a set of road segment ids. The resulting group properties of the simple use case contain a single button label and a single road segment id because a button click is only related to a single road segment and not to a subgraph of the whole road network.

C. Situation Discovery

Finally, the Situation Recognizer uses the group properties to parametrize a prespecified temporal event pattern in order to extract similar situations. Thereby, the group properties are interpreted as a description of the environment the actions occurred in. If a similar situation is found the HMI will be informed about the group properties containing the environmental aspects as well as future values of use case relevant attributes.

The temporal event pattern that triggers a notification of the complex example use case may look like this:

$$StationChange\ and$$

$$StationChange.station\ != \ Group.station$$

$$\rightarrow$$

$$Context.location == Group.location$$

In words, if the last radio station change switched to a radio station that is different to a radio station property value of a discovered group and if the location of the current context is similar to the location of the same group then notify the HMI about the detected situation. The temporal event pattern of the simple use case contains only the location comparison. An overview about all steps discussed in this section is given in Figure 3.

So far, a situation notification is only triggered in case a significant group was found. This implies the instantiation and parametrization of a temporal event pattern for the detection of the corresponding situation in case a certain amount of similar actions occurred within the same environment. Considering the radio example such a notion of the statement ”likely to happen” could be adequate but insufficient for the dynamic buttons example. For the execution of resize commands it is necessary to know the probability of a button click within a certain environment cause the different buttons are influencing one another concerning their sizes. To calculate the probability of an occurrence it is necessary to relate the amount of occurrences to the amount of non-occurrences. But how does the prototype discover non-occurrences? It is impractical or almost impossible to detect and record all non-occurrences of actions that are unknown so far. Therefore, the prototype does only observe a non-occurrence of an action if the action happened at least once. Additionally, the detection of non-occurrences will only be constrained to an environment defined by the temporal event pattern of the corresponding situation. Thus, the detection of non-occurrences starts with the detection of a situation and ends if the environment changes concerning the use case relevant context attributes.

D. Implementation

The Situation Recognizer component is implemented using Java with Esper as its event processing unit for java applications. At startup time the use case descriptions are parsed and all action discovering patterns are generated as EPL statements. Additionally, the global context that gets appended to each incoming event gets initialized. This guaranties that all use case relevant environmental aspects are related to single events. The global context definition is part of each use case description. A preprocessing step for the context in form of a road detection algorithm is outsourced into a separate library described in the next Section.

VI. Map Matching

In order to group actions by their occurrence on certain roads it is necessary to first define the type of information we like to group. In mobile applications the location information often comprises of a latitude and a longitude value. Applying well-known density-based clustering algorithms like DBSCAN [13] or MRStream [14] would result in arbitrary clusters like the sum of all locations or groups of rectangular regions. In case of an automotive environment the clusters do not consist of arbitrary two dimensional regions because the car movements are always constrained by roads. Therefore, a cluster in the sense of a car is either a road segment or subgraph of the whole road network. However, GPS sensors provide latitude and longitude values that need to be transformed into a unique road id identifying a certain road segment. With the help of the road segment id it would be possible to define a notion of similarity between road segments by considering their road network distance.

The transformation of the latitude and longitude value into a position that lies on the road network is called map matching. The presented prototype includes an implementation of a map matching algorithm that is used to preprocess the location data before starting the action grouping process. It is essentially the first part of an algorithm proposed by Hummel [15]. For our purposes this rather simple implementation is sufficient.

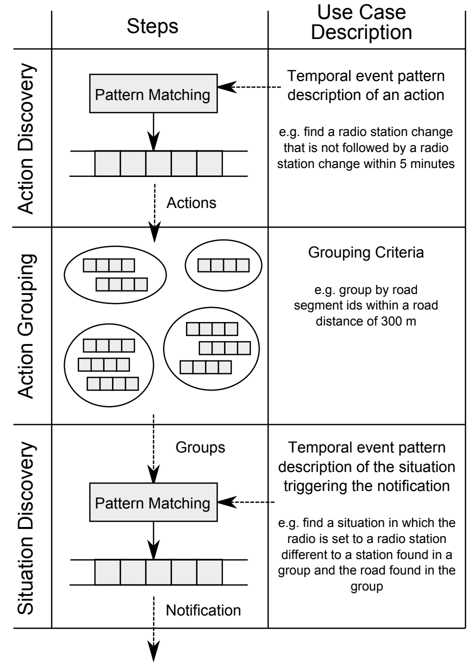


Figure 3. Overview Situation Recognizer

The topological data, needed to perform map matching, is based on the open source data provided by the OpenStreetMap project [16]. Before this data can be used, it has to be converted and inserted into a PostGIS database. The conversion was done with the free version tool called OSM2PO [17].

As mentioned above, the similarity between road segments needs to be calculated to decide if two road segments belong to the same group or not. Two road segments are declared to be similar if the road distance in between does not exceed a certain amount. The calculation based on the topological data is realized by pgRouting presented in [18].

It is also necessary to present the results in a human-readable way to be comprehensible for the HMI user. Therefore, a the street name needs to be extracted from the resulting road segment id. This service is also provided by the Map Matching library.

VII. Conclusion

We have presented a highly personalizable in-car-infotainment prototype that is capable to adapt itself based on the detection of regularities of its use in recurring environments. Architectural aspects as well as implementation details were discussed in order to describe primarily the technical problems.

The examples throughout this paper were mainly focused on use cases that try to adapt the HMI based on regular user

 ISBN: 978-1-61208-117-5

interactions at certain roads. However, it is up to the expert to decide, which dependencies should be used for a use case. For this purpose, we introduced the use case descriptions that are specified by an expert and used by the Situation Recognizer to control the whole process. While specifying the use cases the expert may decide to group button clicks only by time intervals and radio changes by time intervals and road segments.

VIII. Future Work

In the near future we plan to implement a feedback mechanism of the HMI in order to skip notifications that are not processed by the HMI. At the current stage, the Situation Recognizer works independently of any real execution of a personalization use case at the HMI. Therefore, it might happen that certain HMI personalization proposals were deactivated by the HMI user without notifying the Situation Recognizer. In this case, it is necessary to inform the grouping process to delete the corresponding groups or/and to prevent the grouping process to rediscover the same groups.

The approach of specifying and configuring personalization use cases in a generic way without the need to reprogram the Situation Recognizer offers a wide range of possibilities for rapid prototyping of personalizable applications. We plan to implement and test several new HMI adaptations that are based on the analysis of sequences of user interactions containing different road segments as well as time intervals. Especially, the prediction of the destination based on regular environments would enhance the usability of future navigation systems.

References

[1] D. Ashbrook and T. Starner, "Learning Significant Locations and Predicting User Movement with GPS," in *Int. Symposium on Wearable Computers*, 2002, pp. 101–108.

[2] G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," in *2nd Workshop on Context-Aware Recommender Systems*, 2010.

[3] J. Shen, E. Fitzhenry, and T. G. Dietterich, "Discovering frequent work procedures from resource connections," in *Proceedings of the 13th Int. Conf. on Intelligent user interfaces*, 2009, pp. 277–285.

[4] O. Brdiczka, N. M. Su, and J. B. Begole, "Temporal Task Footprinting: Identifying Routing Tasks by Their Temporal Patterns," in *14th Int. Conf. on Intelligent User Interfaces*, 2010, pp. 281–284.

[5] O. Coutand, S. Haseloff, S. L. Lau, and K. David, "A Case-based Reasoning Approach for Personalizing Location-aware Services," in *1st Workshop on Case-based Reasoning and Context Awareness*, 2006.

[6] M. Baldauf, S. Dustdar, and F. Rosenberg, "A survey on context-aware systems," *Int. Journal of Ad Hoc and Ubiquitous Comput.*, vol. 2, no. 4, pp. 263–277, 2007.

[7] A. K. Dey and G. D. Abowd, "The context toolkit: Aiding the development of context-aware applications," in *Proc. of the Conf. on Human Factors in Computing Systems*, 1999, pp. 434–441.

[8] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello, "Extracting Places from Traces of Locations," in *Proc. of the 2nd ACM Int. Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, 2004.

[9] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering personally meaningful places: An interactive clustering approach," *ACM Trans. Inf. Syst.*, vol. 25, p. 12, 2007.

[10] L. Liao, D. Fox, and H. Kautz, "Location-Based Activity Recognition using Relational Markov Networks," in *Proc. of the Int. Joint Conf. on Artificial Intelligence*, 2005.

[11] S. Rogers, C. nicolas Fiechter, and C. Thompson, "Adaptive User Interfaces for Automotive Environments," in *Procs. IEEE Intelligent Vehicles Symposium 2000*, 2000, pp. 662–667.

[12] EsperTech Inc., "Complex Event Processing," http://esper.codehaus.org/, Last access: 2010-12-20.

[13] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of 2nd Int. Conf. on Knowledge Discovery in Databases and Data Mining*. AAAI Press, 1996, pp. 226–231.

[14] L. Wan, W. K. Ng, X. H. Dang, P. S. Yu, and K. Zhang, "Density-based clustering of data streams at multiple resolutions," *ACM Trans. Knowl. Discov. Data*, vol. 3, no. 3, pp. 1–28, 2009.

[15] B. Hummel, "Map Matching for Vehicle Guidance(draft)," 2006, http://www.mrt.uni-karlsruhe.de/z/publ/download/hummel2006b.pdf, Last access: 2010-12-20.

[16] OSM Project, "Open Street Map," http://www.openstreetmap.org/, Last access: 2010-12-20.

[17] C. Moeller, "OSM2PO," http://www.osm2po.de, Last access: 2010-12-20.

[18] pgRouting Project, "pgRouting," http://www.pgrouting.org/, Last access: 2010-12-20.

 ISBN: 978-1-61208-117-5

Virtual Reality Technologies: a Way to Verify Dismantling Operations

First application case in a highly radioactive cell

Caroline Chabal, Alexandre Proietti, Jean-François Mante, Jean-Marc Idasiak
CEA, DEN, SDTC
Bagnols-sur-Cèze, France
caroline.chabal@cea.fr, alexandre.proietti@gmail.com, jean-francois.mante@cea.fr, jean-marc.idasiak@cea.fr

***Abstract*—Dismantling is a major challenge for nuclear companies, which are faced with the clean-up of former nuclear sites. In order to increase efficiency, optimize costs and planning, intervention designers must verify scenario key points, take into account unexpected situations and provide technical answers. Simulation is a good means of visualizing and therefore understanding constraints, of testing different alternatives, and a way to train workers prior to interventions. This paper describes an application of such a technology: dismantling a chemical cell in the APM (Marcoule Pilot Workshop) facility at Marcoule (France). This highly radioactive cell will be dismantled by a remote handling system using the Maestro slave arm. An immersive room has been used to design the dismantling scenarios. For these, the Maestro slave arm has been coupled with a haptic interface and, thanks to force feedback and visual immersion, accessibility, operational trajectories and maintainability on the carrier have been verified. If problems are found, updates of the carrier design are carried out before its final construction to guarantee the system will work properly. We describe the processes of building the 3D model and verifying scenarios. Finally, we present the first results, which are encouraging, and the perspectives for the project.**

Keywords-virtual reality; dismantling operation; haptic interface; accessibility study; remote handling; collision detection; interactivity; real-time

I. INTRODUCTION

The CEA (French Atomic and Alternative Energies Commission) dismantling division runs an R&D program to provide innovative simulation tools for decommissioning projects. The various software and hardware tools are based on Virtual Reality (VR) technologies, and enable a user to interact with a computer-simulated environment, whether that environment is a simulation of the real world or of an imaginary world. VR environments, mostly based on visual immersion and displayed either on a computer screen or through stereoscopic displays, can also include additional sensory information, such as sound or touch [1].

This paper describes how VR technologies, adapted to the nuclear decommissioning context, can provide useful support to engineers in charge of scenario design. Before beginning the actual operations, such a set of tools is also well adapted to communicating and sharing information during project reviews, or to training workers and ensuring they are aware of the risks they could be exposed to. It is the first time in the world that such technology has been used to define decommissioning scenarios. VR showed its advantages in other industrial or medical fields, like automotive conception or surgical training [2].

First, the chosen VR technologies will be presented. Secondly, the first application case will be presented and explained as well as the nuclear environment and the remote handling system used for dismantling. Then, we will describe the simulator developed to validate scenarios.

In the last section, we will describe our first results and the perspectives.

II. THE MARCOULE IMMERSIVE ROOM

The CEA created the Marcoule immersive room at the end of 2008 in order to validate maintenance or dismantling operations. It is a resource shared by all the CEA decommissioning projects (about ten), and can be used for project reviews, for accessibility, ergonomics or scenario feasibility studies, and for training workers [3].

The team works on new plant design as well as dismantling projects.

A. The hardware equipment

1) Stereoscopy

The room is equipped with a very large screen, 3.7m by 2.3m. After examining the options available, we have chosen the Infitec (*INterference FIlter TEChnology)* passive stereoscopic technology and use two projectors from the constructor Projection Design to display the images by back projection. The largest resolution is 1920x1200. Special interference filters (dichromatic filters) in the glasses and in the projector form the main item of technology and have given it this name. The filters divide the visible color spectrum into six narrow bands - two in the red region, two in the green region, and two in the blue region (called R1, R2, G1, G2, B1 and B2 for the purposes of this description). The R1, G1 and B1 bands are used for one eye image, and R2, G2, B2 for the other eye. The human eye is largely insensitive to such fine spectral differences, so this technique is able to generate full-color 3D images with only slight colour differences between the two eyes [4].

The main advantages of this technology are the very light weight of the glasses and it allows the user to turn his head with keeping a good stereoscopy. The size of the screen is very comfortable to work on life-size simulations.

 ISBN: 978-1-61208-117-5

2) Motion capture

A motion capture system based on four IR cameras manufactured by AR-Tracking is used to track the position of the specially-equipped glasses in real-time. As a result, when the user moves his head, the point of view of the simulation changes as if a genuine movement had taken place within the VR surroundings.

3) Haptic device

The room is also equipped with a haptic interface, the Virtuose 6D35-45 (Fig. 1), which enables the manipulation of virtual objects and force feedback to simulate the contacts. This device is developed by Haption, a spin-off of the CEA and it is the only product on the market today, which offers force feedback on all six degrees of freedom (DOF) together with a large workspace and high torques [5].

Figure 1. Virtuose 6D35-45.

B. *The software tools*

We use TechViz XL [6] developed by TechViz, a French company in order to catch the OpenGL flow of an application, generate stereoscopic images and send it to both projectors. TechViz XL notably works with 3DSMax, SolidWorks or Virtools.

Virtools produced by Dassault Systèmes is used to manage a simulation. It offers a development environment to create 3D real-time applications and thanks to its Software Development Kit (SDK), we can add our own functionalities.

III. First Application: Cell 414

A. *Presentation of the project*

The APM facility is the Marcoule Pilot Workshop. It was a prototype plant for reprocessing spent fuel, first commissioned in 1962, with production activities shut down in 1997. The plant is currently undergoing clean-up and dismantling.

Cell 414 (Fig. 2) is a part of the APM and was a chemical unit used to process liquids from irradiated fuel dissolution operations. It is a particularly large cell: 20m long, 4m wide and 6m high. There is approximately 5km of pipes. The present high level of radioactivity rules out direct manual dismantling, so the choice of a remote handling system called Maestro has been made.

The first step of decommissioning is to remove high level radioactivity (hot spots). Data was gathered from an initial inventory: hot spots were identified with a gamma camera. These hot spots like the centrifuges or some parts of the pipes have to be removed first in order to reduce cell radioactivity.

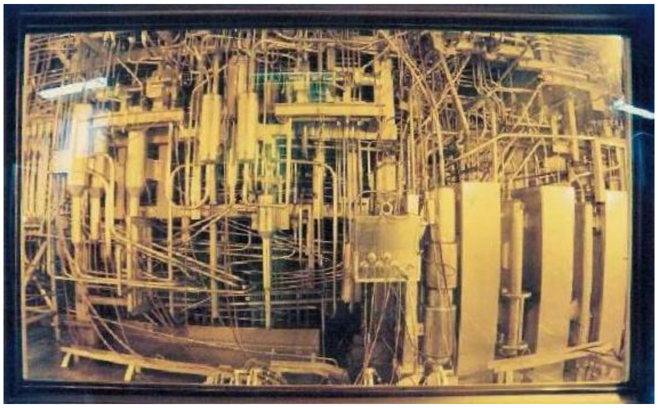


Figure 2. Cell interior seen from a porthole.

B. *The remote handling system*

1) The Maestro system

The Maestro system is the result of 10 years of collaboration between the CEA and Cybernetix in charge of its manufacturing [7]. This advanced remote manipulator is used when human intervention is not possible as if to operate in nuclear or offshore hostile environments. Maestro is dedicated to many tasks like inspection, maintenance, dismantling, cleaning…etc. Dexterity, accuracy and strength are its main advantages. It can be used in either robotic mode (automatic sequence) or in manual remote control mode with or without force feedback management.

This system is made up with two parts: the master arm and the slave arm. The master arm is a device allowing the control of slave arm end-effecter in Cartesian mode with a complete force feedback. This device is a Virtuose 6D40-40 from Haption. The slave arm is a hydraulic robot with six degrees of freedom (Fig. 3).

The cell 414 dismantling project will be the first working site where Maestro will be used to dismantle a whole cell.

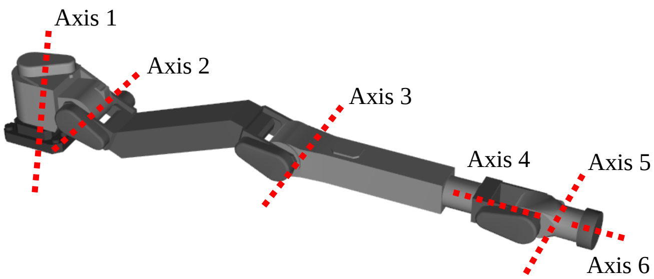


Figure 3. The Maestro kinematics

2) The carrier

The carrier was especially designed for Cell 414 dismantling, and will enable the Maestro system to reach all parts of the cell.

It works on three axes, using existing rails to move along the cell (20m), with vertical (3m) and rotating movements. A crane-type handling bracket is also set up on the carrier to hold parts during dismantling and for other handling operations. This carrier is currently undergoing tests.

The carrier-crane combination also has six degrees of freedom (Fig. 4), with four translation axes and two rotation axes.

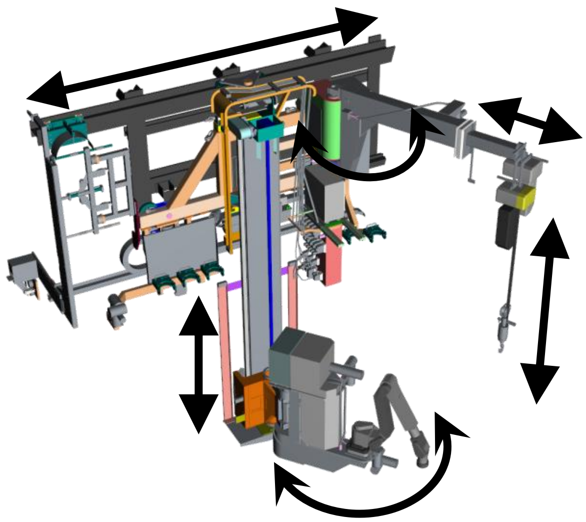


Figure 4. The carrier kinematics

IV. From real to virtual : the steps to build the simulation

In order to verify accessibility and maintainability on the carrier and to validate technical choices, it was decided to design the dismantling scenarios using a simulator and the VR technologies available in Marcoule.

A. Step one: build the 3D models

1) The cell 414 and the APM facility

First, 3D models of the environment had to be built. As the 2D drawings of facility inside were not sufficiently up-to-date to design a precise digital mock-up, a photogrammetric technique was used. A remote controlled camera was brought in the cell to take a set of pictures from different points of view. Then the 3D model was built using a semi-automatic process, to produce a model compatible with CAD or modeling software. About 200 pictures were necessary for the 3D reconstruction and only the internal parts were modeling in this way. This step was made by an external company.

Secondly, the modeling of the building containing the cell 414 was made with the plan of construction.

Finally, we merged these two parts to get a whole model into the 3DSMax software. The images below enable the comparison between a real photo and a VR view of the same scene. We can see that the 3D simulation is very close to reality (Fig. 5).

Figure 5. A real photo (left) and 3D view (right)

2) The robots

Dealing with the robots previously presented, we obtained the CAD model by the manufacturer of the real robots: Cybernetix. The modeling is in SolidWorks format and we did the necessary conversions to use it into 3DSMax.

B. Step two: develop the simulator

1) The Virtools environment

Our simulator is based on Virtools 5.0. It uses a specific script language and functions called Building Block (BB), in order to provide the interaction between the different objects, create menu, play sound, etc.

To import 3D models into Virtools, they must be in a specific format. That is why we use 3DSMax because it provides an exporter from .MAX format to .NMO format used by Virtools.

Virtools owns a basic physics engine but we use a third party physics engine because it is more precise and adapted to the force feedback. It will be presented in the following paragraph.

2) Physics engine: IPSI

IPSI [8] is a physics engine provided by Haption. It enables the testing of intersections between volumetric solids, in order to calculate trajectories and impact points. The real-time collision detection disables penetration between objects. Volumetric existence is given to geometric objects by taking the external skin of each 3D object.

IPSI allows the creation of the kinematics of a robot: hierarchy between the different segments of a robot, the kind of degrees of freedom and its values, etc.

Finally, IPSI can attach a virtual object to the Virtuose device and calculate the force feedback in case of collisions.

3) The simulator

Thanks to the Virtools SDK and the IPSI API, we can use Virtools for the graphical part and IPSI for the physical part. We created a Dynamic Link Library (DLL) for Virtools to add the abilities of using the functionalities of IPSI.

In the initialization of the simulation, all the 3D objects we want to add to the physical simulation are sent to IPSI as well as the information about robots (hierarchies, degrees of freedom, etc). So we create kinematics of the Maestro arm

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-117-5

and the carrier. As a matter of fact, we can consider both objects as robots and describe very precisely their motion. It consists in giving properties, by defining rotation or translation axis and end stops given by Cybernetix. Virtual robots can then be manipulated with their constraints as in real.

The graphical representation of the objects is updated in Virtools by IPSI which calculates the new position in real-time. During the simulation life, we use a callback function to match graphical and physical objects.

C. Step three: control the simulation

To control the robots, two gaming joysticks are used to pilot the carrier and the crane. The first one controls the three DOF of the carrier and the second one those of the crane. These controls are very similar to the interface which will be used for the final dismantling system.

The Maestro arm has been coupled to the Virtuose 6D 35-45 haptic interface. The Virtuose enables manipulation of the Maestro end-effecter, and thus control of the Maestro extremity while respecting the kinematics chain and all the end-stops.

The entire robot can be maneuvered in real-time with these devices.

D. Step four: add interactive functionalities

An interactive real-time simulator was developed where the whole cell, the Maestro slave arm and the carrier are loaded. The Maestro arm and its carrier can be maneuvered using the joysticks and Virtuose. Any of the nine available tools like drill or angle grinder can be connected to the Maestro arm or changed, as necessary.

The points of view of the six cameras (two available in the cell and four embedded on the carrier) are also displayed in the simulator. It has been checked that every part of the cell is visible. Indeed, the operator will not perceive the cell directly during the dismantling. He will be in a room with six monitors to view the pictures being transmitted from the six cameras. There will be also a microphone in the cell, so the operator will can hear the sound of collisions in the monitoring room. Therefore we associated a sound with collision in the simulator, to enhance the information sent to the user.

Lastly, automatic scenarios, such as the carrier entry in the cell, are programmed.

V. FIRST RESULTS

Tests carried out on the system had two objectives : first, to check that the carrier design was suitable for the Cell 414 environment, and second, to verify the whole dismantling operation design [9].

Two interface problems preventing the forward movement of the carrier were quickly identified: while the first obstacle could be avoided by raising the Maestro base, the second will have to be dismantled by existing in-cell equipment, before the carrier enters the cell.

We have then verified the detailed dismantling scenario from the carrier entry to the cutting of centrifuges. We found a lot of technical key points which need to be clarified because the feasibility of the task is not easy. For example, the tool grasping seems to be a problem because the Maestro slave arm is in a limit configuration and the carrier has to be in a specific position (Fig. 6). It disables the grasping in some parts of the cell. This kind of problems was not yet identified and the manufacturer has to take into account these issues.

The last work we have done focuses on the centrifuges dismantling. We proved that the planed scenario was not feasible. With the simulation we brought alternative solutions.

Thus, from the first simulation runs, the project has already brought vital information to implement in its dismantling scenarios. The chosen VR technologies have proved their worth, and the various capabilities of the Maestro system and carrier will continue to be tested as the dismantling project enters its next phase.

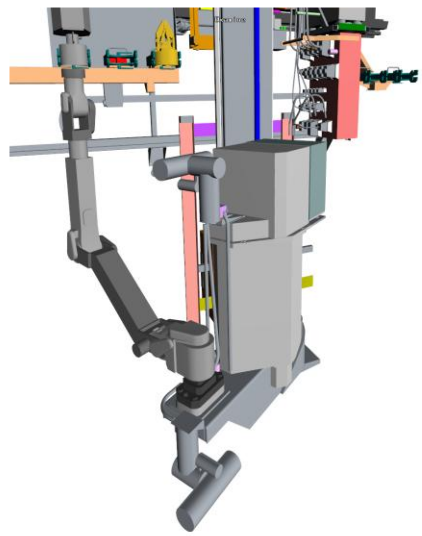


Figure 6. The tool grasping

VI. LIMITS AND PERSPECTIVES

A. Current limits

First, the mismatch of information relevant to reality can affect safety and performance. For instance, if the modeling accuracy of the robot or the cell is not enough, we cannot be sure that the scenarios that have to be tested with the simulator are reproducible in practice. The robot model directly comes from the CAD model of the manufacturer, so it can be considered as identical as the real one. The modeling incertitude is brought by 3D reconstruction. As a matter of fact, photogrammetry is accurate with 5cm precision. The most difficult task is to obtain a right model of the cell. The actual model created by photogrammetry is accurate enough for the first steps of scenario study but because of the layout of the cell, the complete model of the pipes could not be rebuilt with this technique. Only the first row of pipes were modeled so we have to update the cell modeling after the first steps of dismantling if we want to fit to the reality.

Secondly, we are limited by the physics engine. In deed, it is directly dependent on the computing power. With the current hardware, we cannot physicalize the robots and the whole cell with a high precision for the collision detection and get a real-time simulation. That is why we only physicalize the robots and some interesting parts of the cell. These parts depend on the tested scenario. Collision detection precision has to be inferior or equal to 10mm so that the accessibility studies could be realistic.

B. Add the radioactivity dose rate information

The CEA in collaboration with Euriware, a French company, developed an application called NARVEOS [10] capable of calculating the radioactivity dose rate. It is specifically used to simulate scenarios in nuclear environments. In NARVEOS, we can import a 3D model of a nuclear facility, specify the kinds of materials (steel, lead, concrete …) and add radioactive sources, protection screens, and measurement points in the 3D model. The software is able to calculate the radioactivity level in these points in real-time. For instance, we can measure the radioactivity dose rate on an operator moving in the facility.

In a near future, we want to assembly the functionalities of NARVEOS with our simulator. In this way, we will be able to follow in real-time the decrease of radioactivity level during decommissioning and calculate the new levels after the removal of hot spots.

C. Train the operators

From the beginning of this project, the idea of training operators was predominant. The models are very closed to the reality and we can work with a life-size simulation. Currently, the control of the robots with the joysticks plus the Virtuose device, allow testing the real robot motion in the cell. For instance, we can find the most fitted carrier positions to work at best with the Maestro slave arm. We can also use the simulation to increase the operators' awareness of the risks they could be exposed to, like collisions between carrier and equipments or robot damage.

Moreover, the main purpose of the training is to avoid nuclear incidents like workers' irradiation. That is why we want to use the radioactivity dose rate simulation to train operators and inform them where the radioactive areas are located.

Another advantage of the training is to show operators that there is no direct vision, so they will get used to working with only video and sound monitoring from the cell.

VII. CONCLUSION

This project shows that VR technologies can contribute to improve knowledge regarding project preparation and validate technical choices. The simulator involved is generic and can load any 3D model of a building. It is already functional and useful for the operator's training. The CEA has also compiled a comprehensive robotics library and can therefore run VR versions of scenarios with any of these systems in order to test alternative solutions.

Given the first results, the CEA proved VR tools open up new perspectives for studies, for decommissioning cost and deadline management, as well as for communication between project teams, contractors and Nuclear Safety Authority.

REFERENCES

[1] F. Gosselin, J. Martin, C. Andriot, and J. Savall, "Large workspace haptic devices for human scale interaction", LIST/Réalité Virtuelle/UE-SKILLS, 2008.

[2] F. Yaacoub, "Development of virtual reality tools for arthroscopic surgery training", Université Paris-Est, Phd thesis in computer science, Paris, France, 2008.

[3] G. Rindahl, T. Johnsen, and Y. Iguchi, "Virtual reality technology and nuclear decommissioning", Int. conference on safe decommissioning for nuclear activities, Berlin, Germany, 2002.

[4] H. Jorke and M. Fritz, "INFITEC- a new stereoscopic visualisation tool by wavelength multiplex imaging", Journal of Three Dimensional Images, vol.19, pp. 50-56, Japan, 2005.

[5] P. Garrec, J-P. Friconneau, and F. Louveau, "Virtuose 6D: a new force-control master arm using innovative ball-screw actuators", ISR, Paris, France, 2004.

[6] C. Limousin, J. Sebot, A. Vartanian, and N. Drach, "Architecture optimization for multimedia application exploiting data and thread-level parallelism", Journal of Systems Architecture, vol. 51, issue 1, pp. 15-27, January 2005.

[7] O. David, Y. Measson, C. Rotinat, F-X. Russotto, and C. Bidard, "Maestro : a hydraulic manipulator for maintenance and decommissioning application", ENC, Bruxelles, Belgium, 2007.

[8] Haption, IPSI datasheet. Available from http://www.haption.com/site/eng/images/pdf_download/Datasheet_IPSI.pdf.

[9] C. Chabal, A. Proietti, and E. Rigaux, "Virtual Reality, a way to optimize decommissioning scenarios", Poster, ENC, Barcelona, Spain, 2010.

[10] J-B. Thevenon, L. Lopez, C. Chabal, and J-M. Idasiak, "Using simulation for intervention design in radiating environment: first evaluation of NARVEOS", GLOBAL, Paris, France, 2009.

 ISBN: 978-1-61208-117-5

Contextual Spaces with Functional Skins as OpenSocial Extension

Evgeny Bogdanov, Christophe Salzmann, Denis Gillet
Computer, Communication and Information Sciences
Ecole Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
{*evgeny.bogdanov, christophe.salzmann, denis.gillet*}*@epfl.ch*

***Abstract*—Portability, flexibility and extensibility are essential features of social media platforms. When such Web platforms are able to take user's context into account, they provide better user experience and enhance the effectiveness of users' actions. In this paper, we discuss an extension to OpenSocial standard, namely contextual space, that shapes the framework, in which people carry out online activities. The proposed contextual space extension defines how a set of OpenSocial widgets are aggregated as a Web environment for a given purpose and with a given functional skin as a user interface. Additionally it allows to create contextualized widgets. In this paper we discuss the proposed extension in details and provide the examples of its use based on real life scenarios. Finally, we detail an implementation scheme.**

***Keywords*-context, contextualized widget, contextual space, functional skins, portability, OpenSocial, social media platform, Web application, Web environment.**

I. Introduction

The number of social media platforms is rapidly increasing. After the advent of popular platforms such as Friendster, Facebook and MySpace, which are oriented mainly on management of users' list of friends, a great number of new social media platforms appeared and continues to appear. These new platforms are often of a narrow orientation: Twitter focuses on micro-blogging, Youtube specializes on online video streaming, Last.fm is a network for people that like music. Despite their different goals, all of these platforms integrate a notion of friendship and/or connections between people.

Some of these social media platforms started to provide APIs to allow external developers to create widgets (small applications that can be embed into a Web page). Such widgets represent usually a combination of HTML and Javascript code and can easily be added by users into their pages. They can as well take into account information about a user from its hosting platform. Widgets added into users' pages extend the default functionalities of a hosting platform and bring greater personalization experience to users. Thanks to these widgets, social media platforms can attract external developers to augment their functionalities.

OpenSocial Foundation provides a standard API [5] to retrieve information about a user from a social media platform. A special widget standard (OpenSocial gadget) was first proposed and implemented by Google and later spread in other platforms. These widgets can be found everywhere ranging from simple blog sites such as Blogger to social media platforms (Myspace, Orkut, Friendster, etc.) and business-oriented networking solutions (LinkedIn, Oracle, XING, etc.). A social media platform that implements support for OpenSocial can ensure that any widget implemented according to the standard will run properly when plugged into this application. In other words, the widget becomes portable and can be run in different social media platforms. With the advent of a reference open source implementation for OpenSocial API (Apache Shindig), any social media platform can quickly start hosting OpenSocial widgets.

Despite the fact that OpenSocial solves the problem of social media platform's extensibility and widgets portability, it has two major limitations. First, it is user-centric but does not take the user's context into account. In many social media platforms context is a crucial component and if widgets were able to retrieve the user's context (from the hosting platform), it would greatly improve user's experiences [9], [15].

To understand the second limitation, one should distinguish between two types of Web applications, namely, Web widgets and Web environments. While Web widgets are usually considered to be small Web applications (weather forecast widget, translator, calendar), a Web environment is a relatively complex Web application such as a forum, a text editor or a personal learning environment. These Web environments can be seen as meta-components since they might have the capability of managing other Web applications. OpenSocial standard provides good support for Web widgets, but lacks support for Web environments.

In this paper, we introduce a notion of a contextual space as an extension to OpenSocial. This extension enables the definition of portable Web environments combining widgets that can be presented to users with different functional skins to take context and preferences into account. In addition, being the representation of a user's context, space permits the development of contextualized widgets. In this paper, we present the extension details and show the scenarios, in which such extension could be beneficial. Furthermore, we describe an implementation that uses the concept of spaces and we demonstrate portable Web environments as proof-of-concept.

There exist different standards for widgets: W3C standard [4], OpenSocial gadget specification and some other proprietary standards. In this paper, we will be using indifferently "widget" to refer to all standards. Similarly, we will refer to OpenSocial widgets for OpenSocial gadgets.

The paper is organized as follows. Section II discusses the OpenSocial specification. In Section III, we discuss the details of th proposed extension. In Sections IV,V,VI, contextualized widgets, portable Web environments and functional skins are defined respectively. Finally, the implementation of the proposed extension is discussed in Section VII. Section VIII concludes the paper.

II. OpenSocial specification

OpenSocial specification [5] provides a set of common APIs for social media platforms. This API standardizes the way information about people, their friends, application data and people's activities are retrieved. This standard solves two problems at the same time. First, a social media platform developer does not have to create a new naming scheme for his/her API and, secondly, widgets developed according to the OpenSocial specification become portable and can run on any Web platform that implements OpenSocial.

The only requirement for a social media platform is that it has to support OpenSocial specification. Thanks to Apache Shindig [1] (open source reference implementation of OpenSocial specification) any social media platform can quickly add support for OpenSocial and be able to instantiate OpenSocial widgets. By implementing this specification, a social media platform automatically gets the opportunity to extend its own functionality by plugging in OpenSocial widgets.

III. Contextual space and OpenSocial

OpenSocial specification focuses mainly on users. It provides APIs to retrieve list of friends for a given user, list of user's activities, albums, list of user's widgets, etc. It is known, however, that depending on user's context, a list of context-specific tools is needed. A list of people can change from one context to another. The same person might have a completely different role in different contexts. Such context could be a workspace, a forum's topic, a friends discussion or a trip planning. Social media platforms often support the context concept in one form or another. As an example, there might be different discussions and sub-discussions in the same forum, each one with its own topic, own list of participants and own list of resources. Groups or events of Facebook is another example of a context. In case of Learning Management System, every user can have several contexts - one for every course the user takes part in. It is important that Web applications get access to user's context and are able to take it into account.

One of the main limitations of OpenSocial is that it does not support user's context. Such concept simply does not exist in OpenSocial. Currently with OpenSocial it is impossible to model, for example, a university course with participants and resources. Contextual spaces as OpenSocial extension are intended to provide support for context to OpenSocial specification and make OpenSocial widgets more useful for people.

Later in the paper, we will narrow down the definition of the context to a contextual space. We define a contextual space as an aggregation unit that includes list of applications that are to be used in the context, a list of people with different access rights sharing the context, resources that can be used inside such context, and possibly some other subspaces that belong to this context. As the 3A-model [11] suggests, such mapping can be done for any social application [8]. Any social media platform's structure can be mapped into three different entities: Actors, Activities and Assets, where Actors represent people, agents, applications; Assets represent resources, documents, etc.; Activities represent an aggregation unit that combines different Actors, Assets and another Activities together into a context. Activity is a contextual space in our definition and represents the user's context. A contextual space can be further enriched with additional fields to better represent the user's context.

It should be noted that support for people, resources and tools already exists in OpenSocial. However, all these entities are centered around the user and not around the user's context. To avoid confusion, we should note that "Activity" exists in OpenSocial specification but this concept is completely different from 3A-model concept. While in 3A-model word "Activity" means an aggregation unit or a contextual space, in OpenSocial it means an action done by people (user sent a message, user became friends with somebody, etc.), which are two completely different concepts.

Table I presents the proposed OpenSocial extension with contextual spaces. In OpenSocial the "People" service is responsible for retrieving a list of people connected to a user. With the contextual space extension, there is additionally a list of people for every space (for example, list of group members). In order to handle both scenarios we suggest to add a field "type" to the API. By doing this, a pair (gid,type) defines either a space or a person, for which a list of connected people has to be retrieved.

"AppData" service of OpenSocial can be extended in a similar way. By default, this request returns application data for a widget that belongs to a user. With the new extension, a widget can belong to either a person or to a space. Thus application data can be retrieved for widgets belonging to either people or spaces. "Type" parameter in the request allows to specify for which item (space or person) application data should be retrieved.

In addition to above extensions of existing OpenSocial services, we suggest to introduce two new services, namely, "Spaces" and "Applications". Requests for "Spaces" are similar to those of "People" service. The first request in

 ISBN: 978-1-61208-117-5

People	**OpenSocial: /people/{guid}/@all** - Collection of all people connected to user {guid} **Extension: /people/{gid}/@all/{type}** - Collection of all people connected to item with id {gid} and with {type} in ("space","person")
AppData	**OpenSocial: /appdata/{guid}/@self/{appid}** - All app data for user {guid}, app {appid} **Extension: /appdata/{gid}/@self/{appid}/{type}** - All app data for item with id {gid} and with {type} in ("space","person"), app {appid}
Spaces	**Extension: /spaces/{spaceid}/@self** - Profile record for space {spaceid} **Extension: /spaces/{gid}/@all/{type}** - Collection of all spaces for item with id {gid} and with {type} in ("space","person")
Applications	**Extension: /applications/{appid}/@self** - Profile record for application {appid} **Extension: /applications/{gid}/@all/{type}** - Collection of all applications for item with id {gid} and with {type} in ("space","person")

Table I. OpenSocial extension with contextual spaces

the third row allows to retrieve detailed information about a space. The second request allows to retrieve list of spaces or list of people connected to a space. The last request is similar to the extension of "People" request, but instead of returning list of people, it returns list of spaces. With these two requests one can get lists of all people and spaces connected to a specific person, as well as lists of all sub-spaces of a space and lists of people for a space (or space members).

The service "Applications" retrieves detailed information about a given application with specific identifier (id). The "Applications" request with "type" parameter in the fourth row allows to get list of applications for a person or a list of applications that belong to a space. This request is similar to previously described requests for "People" and "Spaces" services.

Interested readers can refer to Apache Shindig wiki [2] and [12], where implementation details are presented.

IV. CONTEXTUALIZED WIDGETS

Assuming that the proposed contextual space concept is supported by the OpenSocial specification, widgets can retrieve user's context information (such as people, applications and resources in this context) from hosting social media platform. This would provide greater widgets personalization.

By taking context into account, one can create contextualized widgets, widgets that adapt their behavior to user's context. Such widgets can better extend the functionality of a hosting platform. Their visual interface, displayed data and functionality can be changed according to the context, in which user currently is interacting. The same widget might display different people and different resources depending on the user's context.

Let us look at a simple example coming from a learning scenario: a person is in the process of learning German and French languages. This person utilizes a widget that lists language documents from a space and another widget that lists space members. With the proposed OpenSocial extension it is enough for a user to create two spaces "Learn German" and "Learn French" in his/her social media platform. Then s/he can put German documents and friends learning German into "Learn German" space and French documents and friends learning French into "Learn French" space. When user enters "Learn German" space in the hosting platform, widgets display German documents and his/her friends from German space, and when user enters "Learn French" space, the corresponding French items are displayed in the same widgets. Thus these widgets become contextualized widgets.

V. PORTABLE WEB ENVIRONMENT

The notion of a space and its support in OpenSocial bring us new scenarios, that were impossible to realize before, namely portable Web environments. Such Web environment can be moved between two different social media platforms [13]. With contextual spaces it is possible to model such concepts as group of people, event, discussion, course, conference, etc. All these concepts have the following characteristics in common: they all serve as an aggregation unit to join together people, resources, applications.

A contextual space defines a Web environment and once OpenSocial specification is extended with contextual spaces, a Web environment can be technically implemented as an OpenSocial widget (such Web environment can be seen as a meta-widget [6], that is a widget that can integrate several widgets). People are used to consider a widget as a small application running in their desktop or their browser page. However, as Wikipedia states, "in computing a Web widget is a portable chunk of code that can be installed and executed within any separate HTML-based web page by an end user without requiring additional compilation. They are derived from the idea of code reuse." Thus, even though a Web widget and a Web environment represent different concepts, they both can be implemented according to OpenSocial widget standard (as "portable chunk of code"), which ensures portability for Web environments.

Let us consider a real life example where such approach can be used. Imagine a Web environment *App1* (that is a groups management application) implemented as follows: the server side code is implemented according to OpenSocial specification extended with spaces. The client has a widget container able to render OpenSocial widgets. Figure 1 depicts the Web environment *App1*. Learning groups are represented as horizontal tabs. When a tab with a group is active, the list of people for this group is shown in the right area and widgets for this group are displayed inside the left area side-by-side. This Web environment is technically implemented as a big OpenSocial widget that receives via OpenSocial list of groups (as spaces) for a logged in user, a list of people and widgets for every group. Since this Web environment is implemented as a widget, its functionality is portable and can be reused in another Web platform. If the owner of *App1* believes such widget (Web environment) to be useful for other people, s/he can add it into one of the available widget repositories (iGoogle, for example).

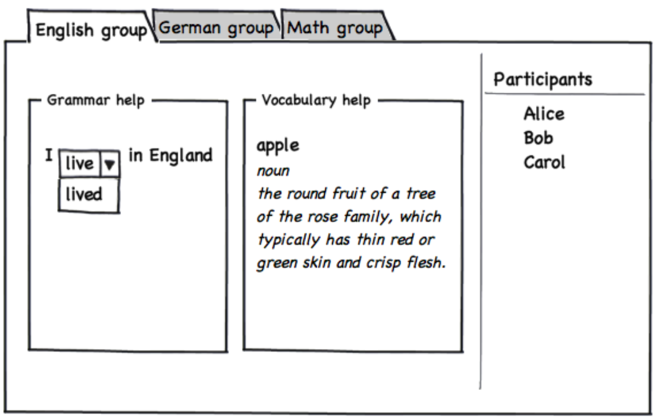


Figure 1. Web environment *App1*

Other people can reuse this Web environment according to the following scenario. A user plans to develop a new Web environment *App2* that will be showing news to people. The user would like to organize the news site in the following manner: s/he wants to have tabs that correspond to the different news topics (music, economics, sport). For every topic the user wants to add some RSS widgets that show news for the topic and a list of people subscribed to this topic. S/he would like to know if similar functionality is already implemented by someone else, so that she does not have to start from a scratch. S/he goes to a widget repository and finds a URI for the OpenSocial application. This application suits perfectly to his/her requirements. The structure and functionality is exactly as expected, however the data are different.

Assuming s/he already has the extended Apache Shindig installed on his/her server, s/he first creates topics (as spaces) and adds widgets and people to every topic. Then, s/he adds the retrieved widget and his/her site starts to function as planned (Figure 2).

This scenario shows that the Web environment functionality becomes portable and reusable. Even though the difference is transparent for an end-user, for Web developers

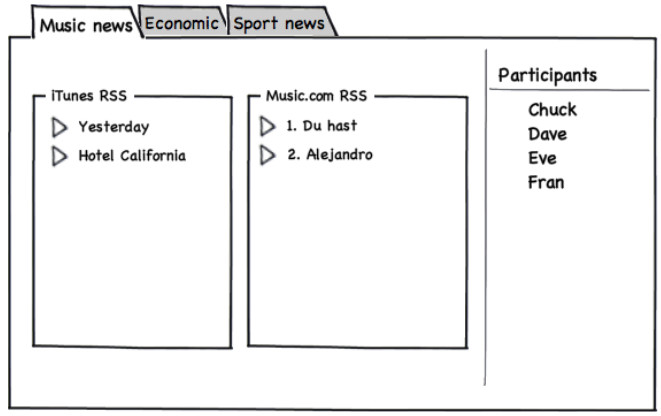


Figure 2. Web environment *App2*

this is a step forward in simplifying the development process. It is true that even without spaces Web widgets are reusable and portable, however without spaces it is difficult to create portable Web environments where the user's context is important (as in the described scenario).

The space extension and portable Web environments can have a great influence on end-user experience too. In the previous scenario we showed how different data can be represented in a similar way through the portable OpenSocial widgets with spaces. The next section presents functional skins, where the interaction with the same data structure can be offered to users differently through OpenSocial widgets extended with spaces.

VI. Functional skins

OpenSocial specification extended with spaces and Web environment (or meta-widget concept) allow people to easily change the way interaction with the information is offered. We define such different interaction schemes as functional skins, where data and data structure do not change, while visual representations and actions that user is allowed to perform with data (functionality) might differ from one functional skin to another.

The main idea behind functional skins is the same as for portable Web environments: Web application implemented as an Opensocial meta-widget that supports spaces and can integrate other widgets.

As an example, we consider the scenario where a knowledgable person in Computer Science decides to create a space to support learners in mastering Computer Science. For this goal, the mentor creates a space "Learn Computer Science". Then learners (Alice, Bob, Chuck) and some subspaces (Introduction, Basic Algorithms, Complex Algorithms) are added to the space. Then the mentor structures information and populates every subspace with widgets and resources helpful to master the Computer Science. The resulting view for a space "Basic Algorithms" is depicted in Figure 3.

 ISBN: 978-1-61208-117-5

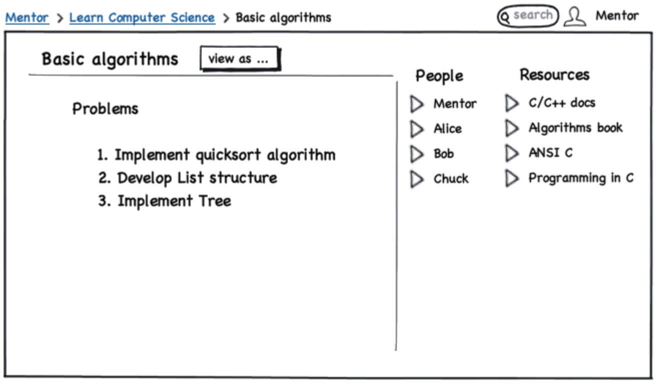


Figure 3. Default functional skin for Basic Algorithms space

During the studying process some of the visual parts are not needed for learners as they might be distracted from a learning process. Thus the mentor wants to provide a special view for learners that provides only the needed functionalities: two widgets displayed side-by-side and a list of resources as a column on the right (Figure 4).

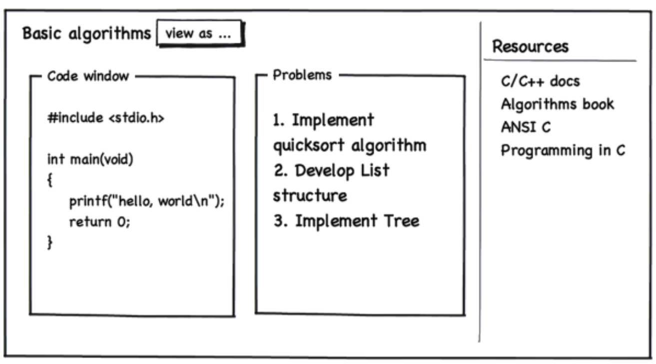


Figure 4. Learning-focused functional skin proposed by mentor

Mentors are provided with a way to change the visual representation of data structure through widgets. The mentor can go to a publicly available widget repository and find a widget that provides a functionality similar to Figure 4 (such widget could be developed either by mentors itself or by some other developers). URI for this widget is retrieved and added into the list of URIs for "view as ..." button in "Basic algorithms" space. Since this widget implements the extended OpenSocial specification, it correctly processes information about space, widgets and resources inside and it displays information as illustrated in Figure 4.

The mentor is only required to know the URI for this widget and no additional implementation is required on the user's side to add this widget as a functional skin for his/her space. Thus, the mentor can easily share this functional skin with other people. In the case of a different Web platform (iGoogle for example) that implements OpenSocial specification, this functional skin can be used also. Another interesting point is that with such technique, the mentor can greatly extend the default functionality by using third-party Web applications.

Moreover, learners can add and change functional skins themselves. If, for example, a learner would like to have an additional area showing a list of people in the bottom-right area under Resources (see Figure 4), s/he can either look for another functional skin in widget repositories or develop one him/herself. Afterwards, a new widget's URI can be added to the list for "view as ..." button, and learner can simply switch between different functional skins.

These new widgets are indeed skins, since, for example, a space can have several different functional skins that would display the learning space and enable interaction in different ways. However, we add the word "functional" because it does not only change visual appearance as skins normally do but functionality might also change from one skin to another.

This technique allows users to have "functional skins" for their Web environments. It provides both a way to change visual representation of data and actions to be performed with this data. One of this approach's main benefits is that user can easily change the provided default functionalities to work with data. This allows easier code reuse since the same functional skin can be used by different people for their own Web environments. People themselves can find and add new functional skins to work with data. Functional skins can be used in different Web platforms, by providing users with the flexibility in choosing a Web platform, in which they prefer to work.

VII. Implementation

Graaasp is a social Web platform with support for widgets and the proposed space concepts (Figure 5). Every user has a list of connected people (friends), list of widgets and list of spaces s/he is a member of. Every space has a list of members, a list of sub-spaces and a list of widgets connected to this space.

Apache Shindig is used to provide spaces extension for OpenSocial. However space extensions are not implemented exactly as described in Table I to preserve compatibility with current OpenSocial specifications. Instead of using "type" parameter, we added a prefix "s_" for spaces and "p_" for people into the "guid" parameter. Then on the server side we take prefix into account and return either a list of items for a space or a list of items for a person.

Graaasp's OpenSocial API extended with spaces is used in another Web platform, namely, *Rolespace* [3]. *Rolespace* implements a visual interface that is completely different from *Graaasp* user interface, however it uses the same data structure, since it takes spaces, people, widgets from *Graaasp* via OpenSocial API. The *Rolespace* can be seen as a functional skin for *Graaasp*'s data structure.

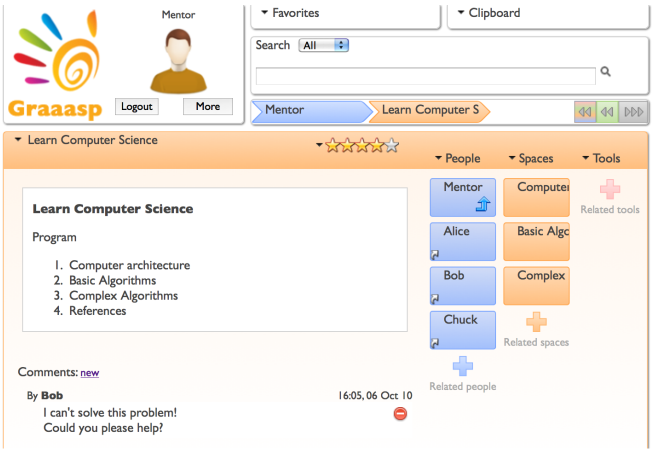


Figure 5. *Graaasp* Web platform

In addition, we developed for *Graaasp* a proof-of-concept implementation for meta-widget that integrates several small widgets. This meta-widget can run on different Web platforms that support OpenSocial widgets.

VIII. Conclusion

This paper presented contextual space as an extension to the OpenSocial specification. It summarizes the limitations of current version of the OpenSocial specification and highlights its user-centric focus without the ability to describe user's context. OpenSocial specification extended with spaces tackles these limitations and provides greater flexibility to users.

These OpenSocial extensions are described in details and three scenarios are presented. As a result, widgets can be contextualized, meaning that they have access to users' context and might adapt themselves to it.

The meta-widget concept combined with the space extension makes it possible to take advantage of functional skins (different representations and modes of interaction with the same data) and ensures portability for Web environments (same representation and mode of interaction with different data).

We conclude the paper with the details of our proof-of-concept implementation for the space extension and meta-widget. Further work is under way to provide a complete implementation for the proposed space extension that will lead to contextualized widgets, functional skins and portable Web environments.

Acknowledgment

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] Apache shindig. http://shindig.apache.org/index.html 13.12.2010.

[2] Apache shindig wiki. https://cwiki.apache.org/confluence/display/SHINDIG/Index 13.12.2010.

[3] Rolespace. http://graaasp.epfl.ch/role_project 13.12.2010.

[4] W3C set of standards for widgets. http://www.w3.org/2008/webapps/wiki/WidgetSpecs 13.12.2010.

[5] Opensocial specification v0.9. apr. 2009. http://www.opensocial.org/Technical-Resources/opensocial-spec-v09/OpenSocial-Specification.html 13.12.2010.

[6] M. Blattner, E. Glinert, J. Jorge, and G. Ormsby. Metawidgets: towards a theory of multimodal interface design. *Computer Software and Applications Conference, 1992. COMPSAC '92. Proceedings., Sixteenth Annual International*, pages 115 –120, sep. 1992.

[7] E. Bogdanov, S. El Helou, D. Gillet, C. Salzmann, and S. Sire. Graaasp: a web 2.0 research platform for contextual recommendation with aggregated data. *CHI*, pages 3523–3528, 2010.

[8] E. Bogdanov, C. Salzmann, S. El Helou, and D. Gillet. Social Software Modeling and Mashup based on Actors, Activities and Assets. *ECTEL - MUPPLE Workshop*, 2008.

[9] A. K. Dey, G. D. Abowd, and D. Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Hum.-Comput. Interact.*, 16(2):97–166, 2001.

[10] K. Fruhmann, A. Nussbaumer, and D. Albert. A psycho-pedagogical framework for self-regulated learning in a responsive open learning environment. *International Conference eLearning Baltics Science (eLBa Science 2010)*.

[11] S. E. Helou, N. Li, and D. Gillet. The 3a interaction model: Towards bridging the gap between formal and informal learning. *ACHI*, pages 179–184, 2010.

[12] R. Lopes, H. Akkan, W. Claycomb, and D. Shin. An opensocial extension for enabling user-controlled persona in online social networks. *4th International Workshop on Trusted Collaboration (TrustCol 09 - in conjunction with CollaborateCom 09)*, pages 1–5, nov. 2009.

[13] M. Palmér, S. Sire, E. Bogdanov, F. Wild, and D. Gillet. Introducing qualitative dimensions to analyse the usefulness of Web 2.0 platforms as PLEs. *IJTEL*, 2010.

[14] S. Sire, E. Bogdanov, M. Palmér, and D. Gillet. Towards Collaborative Portable Web Spaces. *ECTEL - MUPPLE Workshop*, 2009.

[15] M. Wolpers, J. Najjar, K. Verbert, and E. Duval. Tracking actual usage: the attention metadata approach. *International Journal Educational Technology and Society 11 (2007) In*, pages 1176–3647, 2007.

 ISBN: 978-1-61208-117-5

Gathering Interaction, Interface and Aesthetics Considerations in Product Design

Analyzing devices related to the accessibility of heritage

Marina Puyuelo Cazorla, Mónica Val Fiel, Francisco Felip Miralles
Escuela Técnica Superior de Ingeniería del Diseño
Universitat Politècnica de València
Camino de Vera, s/n
46022 Valencia, Spain
e-mail: mapuca@ega.upv.es, movalfie@ega.upv.es, frafemi@doctor.upv.es

***Abstract*—Interaction aesthetics is a substantial part of the design process of products and devices, especially those related to enhance the accessibility of heritage. This paper offers a perspective on the breadth of the concept interactivity through an extensive literature review and analysis of some of the areas closely linked to the concept: the wayfinding experiences, the visual and tactile perception of devices and the computers' user interfaces. All these will help to evaluate the efficiency of existing products, in order to outline some directions that allow designers to develop new devices that make information and cultural content accessible to all users. We analyzed the most common devices used in cultural heritage sites, based on various parameters relating to the level of interactivity. The main results show that most devices are dynamic and designed to facilitate mobility, but they are not interactive. Therefore, designers must continue their work in this direction.**

Keywords- aesthetics; interactivity; interface design; product design; built environment

I. Introduction

The concept of interactivity has become worn and trivialised in recent years, due especially to the enormous quantitative leap in the field of domestic technology. Any electronic device with an interface appears to take on the characteristic of being interactive, and the term is often simply a feature added to product advertising as a commercial strategy. Many of these devices are presented as being accessible and interactive, but in most cases the relationship users manage to establish with them hardly helps to access content or understand the device, so that sometimes the activity of the device is limited to offering a simple reactive response to a stimulus. This lack of definition of the term creates difficulties when it comes to assessing, comparing or questioning interactivity as a quality in object design. It also comes into serious conflict with interaction aesthetics because the opportunities for developing the aesthetic component will depend on factors intervening in the interactive process (object-user-context).

In view of the above, there is a need therefore to establish the limits of the term interactivity and study its real scope for the user in the process of relating to objects and the contents they offer. This work aims to reflect the concepts which intervene and merge in the subject of interaction differentiating three main levels to which the term refers, for better understanding of its particular interest for proposing or considering interaction aesthetics as an integral part of the design process.

Another stage of research and analysis would attempt to focus on the design of accessibility products, which provide an appropriate level of interactivity required to ensure that specific groups of users can use and understand given contents. Nowadays, accessibility involves planning interactive resources. Designing access to a functional, cultural or information resource involves considering the real figure of the user. Users can only participate in content and enrich their knowledge by establishing interaction with devices and experiencing correct feedback. It is worth asking to what extent current devices effectively ensure that access and which devices are more suitable in certain situations.

By identifying and clarifying interaction types, it will be possible to organise typologies of accessibility products in general and consider in particular products intended to bring heritage sites closer to the visitor.

The structure of this paper starts with the definition of interactivity and the study of the relationship with the concept of wayfinding. Then we analyze the influence of aesthetics in the interactive experience and the connection between accessibility, interface and interaction design. Finally, the paper ends with a classification and analysis of the most common devices related to accessibility to heritage. The strong and weak points in each design can be observed and identified to establish directions for future projects.

II. Definition and Levels for Interactivity

The simplified idea of interactivity as a stimulus-response cycle is very widespread but is somewhat limited as it identifies reaction with interaction without considering the extent to which feedback is involved. Any interaction refers to an information exchange and this happens when a user input has to produce a device output which in turn will condition the following input. Researchers who have referred to this basic interaction model include Maldonado and Bonsiepe, who proposed the idea of the feedback loop between machine and user [1]. Norman's more recent interaction models are based on that same idea [2].

Interactivity is a complex concept related to communicational exchanges which can involve human, objectual or electronic agents at the same time. Shedroff refers to the need for at least two independent participants in any interactive process, so that the response generated by one action provides relevant information and motivates a subsequent action [3]. Rafaeli has defined interactivity as 'an expression of the extent that in a given series of communication exchanges, any third (or later) transmission (or message) is related to the degree to which previous exchanges referred to even earlier transmissions' [4].

In view of the above, interaction can be defined as a process of continuous action and reaction between two or more agents who participate alternately, helping to create experiences through the information exchange.

Taking into account these theories which relate feedback to the interactive process, very different situations can be found. In particular is the fact that firstly many objects are used in a basic, unilateral way, without exploiting their particular features. Secondly, other products and devices intended to make content accessible do not offer the possibility of establishing a valid dialogue between user and system.

In a first step towards establishing a classification of interaction levels, there is an interesting proposal [5] that distinguish between two types of systems: static systems, which cannot act on their own or influence the environment and dynamic systems which act and influence the environment, either from mere reaction to a stimulus or from a more complex interaction. Thus, objects which admit manipulation and react to it are dynamic, although not all such objects can be considered interactive, because guaranteeing user response is not enough.

Another main characteristic of interactivity is the opportunity for the user to act on all the agents involved in the process. As Liu and Shrum [6] pointed out interactivity can be defined as 'the degree which two or more communication parties can act on each other, on the communication medium and on the messages and the degree to which such influences are synchronized' (p. 54). Similarly, Bettetini's definition highlights the multidirectionality of interactions, the user's active role in choosing the required information and the particular pace of communication [7]. Three typologies of interactivity can be established according to these agents:

- According to the level of control over structure and content [8]: reactive, where the user has little control over structure and content; coactive: provides control over sequence, pace and style; proactive: control over structure and content.
- According to the relationship between messages [4]: not interactive: a message does not relate to a previous message, reactive: a message relates only to the previous message, interactive: a message relates to the previous message and preceding messages.
- According to the participant agents [9]: user-machine interaction: the computer must respond to user actions, user-user interaction: communication between persons, user-message interaction: skill of the user to modify messages.

Thus, any discussion of interactivity involves considering these three areas and qualifying the degree to which the user manages to influence and modify each aspect related to the process.

Interactivity, as a communication-related phenomenon, is linked to the acquisition of knowledge and the development of cognitive skills [10], and so it is important to ensure correct design of interactive resources. From the point of view of the product, interaction is the physical relationship a user establishes with a given device or object and therefore interactive design must consider various disciplines such as usability and ergonomics.

Usability brings us back again to a broad conceptual framework. Its definition [11] shows that interactivity and usability share the same consideration towards the figure of the user, whose special features should be studied to gain a better understanding of the interactive process.

III. Interactivity and Wayfinding

Wayfinding originally meant orientation in an urban environment by interpreting signs [12] and has been expanded to include all inner orientation strategies in an architectural space [13], with reference to maps and directories. Assimilating this information is key for making decisions about where and how to move around and putting the decision into practice. Arthur and Passini define these two stages in the process highlighting the formulation of cognitive strategies inherent in complex decision-making [14]. Thus it can be said the appropriate design of a wayfinding system can facilitate cognitive accessibility to an environment.

Wayfinding resources can be architectural, graphic, auditory or tactile and therefore the range of competencies covers all disciplines working with these resources. These subjects have been studied by architecture with more focus on the legibility aspects of the built environment, aspects of image and structure [12][15], geography developing symbolic content [16], environmental psychology which has investigated perception of the environment [17], cognitive representation and orientation of behaviour processes [18][19] and many others. Each of these different approaches is nourished by references and research from complementary approaches.

Ensuring accessibility to informative content requires devices (static or dynamic) able to notify users of specific information efficiently. In interactive devices, the interface must show the user what actions are available at all times in order to establish a dialogue with the content or activate a helpful response for subsequent decision-making. This process is also key for developing teaching resources, as user learning depends on establishing correct feedback through the interaction.

Therefore an interactive environment requires a sensory map (visual, auditory and/or tactile) which shows users the content and/or actions to be taken and guides them through the process. The wayfinding concept applied to orientation in

 ISBN: 978-1-61208-117-5

a physical space may seem insufficient to establish parallels with interactivity applied to electronic devices as this happens more dynamically in a conceptual space where users do not have to move around a physical space, but around information. Nevertheless, to navigate and access information the user adopts similar behaviours to those based on feelings of confidence, efficiency, effectiveness, exploration or enjoyment, depending on user intention and the characteristics or conditions at each moment [20].

So, the concept of interactivity is linked to wayfinding in that users must be able to access and explore different alternatives or information from an interface on the path to making decisions towards an action and must obtain a response from the device which orients them.

IV. AESTHETICS AND INTERACTIVE EXPERIENCE

Aesthetic components are immanent specific values characteristic of design activity, capable of increasing the potential for interaction with objects. The conception of a product must integrate aesthetics in a way which is intrinsic to its functional resolution and that involves bearing in mind the elements which secure particular emotions. Danielle Quarante clarifies the factors which determine these elements and appeal to different orders: purely emotional factors, related to subjectivity; cognitive factors concerning our knowledge and culture; intellectual factors, which lead to the logical satisfaction of understanding a product; and psycho-physiological factors which lead to aesthetic pleasure in relation to the quality of our perceptions [21]. Aesthetic features are certainly not quantitative but related to taste, pleasure, sensation and many other individual, social, cultural and historical parameters which merge in the special poetics of the object.

The degree of satisfaction users obtain from interacting with a product is structured on different levels: all experiences a user has with a product influences to some extent the degree of realisation and development of the ego, touching its different facets [22]. These experiences can be related to the different types of pleasure products can provide: physical, psychological, social and intellectual [23][24], differentiating a lower or physical level and a higher or mental level [25].

At this point it is worth offering some considerations on visual and tactile perception. In recent decades we have extended the field of action and our perception has evolved in these areas as a consequence of the technological revolution and materials engineering which enhance the qualities of objects and environments favouring orientation, use and legibility of the built medium, enabling more spontaneous use. Although emphasising visual factors, in the 1960s Lynch was already pointing out the significance of these factors with reference to the built environment [12]: 'Structuring and identifying the environment is a vital ability […] Many kinds of cues are used: the visual sensations of color, shape, motion, or polarization of light, as well as other senses such as smell, sound, touch, kinesthesia, sense of gravity, and perhaps of electric or magnetic fields' (p. 3).

Pallasmaa's contribution is important as it argues for the primacy of the sense of touch as the first contact with the medium on which all the other senses are based [26]. He also refers to the role of peripheral and unfocused vision in our experience of the world and how it integrates us in the space: 'The very essence of experience is moulded by hapticity and by unfocused peripheral vision [...] In addition to criticising the hegemony of sight, the very essence of vision must be reconsidered' (p. 10).

When we attempt to reflect these issues concerning the different levels of perception individuals obtain from their relationship with the physical medium, then the variety and diversity of stimuli and responses to which they are exposed immediately becomes apparent. The question is how designers can implement what is truly meaningful [27][28] to establish an optimum emotional relationship between user and product or environment.

The concept of aesthetics applied to interactive products is a complex term, but necessary to establish an essential affective link between object and user. In this meaning, it is important to emphasise that aesthetics is not only based on the sensations deriving from surface appearance [29], but is conditioned by empirical factors related to use and function closely linked to a specific user and the task to be performed. As stated in recent researches [30], this interaction is also linked to product personality, which is important in order to make a consistent use. The interactive event provides the user with an expansive perception of the product which goes beyond the perception based on its appearance as it also concerns psychological and cognitive aspects related to achievement of the task and possible output.

V. ACCESSIBILITY, INTERFACE AND INTERACTION DESIGN

We have seen that by interacting with a device it is possible to consult and handle the information it contains, making accessible its functions, the tasks it can carry out and the knowledge it offers. Defining the concept of accessibility is not an easy task, as it depends on the specific context in question. We can consider accessibility as the set of all those conditions in the surrounding environment and the user which favour the achievement of a goal, whether it be economic, social, or in the case of our study, cultural.

One of the most significant agents for accessibility is interface design, which is the key to appropriate interaction. We can initially consider the interface as a multiple space between two agents which enables communication by performing codification and decodification operations, thereby guaranteeing an information exchange. Limiting the definition for electronic environments, Mandel [31] defined the interface as 'both computer hardware and software which presents information to users, enabling them to interact with the information and the computer' (p.14), thereby defining it as a medium through which users and computers communicate with each other, that is, the main agent for human-computer interaction.

The concept although with nuances between authors, refers to two key ideas: the interface as contact surface or place between two entities which are not necessarily physical [32] and as key agent in the exchange of information between the user and the computer, enabling a communication relationship to be established. As Gui Bonsiepe (1999) has pointed out [33]: 'A human interface is the sum of communicative exchanges between the computer and the user. It is what presents the information to the user and receives information from the user' (p. 42).

The interface is a common space for both interactors (a logical mechanical entity and an organic one), it determines the typology of the actions and therefore the nature and scope of the communication. Multisensory experience (visual, tactile and sound resources) will give greater scope to the interface, while extending its use to more users.

The construction of the interface involves considering such diverse fields as those touching users' perceptive, psychological and cultural peculiarities, the technical limiting factors inherent in the nature of the interface and the establishment of a structured taxonomy of the tasks it will enable. Thus, Rheingold maintains that the interface should be transparent, that interaction should not take place on the programme, but on the task itself [34]. Bonsiepe coincides with Rheingold and adds that the interface is nothing more than an intangible tool whose true raison d'être is to help users build a personal mental model to help them understand how the program functions, but without manipulating it directly [33]. Far from being a one-way communication system, an interactive interface -which now integrates both visual and audio signals- proposes a game with users which opens and amplifies their cognitive abilities [35] hence its importance for the subject of this study.

A. *Interface design: related concepts*

There are studies on interface design in the literature dating from the 1970s [36], followed by Apple publications [37], Open Software Foundation [38], IBM [39] and Microsoft [40][41]. All these studies highlight the importance of users as active participants in the design of any communication system, since knowing their modus operandi will enable a more efficient interface system to be generated thereby guaranteeing better accessibility to content.

User acceptance of the interface largely depends on its practical operational ability. This concept known as cognitive compatibility [42] refers to complex information processing, accepted as a determinant part of human behaviour since the cognitivism emergence in the 1960s.

Bruce Tognazzini [43] proposes 16 principles for interactive graphic interface design, that are based on the design of tasks to help users achieve certain goals. Cooper [44] provides another approach to interaction design, based on prior identification of user goals and designing the simplest route to achieve them in five stages, applying the five interaction design activities defined by Gillian Crampton and Philip Tabor [45]: understanding, abstracting, structuring, representing, and detailing.

Above all these considerations, other authors [46] highlight the importance of matching user's expectations and actual experiences in interactive systems in order to guarantee a pleasant interactive process. These ideas refers to artistic spheres, but can be applied in all fields related to computer interface, because both interactive artwork and HCI aesthetics design 'focus on users' perceptions and psychological states during experience' (p. 526).

In short, we can say space for interface design is in the combination of user knowledge and goals, the tasks to be carried out and the necessary tools for that purpose.

B. *Basic guidelines*

We can state that treatment of the image, text and navigability must ensure formal and semantic unity throughout the interface. A solid organisation of content, tasks, functions and an appropriate navigation architecture, will speed up users' knowledge of the system. The practice of object-action syntax, the introduction of browsing and the adoption of graphic vehicles familiar to users also intervene to accelerate the learning required for control. Thus a space of action for the designer is establish where ordering logic and economy of resources should be directed at promoting communicative efficiency.

VI. Classification of Devices Related to the Accessibility of Heritage

It is clear that many of the devices used in cultural sites don't guarantee a regular level of interactivity for all users. Disability or physical limitations prevent access to culture or understand the contents on equal terms, since these users assume a different perception of the environment, so that the most common interaction mechanisms implemented in the heritage access devices may no longer be effective.

Realizing the importance of this issue, and considering the information previously treated, we have carried out a catalog of all accessibility related devices in the market, used in both cultural spaces as in other social areas. This section discusses a selection of the most common devices used in cultural enclaves. The aim is to establish a classification based on various parameters relating to the level of interactivity studied in previous sections, and thus better understand the scope of each of them, their limitations and appropriateness of their use for the fields of cultural heritage.

For this purpose we have considered four parameters. First, the static or dynamic nature of the device, depending on the response to external stimuli and the ability to provide feedback to the user. Secondly, a classification based on the type of output information: touch, sight, hearing, as well as those designed for solving mobility problems. The third classification refers to passive or active nature of the device: input device, if it is able to receive new information for treatment and subsequent response, or output device, if only sends information to the user. Finally, a classification based on the level of interactivity within the parameters discussed in the literature review: according to the actors involved, according to the relations between messages, and according to the control over the writing and content.

The results show that about half of the main devices used in cultural heritage sites present a dynamic behavior, that is, they have an output reaction caused by an input of the user. Many of these devices are designed to facilitate mobility, or require visual perception for use. Similarly, there are very few devices that offer full interactivity with the information they present: in most cases they only offer a reactive

 ISBN: 978-1-61208-117-5

response to user action, but they don't allow to interact with content. Thus, it remains necessary to focus the work of design in this field, in order to create applications that guarantee access to useful content and an appropriate response to a specific requirement.

TABLE I. CLASSIFICATION OF DEVICES RELATED TO THE ACCESSIBILITY

Device	Classification																
	Related to response[a]		***Related to perception***[b]				***Input / output***[c]		***Related to interactivity level***								
									Actors[d]			***Messages***[e]			***Control***[f]		
	r_1	r_2	p_1	p_2	p_3	p_4	*i*	*o*	a_1	a_2	a_3	m_1	m_2	m_3	c_1	c_2	c_3
Pictograph	•			•				•			•	•			•		
Information pannel	•			•				•			•	•			•		
Brochure	•			•				•			•	•			•		
Wheelchair		•				•	•	•	•			-	-	-	-	-	-
Walker	•					•		•	•			-	-	-	-	-	-
Ramp	•					•		•	•			-	-	-	-	-	-
Railing	•					•		•	•			-	-	-	-	-	-
Tactile paving	•					•		•	•			•			-	-	-
Adapted elevator (accessible buttons)		•				•	•	•	•			•			-	-	-
Interactive table		•		•	•		•	•	•					•		•	
Adapted counter	•			•		•		•	•			-	-	-	-	-	-
Tactile model	•		•					•	•			-	-	-	-	-	-
Tactile image	•		•					•	•			-	-	-	-	-	-
Audio description		•			•			•	•			•			•		
Audio guide		•			•			•	•			•				•	
Video guide		•		•				•	•			•				•	
Digital talking books		•			•		•	•	•					•		•	
Anti-obstacles lenses		•			•		•	•	•				•		•		
Ergonomic keyboard		•	•	•		•	•		•			•			-	-	-
Virtual keyboard		•		•		•	•		•			•			-	-	-
Braille keyboard		•	•				•		•			•			-	-	-
Signaling hardware		•	•			•	•		•				•		•		
Scanner: audio output		•			•		•	•	•				•		•		

a. r_1 estatic, r_2 dynamic; b. p_1: tactile, p_2: visual, p_3: audio, p_4: mobility; c. i: input, o: output; d. a_1: user-device, a_2: user-user, a_3: user-message; e. m_1: non interactive, m_2: reactive, m_3: interactive; f. c_1: reactive, c_2: coactive, c_3: proactive

DISCUSSION

The quality of being interactive is frequently used to describe very different products and in some cases, in a contradictory way. If the term is overused, improperly used, or without reference to usability features, it can result in an initial sensation of rejection of the product. So, there is an obvious need to clarify what we mean by interaction as a first step towards specifically assessing this characteristic with quality criteria in different types of objects or interfaces.

This paper has attempted to offer a perspective on the breadth of the concept of interactivity through an extensive literature review and analysis of some of the areas of study closely linked to the concept.

In the last decade it can be seen that most studies related to interactivity focus on technological resources mainly in relation to the use of computers and graphic interfaces for digital applications where the visual and sequential organisation of the perceptive scene is particularly important. However, in order to highlight the role of people in this process it has been considered important to recover classifications which contemplate the set of participants in the process: people, messages and message structure. According to these agents three levels of interaction are established, which could guide methodical and empirical analysis of products and devices already implemented. Any device that focus on disabilities can be studied from this point of view, examining their potential for interaction.

The review of other concepts and research related to interactive experience enables observation of the size and interest of this topic of study for architects and designers and the importance of aesthetic components as an inescapable part of objects and places.

With regard to the design of user interfaces for computers, interaction aesthetics is a determinant element in the nature of the dialogue between user and information, opening a broad field for artistic practice based on multimedia experiences or providing the user with more efficient access to information.

In the literature reviewed, we found there are numerous interaction studies based solely on visual attributes, ignoring elements that could enhance the interaction or lead to the development of new uses for interfaces. The wide expressive range of the movements of the human body, for example, can be used to establish a more dynamic relationship with the devices: speech, hands and finger movements to point or to drag objects, can certainly promote more intuitive contact with the contents and should be taken into consideration in any attempt to improve accessibility.

All these approaches link multisensory perception, aesthetics and interaction as a substantial component in all kinds of products, places and services directed at people.

This idea must be kept in mind when designing for accessibility or for disabled people.

We conclude this paper by stating that some major challenges for interactive product design in the XXI century include studying the potential of interactive images, incorporating the tactile qualities of materials, using the expressive and movement qualities of the human body, and giving the user the ability to discriminate between information, making content more pleasant and accessible.

REFERENCES

[1] T. Maldonado and G. Bonsiepe, "Science and Design", *Ulm 10/11, Journal of the Ulm School for Design*, pp. 8-9, 1964.

[2] D. A. Norman, The Design of Everyday Things. Basic Books: New York, 2002.

[3] N. Shedroff, N., Experience Design 1. Indianapolis, Indiana: New Riders, 2001, p. 142.

[4] S. Rafaeli, "Interactivity: From new media to communication" in Sage Annual Review of Communication Research: Advancing Communication Science: Merging Mass and Interpersonal Processes, vol. 16,1988, pp. 110-134.

[5] H. Dubberly, U. Haque and P. Pangaro, P., What is interaction? Are there different types? Retrieved February 2, 2010, from http://www.dubberly.com/wp-content/uploads/2009/01/ddo_article_whatisinteraction.pdf.

[6] Y. Liu and L. J. Shrum, "What Is Interactivity and Is It Always Such a Good Thing? Implications of Definition, Person and Situation for the Influence of Interactivity on Advertising Effectiveness", Journal of Advertising, vol.31, n.4, 2002, 53-64.

[7] G. Bettetini, "Tecnología y comunicación", in Las nuevas tecnologías de la comunicación, G. Bettetini and F. Colombo, pp. 15-38. Barcelona: Instrumentos Paidós, 1995

[8] D. M. Rhodes and J. W. Azbell, Designing interactive video instruction professionally", Training and Development Journal, vol. 39, n.12, 1985, pp. 31-33.

[9] C-H. Cho and J. D. Leckenby, "Internet-Related Programming Technology and Advertising", Proc. 1997 Conference of the American Academy of Advertising, M. C. Macklin, Ed. Gainesville, FL: American Academy of Advertising, 1997, pp. 162-179.

[10] P. Barker, "Designing Interactive Learning", in Design and Production of Multimedia and Simulation-based Learning Material, T. De Jong and L. Sarti, Eds. Dordrecht: Kluwer Academic Publishers, 1994.

[11] International Standard, ISO 9241-11:1998(E). Retrieved July 13, 2009, from http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/ISO9241part11.pdf

[12] K. Lynch, The Image of the City. Cambridge: The MIT Press, 1960.

[13] R. Passini, Wayfinding in Architecture. New York: Van Nostrand Reinhold, 1984.

[14] P. Arthur and R. Passini, Wayfinding: People, Signs, and Architecture. New York: McGraw-Hill Book Companies, 1992.

[15] J. Weisman, "Evaluating Architectural Legibility: Way-Finding in the Built Environment", Environment and Behavior, vol. 13, 1981, pp. 189-204.

[16] D. Lowenthal and M. Riel, "The nature of perceived and imagined environments", Environment and Behavior, vol.4, 1972, pp. 189-207.

[17] L. M. Ward and J. A. Russell, "Environmental Psychology", Annual Review of Psychology, vol. 33, 1982, pp. 651-689.

[18] A. Stevens and P. Coupe, "Distortions in Judged Spatial Relations", Cognitive Psychology, n.10, 1978, 422-437.

[19] P. W. Thorndyke and S. E. Goldin, "Spatial Learning and Reasoning Skill", in Spatial Orientation: Theory, Research, and Application, H. L. Pick and L. P. Acredolo, Eds. New York: Plenum Press, 1983, pp. 195-217.

[20] Y. Wang, "Taxonomy of Wayfinding Experiences", Design Principles And Practices: An International Journal, vol. 1, n. 2, 2007, pp. 83-91.

[21] D. Quarante, Diseño Industrial, Elementos Introductorios. Barcelona: Ceac, 1992.

[22] A. G. Greenwald, "A social-cognitive account of the self's development". In Self, ego, and identity: Integrative approaches, D. K. Lapsley and F. C. Power, Eds. New York: Springer-Verlag, 1988, pp. 30-42.

[23] L. Tiger, The pursuit of pleasure. London: Transaction Publishers, 1992.

[24] P. W. Jordan, P. W., Designing pleasurable products: An introduction to the new human factors. London: Taylor & Francis, 2000.

[25] Seligman, M. (2002). Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment. New York: Simon & Schuster.

[26] J. Pallasmaa, J. Los ojos de la piel. Barcelona: Gustavo Gili, 2006.

[27] A. Antonovsky. Unrevealing the mystery of health – how people manage stress and stay well. London: Josey Bass, 1987.

[28] S. Ilstedt-Hjelm, Making sense. Design for well-being. PhD thesis. Stockholm: NADA, Royal Institute of Technology, 2004.

[29] W. C. Chang and T. Y. Wu, "Exploring types and characteristics of product forms", International Journal of Design, vol.1, n..1, 2007, pp. 3-14.

[30] P. Desmet, J. C. Ortiz and J. Schoormans, "Product personality in physical Interaction", Design Studies, vol.29, n.5, 208, pp. 458-477.

[31] T. Mandel, The element of user interface design. New York: Wiley, 1997.

[32] B. Laurel, "What is Interface?", in The Art of Human-Computer Interface Design, B. Laurel, Ed. New York: Addison Wesley, 1990.

[33] G. Bonsiepe, Del objeto a la interfase. Buenos Aires: Infinito, 1999.

[34] H. Rheingold, "An interview with Don Norman" in The Art of Human-Computer Interface Design, B. Laurel, Ed. New York: Addison Wesley, 1990.

[35] T. Leary, "The interpersonal, interactive interdimensional interface", in The Art of Human-Computer Interface Design, B. Laurel, Ed. New York: Addison Wesley, 1990.

[36] W. J. Hansen, "User Engineering Principles for Interactive Systems", in Fall Joint Conference Proceedings, vol. 39 Moltvale: AFIPS Press, 1971, pp. 523-532.

[37] Apple Computer, Inc., Apple Human Interface Guidelines: The Apple Desktop Interface. Massachusetts: Addison-Wesley, 1987.

[38] Open Software Foundation, OSF/Motif style guide. New Jersey: Prentice Hall, 1992.

[39] IBM, Object-Oriented Interface Design: IBM Common User Access Guidelines. Carmel: Que, 1993.

[40] Microsoft Corporation, The GUI Guide: International Terminology for the Windows Interface. Redmond: Microsoft Press, 1993.

[41] Microsoft Corporation, The Windows Interface Guidelines for Software Design: An Application Design Guide. Redmond: Microsoft Press, 1995.

[42] J. Nielsen, Usability engineering. Boston: AP Professional, 1993.

[43] B. Tognazzini, "First Principles of Interaction Design", 2003. Retrieved February 2, 2010, from http://www.asktog.com/basics/firstPrinciples.html.

[44] A. Cooper, R. Reimann and D. Cronin, About Face 3: The Essentials of Interaction Design. Indianapolis: Wiley, 2007.

[45] G. Crampton and P. Tabor, "The Role of the Artist Designer", in Bringing Design to Software, T. Winograd, Ed. New York: Addison-Wesley, 1986.

[46] Z. Bilda, E. Edmonds and L. Candy, "Designing for creative engagement", Design Studies, vol.29, n.6, 2008, pp. 525-540

Digital Reconstruction of a Historical and Cultural Site Using AR Window

Jiyoung Kang
Graduate School of Culture Technology
KAIST
Seoul, Korea
e-mail: lov2jung@gmail.com

Jung-Hee Ryu
Graduate School of Culture Technology
KAIST
Seoul, Korea
e-mail: ryu@kgsm.kaist.ac.kr

***Abstract*— There are many historical and cultural sites which have been demolished due to the reckless urban development projects. 'Pimatgol' is one of those with meaningful history and culture. There were many attempts to reconstruct the sites using technology to remember the important places. AR Window is an Augmented Reality application for reconstructing 'Pimatgol' for smart phones. User can see the past scenes over the present scenes through AR Window. User can attempt multiple interactions like blowing a breath, wiping with a finger and touching buttons on a phone. We concentrated on how the multiple interactions affect interests and level of awareness of the users. AR Window will open a new path of digital reconstruction for smart phones through the multiple interactions.**

Keywords- Case studies; Multiple interaction; Emotional effect; AR

I. INTRODUCTION

Due to the reckless development project in many cities in Korea, historical and cultural sites have been destroyed. 'Pimatgol' with over 600 years of history is one of the historical and cultural sites in Seoul that have embraced sorrow and happiness of the people throughout the time. It has been a very meaningful and precious place to our community. The name 'Pimatgol' is derived from the word 'Pima' which means avoiding horses. The alleys from behind the Kyobo Building in Gwanghwamun to the YMCA Hotel in Jongno that make up Pimatgol have their roots in the Joseon Dynasty (1392-1910). Pimatgol is thought to date back to the Joseon Dynasty when commoners and lower-class government officials had to kneel before passing high-level officials. The alleyway is so narrow that a person can scrape his/her hands along the walls on each side. The air is filled with the tangy smell of grilled mackerel and the stale leftover smell of makgeolli, a kind of rice wine. This is 'Pimatgol', the alley that runs east-west alongside Jongno in central Seoul. People can feel they've stepped back into another generation in this place. But the place was forcibly demolished by the City of Seoul as a part of Jongno Redevelopment Plan launched in 2003. Regular customers and the public mourned the store's closing as the end of an era. Many of the general public felt sorry for the habitants of Pimatgol and hoped that the site will be preserved.

There have been many efforts to preserve these kinds of historical and cultural sites. There, especially, have been many projects using Virtual Reality technology for digital recreation of the sites. For example, they measure all the remains from the historic sites and collect all the data to recreate digital images of the past. However, this has some problems like lacking interactions, necessary equipments or lack of reality. Due to these, we need to find another way to recreate our historical and cultural sites.

As members of the part who feels sorry for the disappearance of the historical and cultural site of 'Pimatgol', we decided to recreate the place using Augmented Reality. AR [1] is one of the most perspective technologies that could combine real and virtual world to show the combination of ancient and current view of the historical site [2]. Compared to the ordinary instructions like static images, videos, sand table etc., AR can provide user with an intuitive and visual experience by digitally recreating the ruins. User can compare past scenes over present scenes immediately by using AR for digital recreation.

(A) (B)

Figure 1. (A) Past photo of one street of 'Pimatgol' in 2007,

(B)Photograph of dismantlement of the same street of 'Pimatgol' in 2010

There are many iPhone applications using AR. For example, there are 'ScanSearch' [3], 'Layar' [4], 'Sekai Camera' [5] etc. Users usually scan, touch or move around the smart phones. Those are typical interactions for AR applications which are now too simple and common. Through this paper, we suggest AR Window the novel digital recreation application for smart phone which has implicit and interesting interactions that will raise people's emotional responses.

II. RELATED WORKS

There have been many works in the field of digital recreations. Virtual reality and augmented reality techniques, especially, are the most progressive and leading methods of digital recreations. Here are few cases of digital recreations using virtual or augmented reality.

A. Hwang-ryong Temple

Hwang-ryong Temple was the most representative temple which was built by the Shin-ra Dynasty in Kyong-ju, Korea. It was built by selected master craftsmen and became, very famous for its beauty and magnificence. However, it was burnt down 800 years ago and there, only remain a site for the original temple. To recover the original temple's beauty, Graduate School of Culture Technology (GSCT), KAIST produced a digital 3-dimentional video of the original Hwang-ryong Temple. They collected all the paper and other materials from the past for recreating the temple. Digital recreation of the temple was evaluated as a very accurate and immersive work due to the effort. However, being as it is a digital video; user can only just experience one way interaction of watching the video.

B. Sungnyemun in Second Life

Sungnyemun is the Number One National Treasure of Korea. It was destroyed by fire in 2008. Every Koreans felt sorry for the disaster. A developer company called 'Acid Crabiz' [6] recreated Sungnyemun in 3-dimentional virtual reality game 'Second Life' as an effort to recreate Sungnyemun. It was built accurately following the same blueprint of the original building, so user can see the realistic and immersive recreation of Sungnyemun. Many people from all around the world can visit and interact with the Sungnyemun in the virtual world. Users have to go into the virtual world 'Second Life' to visit the building. Moreover, it only exists in the virtual world so there is no reality.

C. AR-View: an Augmented Reality Device for Digital Recreation of Yuangmingyuan

Yuanmingyuan was a vast and magnificent royal garden built by several emperors over a period of time during the Qing Dynasty. It was burnt down by the Anglo-French forces in 1860 and suffered continual damages later on. Beijing Institute of Technology digitally recreated Yuanmingyuan by designing and manufacturing a fixed-position device AR-View to provide a combined real and virtual image of Dashuifa (Great waterworks), a symbol of Yuanmingyuan. With the help of AR-view, the original exquisite architectures and fountains are superimposed upon the current ruins. However, user needs special equipment for AR-View to see the beautiful garden. User' sight is also very limited because the AR-View is fixed in one place that cannot be moved around.

III. OUR WORK

We created AR Window to recreate the historical and cultural site of 'Pimatgol'. AR Window is an application for iPhone. AR Window is using AR for combining past scenes over present ruins of 'Pimatgol'. AR Window provides user with immersive, intuitive and entertaining experience.

A. System Concept and Design

There were many cases of digital recreations, but none concentrated on responses from the user. Maybe it is because they surveyed the satisfaction from the user after experiencing the recreations. However, we concentrated on how this case affects user's emotion. We wanted to raise the level of attention and affection to our historical and cultural site of 'Pimatgol' from the users. Interactive input interactions were given to users to give implicit meaning. The interaction is called 'window wiping' comprising of actions of wiping the AR Window. This is a metaphoric interaction of wiping the actual window to see the outside well. Novel interactions include the blow-and-wipe transition between the current camera views. A user blows a breath to the scene shown at 'Pimatgol' where the user is now on a smart phone. The AR Window will bring the "fogged-up" view of historical photos of the site with the blow. Then the user wipes the blurry view with his/her fingers and the past photos become clearer. Finally, the user can see the past scenes of the site clearly. User needs to perform 'window wiping' interactions to see the past photos of present ruins. Norman [7] proposed the theory of seven stages of action.

- Forming the goal
- Forming the intention
- Specifying an action
- Executing the action
- Perceiving the state of the world
- Interpreting the state of the world
- Evaluating the outcome

It is important to notice that Norman's theory of action focuses on user cognition. User can get used to the idea of blowing and wiping easily because they already know the why. It is called Expressive Representation(ER) according to Ulmer [8]. We 'read' and interpret representations, act on, modify, and create them in interactions. In restricted or mobile circumstances like in this case, it integrates complementary modalities, by having multiple interactions like seeing, blowing and wiping, [9], in a manner that brings a synergistic blend in such as way that each mode can be capitalized upon and used to overcome weaknesses in other modes.

ISBN: 978-1-61208-117-5

Figure 2. Exact address of the marked spot

Through this implicit input, user can get explicit outcomes that are photos from the past. Those photos are from blogs of general public. They also feel sorry for the disappearance of 'Pimatgol' or having their personal memories for the place. We can find the fact from their texts and photos from blogs of users. We manually selected photos from the community sites like 'Naver', 'Nate' and 'Daum' which form over 70 percent of Korea blogs. Due to this, user can interact and share thoughts through the photos. User can get information awareness from their sites which have text and photos of the place.

Using augmented reality for AR Window application, user can get context awareness in outdoor environment and location awareness (GPS) like in 'Human pacman' [10], because user can see the exact addresses and a map where they are seeing. They can see the outside environments over the current photo of the smart phone.

B. *System Description*

It has three menus, the first is a map view, the second is a camera view and the last is the list of the places. There are two methods of using 'AR Window'. User can visit the place and connect to the application with the first method. User will select a camera view and they will see the markers by scanning the site with their phones. The markers contain the past photos of the marked spots. By touching the markers, the user can start the multiple interactions with the application. It will ask the user to blow on the window to see the past scenes of the spot. When the user performs the blowing action, the screen on the phone will show the past photos taken by other community members. The blurry scenes will clear by wiping the screen with their fingers like wiping the window. The clear past photos overlay the present ruins or newly constructed buildings. User can enjoy more past photos of the place taken from 2004 to 2010 by repeating these interactions. For the second way, user can explore the site without going there. User just stays where they are, and they can use the application by using the 'Map View' of the application. In the map view, user can see marked places where they can see the past photos on the Google map. User can see past photos of the markers by touching the markers.

In addition to that, if the users want to know about the place more, they can visit the source blogs where the photos came from by touching the 'visit blog' button on the menu. They can read and see more information and share thoughts about the place on the blogs. Through this, users can get information awareness.

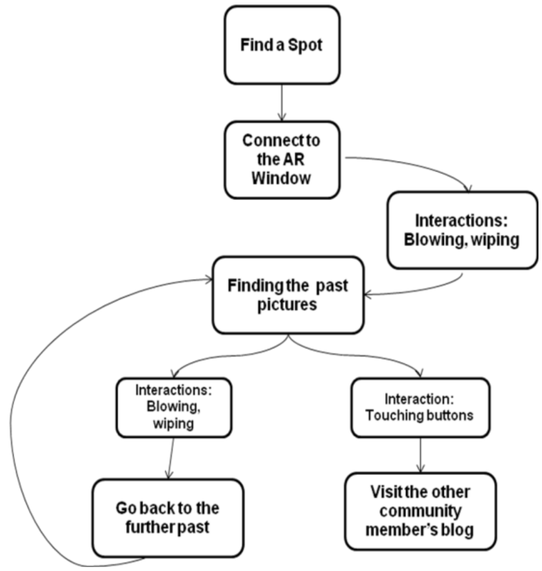
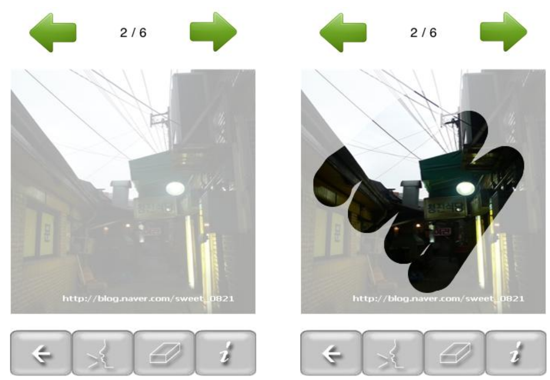


Figure 3. System flowchart



(A) (B)

Figure 4. (A) Blurry photo by blowing a breath, (B) A photo wiped with a hand

1) *Spot Selection*

There are 20 spots in 'Pimatgol'. Selected spots were mostly popular restaurants in 'Pimatgol', because they were well known street for restaurants for general public. Most of them have over 50 years of history, but some of them moved to a new building and others disappeared.

2) *Photo Selection*

We selected photos of this system manually. There are about 6 pictures in average for one spot. We put the word 'Pimatgol' in web search engines like 'Naver' and 'Daum' which are the top two in Korea for searching for photos. And we only picked photos from personal blogs where the community members expressed their thoughts and emotions. We got necessary permits from the bloggers for using the pictures. The blogs only have photos from the year 2004, so we picked the best photos that show the exterior well. Also, the blogs that we picked include their information, memories and thoughts about 'Pimatgol' which the other user can view and share their thoughts.

IV. TECHNICAL DESCRIPTION

We collect all the location information including latitude and longitude from 'Google maps' for places that we picked for reconstruct based on their addresses. The two

(A) (B)

Figure 6. (A) Map View, (B) Camera View

different views in the 'AR Window' are map view and camera view.

Map view is based on the Google map and the camera view shows the current views with markers indicating the presence of photos of the past. SDK 'Mapkit', a basic framework for i-phone, was used for the map view. Pin markers were assigned to the spots on the location information. For the camera view, we used an open library 'ARKit'. Location information was assigned as in the map view, and UIButton was designed. The size and direction of UIButton is updated in the camera view according to the user's current location and direction.

For the blowing action, the microphone checks the decibel of the user's breath. When the decibel is over -5DB which means light breath, the screen will bring blurry past photos. Then, by wiping the blurry photo with fingers masking technique will make the photos clear.

V. USING THE TEMPLATE

Testers were selected randomly from near 'Pimatgol'. They were divided into two groups to compare the effects of different interactions. Each group consisted of thirty different users. First group (A) did multiple interactions like blowing a breath, wiping with a hand and touching buttons on a mobile phone. The second group (B) just touched the buttons on a mobile phone. There was no time limit, so that users can experience the AR Window as much as they wanted. We concentrated on how the multiple interactions affect interests and level of awareness of the users.

A. User Interest

We measured how much time they spend to explore the application, how many spots they visited, how many blogs they visited to get more information about the place and we measured the degree of testers' interests by measuring how many places they remembered.

The first result for performing time showed that the average time for the first group (A) was 175 seconds and the second group spent 91 seconds. Even though, multiple interactions originally take more time to perform, 84 seconds of difference means that the users took much more time for exploring the past 'Pimatgol' through the implicit interactions.

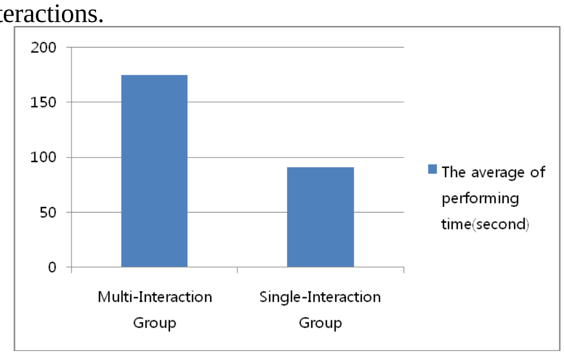


Figure 7. Result of Performing Time

The results for visiting places, the first group (A) was a little bit higher than the second group (B). Group A visited 4.8 places in average. Group B visited 4.4 places in average. The result showed that multiple interactions affected people to visit more places.

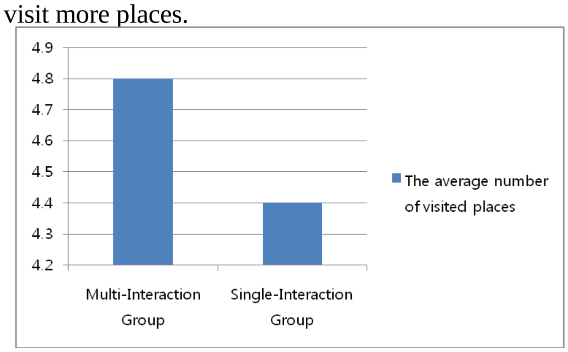


Figure 8. Result of visiting places

For the results for visiting blogs, the first group (A) visited 2.3 blogs in average and the second group (B) group visited 1.7 blogs in average. A group who experienced various interactions had more interests on the place, so they visited the blogs where the photos came from.

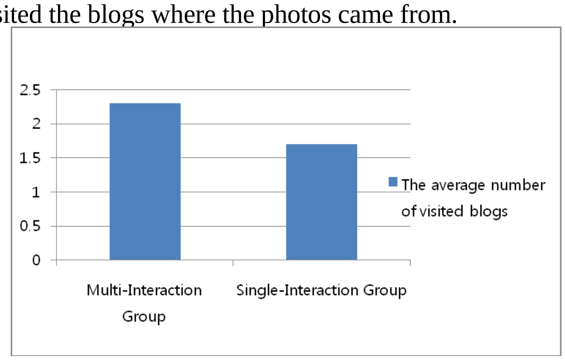


Figure 9. Result of visiting blogs

 ISBN: 978-1-61208-117-5

The number of places remembered by the first group (A) was 3.7 places in average and the second group (B) remembered only 2.1 places in average. This showed that testers can remember more places through various implicit interactions than simple touching. While they experienced various interactions to see the photos, they were more affected by the places.

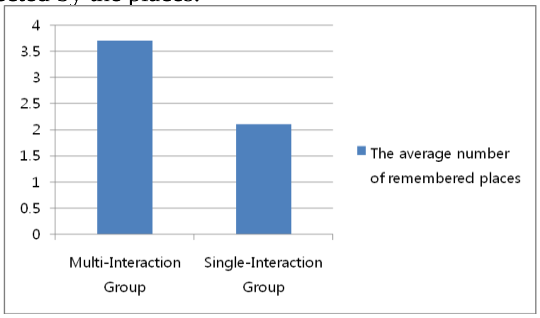


Figure 10. Result of remembered places

All the results from measuring interests showed that more than average experienced multiple interactions from the first group. These results showed that user who performed various metaphoric interactions like blowing a breath, wiping the window with their fingers were more interested on the place. The more involved interaction seems to have had much many positive effects on user interest.

B. *Level of Awareness*

Two questions were asked to two groups to compare the changed awareness on 'Pimatgol'.

(1) How much money will you contribute for preservation of the 'Pimatgol'?

(2) If you were the mayor of Seoul, how much do you agree to preserve the 'Pimatgol'? (in percent)

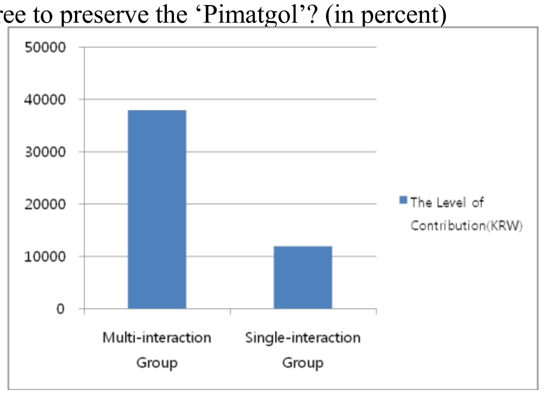


Figure 11. Result of contribution

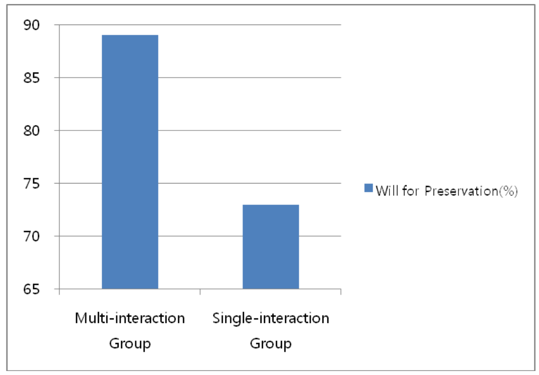


Figure 12. Result of preservation

From the result, we can see that the group who experienced multiple interactions was willing to contribute more money than the simple touch group. The first group was willing to contribute 38,000 Won for preserving the 'Pimatgol' but the second group was only willing to donate 12,000 Won which was significantly less than the first group. Also 89 percent of the multiple interactions group were willing to preserve the 'Pimatgol', while 73 percent of the single interaction group agreed to preserve. It means that metaphoric multiple interactions raised the level of awareness on 'Pimatgol' more besides the user interest. The metaphoric interactions reinforced people to remind the importance of historical and cultural site of 'Pimatgol', so it made the testers to contribute more to preserve the place.

VI. CONCLUSION

We have presented the design of an AR application for digital reconstruction of 'Pimatgol'. Experiments in the site showed that multiple and implicit interactions of 'window wiping' gave more entertainment and emotional effects on the user compared to simple operation of buttons. It also raised more awareness on the historical and cultural site by blowing a breath and wiping the screen with a finger like window wiping. Also, user can see the past scenes over the present ruins or newly constructed through past photos of the spot from other community member's blogs. Also user became more aware of the location through the map which showed where they were. In addition to that, user can get context awareness through their outdoor environment. Lastly, they got information awareness from texts and photos from the blogs of other community member on the place. Although the AR Window application should be improved further, especially by adding more sites and photos; it opened a novel way of digital reconstruction by raising the level of attention and affection to our historical and cultural site.

REFERENCES

[1] Azuma R., Baillot, Y., Behringer R., Feiner S., Julier S., and MacIntyre B., Recent advances in augmented reality, In: Computer Graphics and Applications, IEEE Volume: 21 page(s): (pp.34-47), 2001.

[2] Fritz, F., Susperregui, A., and Linaza, M.T., "Enhancing Cultural Tourism experiences with Augmented Reality Technologies", In :

 ISBN: 978-1-61208-117-5

The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST , 2005.

[3] ScanSearch Website, http://www.scan-search.com/, May 2010.

[4] Layar Website, http://www.layar.com/ May 2010.

[5] Sekai Camera Website, http://www.sekaicamera.com/ May 2010.

[6] Acid Crebiz Website, http://www.acidcre.biz/ May 2010.

[7] Norman, D. A., The Design of Everyday Things. Doubleday: The Design of Everyday Things. Doubleday: (pp.46-48) New York, New York, 1990.

[8] Ullmer B. and Ishii H., “Emerging frameworks for tangible user interfaces”. IBM Systems Journal 39(3- 4), (pp.915-931), 2000.

[9] Oviatt, S. L., “Multimodal Interface Research: A Science Without Borders,” In B. Yuan Huang & X.Tang (Eds.), Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000), Vol. 3, (pp. 1-6). Beijing , 2000.

[10] Cheok, A., Goh, K., Farbiz, F., Fong, S., Teo, S., Li,Y., and Yang, X., Human Pacman: A mobile, wide-area entertainment system based on physical, social *and ubiquitous computing.* Personal And Ubiquitous Computing 8, 2. , Springer-Verlag, (pp. 71–81), May 2004.

Creating Added Value for Smart Card Applications: The University as a Case Study

Hsiao-Chi Wu
Dept. of Information Management
National Taiwan University
Taipei, Taiwan
hciwu69@gmail.com

Jen Wel Chen
Dept. of Information Management
National Taiwan University
Dept. of Business Administration
Chinese Culture University
Taipei, Taiwan
d95725009@ntu.edu.tw

Ching-Cha Hsieh
Dept. of Information Management
National Taiwan University
Taipei, Taiwan
cchsieh@im.ntu.edu.tw

***Abstract*—Studies on smart card applications, in addition to focusing on instrumentality and practicality, should also consider the importance of social construction. The implications of actors with different roles on technology directly influence the developmental direction of smart card applications. This study examines smart card applications, using a private university in Taiwan as a case study. It attempts to understand and interpret how the university handles problems arising from smart card applications, from the perspective of schools and suppliers, during the planning and design phases of implementation. Additionally, it uses the Social Construction of Technology (SCOT) as an analytic lens. The results of this study show that when the university developed a new interpretation of smart card applications, it focused on value-added e-services. This emerged and unexpected consequences arose from actual work-practice situations; smart student ID cards were not merely tools for identification and payment, but also the means to shape student lifestyles.**

Keywords - smart card application, SCOT, e-service, university

I. Introduction

Smart card technology is widely applied within our daily lives, from electronic payments, transportation, and telecommunications to healthcare, entertainment, and education [1, 3]. Smart cards are capable of accessing, storing, and calculating data, as well as providing an immediate exchange of necessary information to facilitate data security, identification, and authentication.

This study discusses the application of smart cards within a university setting. Existing studies on smart card applications in schools [5, 6] and students' technological acceptance of smart cards [1, 7] focus primarily on users and usage. Villano [8] claims that between 1997-2002, there were nearly 50 schools implementing smart card programs in the U.S.; however, only a few actually benefited from their advantages [4, 5].

Holmstrom & Stalder indicated that the drifting of technologies depends on actor interactions. Implementation of technologies do not always follow the original plans [2]. In schools with smart card applications, although students are the primary users, the perspectives of schools, designers, and suppliers should also be considered. During the planning and designing stages of smart card applications, schools and suppliers play particularly important roles. Past research focuses primarily on technological aspects of implementation, especially the instrumentality and practicability of smart cards in schools, without considering the importance of social construction. For example, smart cards could be implemented in a variety of ways, based on the interests and implications of technology on actors with different roles. Some researchers tend to neglect the interpretation of design and configuration of technology, and thus fail to consider certain practical relations [12].

The questions posed by this study are: why do schools issue smart cards? What are the obstacles? How should schools adjust to their implementation and take advantage of their added-value? In answering these questions, this study examines the first case of smart student ID cards at a Taiwanese university. Discovering three different rollouts of student ID cards, it explains how the university treated and coped with the problems arising from the smart card applications. Furthermore, this study uses tools of Social Construction of Technology (SCOT) as an analytic lens.

The paper is organized as follows. Section II reviews the literature related to smart cards in university applications. Section III describes the data collection strategies and introduces the analytic lens -- SCOT. Section IV explains case background. Section Ⅴ analyzes the case and discussion follows in section Ⅵ, and then conclusions are offered.

II. Smart Card in University Applications

Smart cards are commonly applied within universities [1, 5, 6], which often provide the best opportunity for promoting smart card technology. For universities, smart cards can enhance administrative efficiency [1], reduce administration costs, and increase incomes [9]. Mirza & Alghathbar [5] and Flwler, Swatman & Welikala [6] investigated smart card applications in universities around the world, including North America, West Europe, and Asia. They surveyed 20 universities with 34 applications, and developed four categories for the most popular applications: student ID cards,

 ISBN: 978-1-61208-117-5

book borrowing, and access to library and printers [5]. Related applications can be generalized into three categories: (1) student ID cards as identification; (2) entrance cards for libraries, buildings, parking lots, etc.; and (3) e-purses for consumption (e-purse [3] is a prepaid card which can be used for payment or small retail instead of coins).

Schools adopt different types of smart cards, such as magnetic stripe cards, contact-less cards, or hybrid-cards, each with different versatilities [6]. Services also vary, as some function as debit cards, have store-value, or other payment types [5]. Clark [9] discussed the key successful factors of e-purse smart card applications, and concluded that cost, interoperability (applications outside of schools to replace coins, or to be integrated with other cards), and a critical mass of users and merchants are most important.

There are many examples of both successful and failed micropayment applications with university smart cards. Mcard of the University of Michigan was launched in 1995, but suffered from poor promotion and was suspended in 2001. Smart cards at the University of Central Florida, first issued in 1998, provide valuable discounts and benefits to students on campus [9].

In some cases, schools use more than one smart card system. The library of the City University of Hong Kong introduced the Octopus card (one of the most successful debit cards in the world [2]) as a payment mechanism for copying, laser printing, and overdue fines, rather than use the existing student CitySmart card. On one hand, the CitySmart system was plagued by a non user-friendly interface and limited numbers of debit devices on campus. And on the other hand, over 98% of the students at the City University of Hong Kong possess the Octopus card [4].

Beyond the focus on the university perspective, other studies [1, 7] examine this issue in terms of student acceptance of technology. In sum, the promotion and development of smart card applications in universities are not simple technical issues, but rather involves social and economic factors, as well as users' behaviors and preferences.

Although these studies have provided an understanding of smart card applications in universities, they have neglected the purpose of schools implementing such cards. Purpose may influence both the functional design of smart cards, as well as the appearance of potential problems and corresponding solutions during implementation [4]. In fact, smart card application has different meanings to different users and usage stages. Different visions for smart card application by schools influence the developmental direction of these cards on campus.

III. RESEARCH METHOD

A. Strategy of data collection

This study focuses on one case study, allowing the researcher to connect research phenomenon with actual situations [11], in order to recognize the dynamics of these phenomenon. The main methods for data collection include participant observation, semi-structural interviews, conference records, and file data (see Table I). One of the authors works at SCE, he has completely participated in the development of smart card applications. We conducted interviews and collected secondary data from March 2006 to May 2010. Through these multiple sources of data collection, the authenticity of the data was repeatedly validated.

TABLE I. THE TYPES AND ILLUSTRATIONS OF EMPIRICAL DATA

Data type	Illustrations
Participation observation	Observation period: 1998.5-2010.5. One of the authors works at SCE.
Semi-structure interview	There are ten person-interviews. Each interview lasts 90 to 120 minutes. Some of the interviewee is interview twice depends on situation.
Meeting minutes	There are 32 minutes, such as cooperation and negotiation of enterprises, technology discussion meeting, managers' meeting, and technology group meeting.
Documents	There are 54 files like project reports, technical documents, memorandums, official documents, presentations, and historical data.

B. Analytic lens—Social Construction of Technology

This research employs SCOT as the analytic lens for understanding how the interpretation of smart card applications by schools and suppliers influences the developmental direction of these cards during planning and design, and how they deal with related problems.

SCOT suggests that technology is a social structure that allows researchers to analyze technological artifacts through social situations. The developmental process is the selection of changes and eliminations. It is based on a multi-directional model, rather than a linear model. In other words, technology can have more than one developmental result. SCOT opens a "black box" of technology and examines the selection process in order to understand how people consider the problems and solutions of technological devices under different circumstances. The main concepts of SCOT are shown below [10].

- Relevant Social Groups (RSG): relevant social groups are key to understanding that technology is a social product. Each social group has a different interpretation on technological artifacts that results in different problems and solutions. For instance, young people may view bicycle riding as exercise, while mothers and older people only focus on riding safety. Thus, the development of bicycles should be versatile. The developmental direction of technology is limited to the shared implications among all members; RSG will define the technological problems and solutions.
- Interpretive flexibility: besides considering how RSG view and interpret technology artifacts, interpreting "flexibility" also refers to the design of technology artifacts; there is more than one best design. Interpretive flexibility helps to explain how different RSG treat and construct the problems and solutions arising from technology in order to expose the "black box" of technology.

- Closure & stabilization: with regard to technology, closure includes artifact stabilization and problem-solving. It refers to whether RSG are certain about problems being solved or whether main problems have been redefined by RSG (indicating that RSG have given a meaning to the solution of a technology artifact). For instance, use of pneumatic tires on bicycles was originally intended for solving problems with vibration, but later contributed to faster bicycle speeds. Higher wheels are used on high-speed bicycles, but are less safe for women or the elderly. For different groups, the development process of bicycles involves different dimensions, and their chosen artifacts vary. Thus there are degrees of artifact stabilization. Relevant social groups will form the regulation and values of artifacts, and further influence the meaning and development direction of artifacts.

IV. CASE BACKGROUND

This study examines a private university in Taiwan (anonymous: CU) as the subject, and ascertains how the IT department developed smart cards to provide on-campus services and position the university as a pioneer of information technology innovation in Taiwan. Since 1995, CU has actively pursued the goal of "lifelong education," particularly within the School of Continuing Education (SCE). There are eight teaching centers in SCE, located in the center of the city with convenient transportation. In order to attract students and provide multiple learning services, the dean of SCE set "*e-campus*" as a major goal of SCE organizational development.

CU issued smart student ID cards to enhance campus service efficiency and quality, as well as school resource management. Since SCE operates independently, smart card applications were first implemented by the teaching centers of SCE, under the planning of the SCE IT department. The development process of CU smart student ID cards includes three stages:

A. Issue of smart student ID cards -- UPass

In 1998, CU was the first university in Taiwan to issue smart student ID cards. A local bank (Bank C) issued UPass smart student ID cards with debit card functions and a magnetic strip. In contrast to laminated paper IDs, smart cards allow data access, storage, and calculation, all of which support e-services on the CU campus. Initially, smart student ID cards (UPass) were designed for identification, entrance access to specific locations, the disbursement of fellowship money and student refunds, book borrowing in libraries, and automatic transfers of tuition. Due to financial regulations at the time, UPass could not support multi-functional transactions, but with changes to regulations in October 2001, transaction features were also added.

B. Multi-purpose multi-function of E-service (Campus's application services)

SCE attempted to develop more services for UPass. In 2002, CU was funded by Bank C to change UPass cards from simple magnetic strip cards to RFID cards with magnetic strip and Mifare standards. The SCE IT department recommended that in addition to access control to buildings and parking lots, UPass should also allow access to reserved venues. Teachers and students could directly reserve classrooms or discussion rooms online, and access these places by UPass. Such functions would save on both managerial manpower and administration costs, as well as improve the efficiency of the campus' learning environment.

SCE and entrance access suppliers cooperated and applied RFID to micropayments for photocopy, vending machines, overdue book fines, and online applications for school services, in order to provide multi-purpose multi-functional campus e-services. They also applied RFID readers to lighting, electricity, and air conditioning controls via iBOX.

C. Changes in debit smart student ID cards

In 2005, CU again served as a pioneer and cooperated with the largest transportation card system (EasyCard) in northern Taiwan. CU changed smart student ID cards(UPass) from smart debit cards into RFID smart cards, which serve as store-value cards with contact-less Mifare standards. E-purse was developed for off-line micropayments. The technical specifications of EasyCard were the same as the original smart student ID cards, thus removing potential technical problems with transferring the original services to the new student ID card system. A required pin code was established to ensure the confidentiality of the cards. Therefore, CU expanded the services of UPass beyond the campus to incorporate public transportation. There are twelve compus services of UPass, as shown in Figure 1, which can be classified into three categories, i.e., identification, administration information services, and spatial security.

CU and EasyCard held a press conference on the integration of EasyCard with campus services, and allowed participants to experience the twelve campus services of UPass (see Figure 1). After that, some schools show their great interests, and they asked CU to promote the UPass' integrative campus services for their own schools. The conference enhanced the competitiveness of service innovation at CU, and enhanced the school's reputation. Since April 2010, EasyCard has extended its services to transactions at over 10,000 locations, including four major convenience store chains, coffee shop chains, drug store chains, restaurants, fast food stores, and parking lots.

CU has actively developed innovative applications of the original RFID technique. For instance, in 2008, SCE cooperated with banks, telecommunication companies, and Austrian scholars to apply for a special government technology program, and proposed plans to incorporate mobile phone devices from plastic student ID cards in order to further facilitate e-commerce. The dean of SCE stated:, "*Our technology is ready. In the future, students can use and purchase things outside the campus through mobile phones as student ID cards, entrance access cards, and e-purse, thus creating ubiquitous services.*"

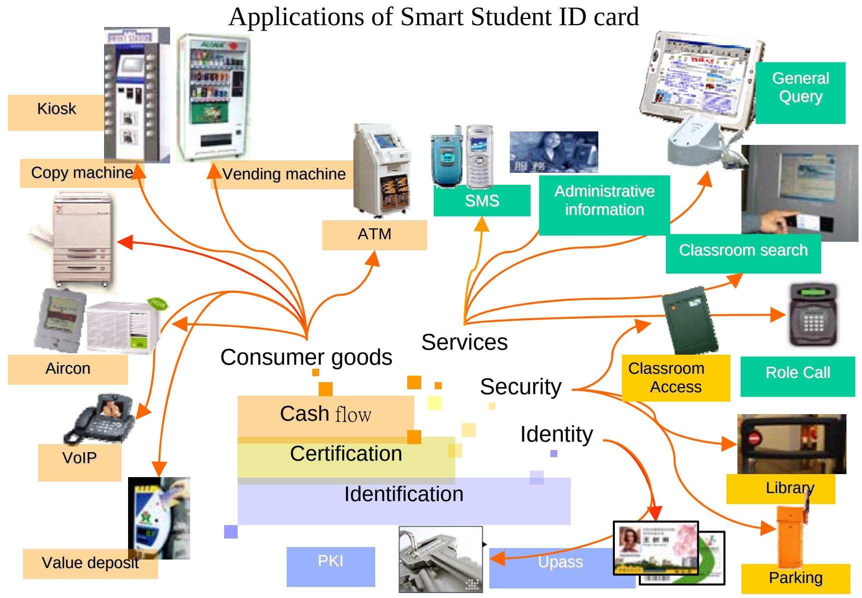


Figure 1. Applications of Smart Student ID Card for CU

V. ANALYSIS

After CU issued the first smart student ID card (UPass) system in 1998, three versions of smart cards have been used, as shown in Table II. This study focuses on the planning and designing stages of smart card applications from the perspectives of schools and suppliers. This section will analyze the development and innovation of CU smart card application using SCOT as the analytic lens.

TABLE II. SMART STUDENT ID CARDS ISSUED BY CU

Stages	Interval	Payment solution	Technical type	Forms of e-purses
1	1998-2002	Debit card	contact	--
2	2002-2005	Debit card	contact contactless (RFID, Mifare)	On-line dollars
3	2005~	Store value	contactless (RFID, Mifare)	On-line dollars Off-line dollars Pre-paid value of EasyCard

A. Relevant Social Groups

In order to realize that technology development is a process of social construction, SCOT prioritizes defining the major relevant social groups within technology development. Groups directly influence technology planning and design. In this case, relevant social groups involved in planning and designing smart card applications are divided into two categories:

- University: the first category includes card holders, including faculty and students; the second involves the IT Department of SCE, which was in charge of planning and implementing smart card applications. Their views on smart card application directly influence development and innovation.
- Suppliers: this includes suppliers who participate in smart card applications, including smart cards issuers such as the bank and EasyCard Company, and suppliers who provide technical support to SCE smart card applications, such as entrance access suppliers.

B. Interpretive flexibility

The views of the university and suppliers on smart card application directly determine planning and design, and influence the direction of development and innovation. Their views are summarized in Table III.

For CU, the issuing of smart student ID cards aimed to realize campus services by an all-in-one card, which provides multi-purpose and multi-function e-services. The IT Department of SCE thus planned and designed smart card applications. However, the IT manager stated: "*the purpose of UPass entrance access is not to control the people coming in and going out; it is about spatial management.*" Thus, e-services incorporating spatial use in teaching buildings were developed, and resulted in the integration of campus information, cash flow, and resources.

Regarding the suppliers of these services (the bank, EasyCard Company, and entrance access suppliers):

- For the bank, it could increase the number of banking customers and provide agency services on tuition, thus earning more income and developing financial transactions on campus.
- For the EasyCard Company, it can utilize EasyCard as a means to connect with campus activities, thus expanding services to fulfill their vision of "*traveling around Taiwan with one card*."

 ISBN: 978-1-61208-117-5

- For entrance access suppliers, it is the opportunity to develop new markets in customized spatial management, in addition to standardized entrance access.

TABLE III. INTERPRETIVE FLEXIBILITY OF RELEVANT SOCIAL GROUPS FOR SMART STUDENT ID CARDS

category	Interpretive Flexibility of RSG
University	SCE: the strategy to develop application services and entrance access is interpreted as spatial management.
	Card holders: convenient; campus pass with one card
Suppliers	Bank: increase income and create new applications for financial services.
	EasyCard: a means to connect with campus activities, and expand the original application service as a transportation card, to multiple services.
	Entrance access suppliers: a way to develop new markets and redefine marketing positioning of the companies.

C. Closure and stabilization

This study aims to understand how key RSG—SCE coped with the problems encountered and redefined them when implementing smart card applications. CU originally expected that UPass could have transaction functions on campus. However, due to the limitations of financial regulations, students could initially only use their cards to pay tuition. Thus, SCE tried to develop other possibilities for smart card applications.

When SCE treated UPass entrance access as spatial management, it redefined the development of smart card applications, seeking the technical support of entrance access suppliers and developing e-services for the spatial use of academic buildings. iBOX is the main engine for entrance access systems that control lighting, electricity, and air conditioning by smart management in order to provide safety services on campus, effectively use campus space, reduce manpower, and avoid wasting resources.

SCE did not give up on the transaction functions of UPass. Since transaction amounts on campus were small and debit smart cards would be more useful, in 2005, they eliminated debit card UPass and chose EasyCard as their new UPass system. When the transaction functions of UPass extended beyond the campus, SCE redefined the ubiquity of its UPass application scope.

VI. DISCUSSION

After three issuances of smart student ID cards, CU adopted different applications and attempted to select the more suitable plan for the university. As mentioned above, SCOT plays a significant role when implementing smart cards. The development of smart cards is not simply interpreted by instrument-oriented technology viewpoint. From SCOT perspective, its social implications can also be examined.

A. Development of emerging and unexpected smart card applications

Technology development and implementation are target-oriented activities. However, when IT design cannot fulfill a goal, there will be unexpected results [15]. SCE encountered difficulties in developing the transaction functions of UPass, and treated UPass entrance access as spatial management to construct the infrastructure of campus services. This outcome was emerged and unexpected given the school's initial plans.

In comparison with the investigation of smart card applications by Mirza & Alghathbar [5] in university settings, UPass' compus services are richer and more widespread (see Figure 1). It is because SCE gives smart card applications new interpretations so that various campus services are created and unexpected results emerge.

The findings of this study verify the situated change model proposed by Orlikowski, who indicated that organizational chage emerges in work-practice situations; it is not a planned or technology-oriented change[13]. Likewise, the development of CU smart student ID cards was not an original goal, and was not led by technology, but rather was an emerging and unexpected result of practical adjustments made by SCE.

B. Creating value-added e-services

Porter suggested that the competitiveness of companies is based on the creation of customer values. One strategy is to develop internal activities of companies through technological development in order to coordinate activities, create value, and realize cost advantages [14].

The findings of this study support Porter's ideas. SCE redefined the meaning of smart card applications as spatial management, cooperated technically with entrance access suppliers, integrated spatial use and managerial activity, and further created value-added e-services, as shown in Figure 2. In addition, for SCE, the planning and design of smart card applications was cost-oriented. They suggested that "*in campus application, managerial costs rather than service costs are high.*" The development of smart card application integrated campus information, cash flows, and resources, thus saving managerial manpower and administration costs for CU. While SCE interprets the function of UPass' entrance access, it means not only to control member's entrance, but also to manage classroom usage, and venue reservations. It enables threefold IT application developments so that SCE further applies the technique of entrance access (e.g., iBOX) to provide safety services on campus and keep creating more value-added e-services.

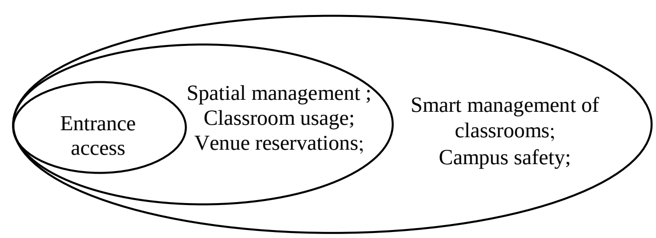


Figure 2 The creation of the value-added e-services from entrance access, spatial management, to campus safety

 ISBN: 978-1-61208-117-5

C. Smart student ID cards: from identification to forms of life

Regarding technological artifacts, Winner suggested that artifacts are not neutral in values, and their design and construction imply specific social intentions. They are treated as forms of life and regulate people's behaviors [12]. Smart student ID cards at CU support this argument [12].

After transforming from paper student ID cards to smart cards, SCE endowed meanings to smart card applications. Besides providing campus services, the smart cards have shaped students' activities, learning, and consumption, both inside and outside of the campus, thus reflecting students' lifestyles. Smart student ID cards are not simply tools for identification and payment, but also could regulate students' disciplines and behaviors. With more services made available by the university and smart card issuers, the regulation of card holders' daily living will continue to expand. Smart ID cards also influenced the work of staff, such as safety control on campus, by helping monitor important systems.

VII. CONCLUSION

Smart cards application in schools is not only for implementing all-in-one card and convenient services. As this case study shows, after the university encountered difficulties in providing transaction services, it redefined the direction of smart card application to spatial management. It further integrated campus information, cash flow, and resources, in order to save managerial manpower and administration costs, thus creating value-added e-services:

- Smart card application is not completely motivated by instrumental characteristics and practicality of technology, rather it is the process of social construction. When relevant social groups give new meanings to smart card application, services emerge from practical situations and result in unexpected outcomes.
- Regarding technology artifacts, smart student ID cards provide campus services, and shape students' lifestyles and behaviors.

Future development strategies of smart card application are as follows. The first is to enhance the mobility of smart card applications, such as technical development of mobile devices and increasing self-service machines on campus for added values and information searches. The second is to develop services upon the activities [17], such as planning daily activities for students and staff on campus; applications for specific locations, such as libraries, gymnasiums, and academic buildings; or socio-spatial dimensions, such as considering urban form, consumer preference, and cultural attributes [18].

Although this study used a university campus as a case study, the findings can also serve as reference to some large corporations that issue smart cards for identification and entrance access. Future studies can discuss how relevant social groups, such as schools (or enterprises) and suppliers, interact with each other to form network relationships by in-depth case analysis in order to contribute to the development of smart card applications.

REFERENCES

[1] C. H. M. Lee, Y. W. Cheng, and A. Depickere, "Comparing smart card adoption in Singapore and Australian universities," International Journal of Human-Computer Studies, vol. 58, no. 3, pp. 307-325, 2003.

[2] J. Holmstrom, and F. Stalder, "Drifting technologies and multi-purpose networks: the case of the Swedish cashcard," Information and organization, vol. 11, no. 3, pp. 187-206, 2001.

[3] J. K. Lee, and C. Y. Yang, "Evolutionary Business Models of E-Cash with Smart Cards," in Proceedings of the International Conference on Electronic Commerce, Seoul, Korea, 2000.

[4] S. Ching, A. Tai, J. Pong et al., "Don't Let Micropayments Penalize You--Experience From The City University Of Hong Kong," The Journal of Academic Librarianship, vol. 35, no. 1, pp. 86-97, 2009.

[5] A. A. Mirza, and K. Alghathbar, "Acceptance and Applications of Smart Cards Technology in University Settings," Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, pp. 746-748, 2009.

[6] D. C. Fowler, P. M. C. Swatman, and J. Welikala, "Issues Affecting the Implementation of Multiple Application Smart Card Systems," "CollECTeR USA" -- 4th CollECTeR Conference on Electronic Commerce, Breckenridge, Colorado, USA, http://www.collecter.org/collUSA/Fowler.pdf., April 11, 2000.

[7] M. Arami, M. Koller, and R. Krimmer, "User Acceptance of Multifunctional Smart Cards," The European Conference on Information Systems, Paper 18, 2004, http://aisel.aisnet.org/ecis2004/18.

[8] M. Villano, "Campus cards: Too smart for their own good," University Business, March 2004, http://universitybusiness.ccsct.com/page.cfm?p=475.

[9] C. Clark, "Shopping Without Cash: The Emergence of the E-purse," Economic Perspectives, Federal Reserve Bank of Chicago, issue Q IV, pp. 34-51, 2005.

[10] T. Pinch, and W. Bijker, "The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might benefit each other," Social studies of Science, vol. 14, no. 3, pp. 399-441, 1984.

[11] R. Yin, Case study research: Sage publications Newbury Park, Calif, 1994.

[12] L. Winner, "Do artifacts have politics?," Daedalus, vol. 109, no. 1, pp. 121-136, 1980.

[13] W. Orlikowski, "Improvising Organizational Transformation Over Time: A Situated Change Perspective". Information Systems Research vol. 7, no. 1, pp. 63-92, 1996.

[14] M. E. Porter, Competitive advantage: creating and sustaining superior performance, The Free Press, 1985.

[15] M. Markus and D. Robey, "Why stuff happens: Explaining the unintended consequences of using IT", The Past and Future of Information Systems, K. V. Andersen and M. T Vendelo ed., Oxford: Elsevier Butterworth-Heinemann, 2004, pp. 61-93.

[16] L. V. Hove, "Electronic purses in Euroland: why do penetration and usage rates differ?," SUERF Studies, M. Balling, ed.: Vienna, Austria: Société Universitaire Européenne de Recherches Financières, 2004.

[17] S. Balasubramanian, R. Peterson, and S. Jarvenpaa, "Exploring the implications of m-commerce for markets and marketing," Journal of the academy of Marketing Science, vol. 30, no. 4, pp. 348-361, 2002.

[18] Y. Aoyama, "Sociospatial dimensions of technology adoption: recent M-commerce and E-commerce developments," Environment and Planning A, vol. 35, no. 7, pp. 1201-1222, 2003

 ISBN: 978-1-61208-117-5

An Architecture of Adaptive Product Data Communication System for Collaborative Design

Bernadetta Kwintiana Ane and Dieter Roller
Institute of Computer-aided Product Development Systems, Universität Stuttgart
Universitätsstr. 38, D-70569 Stuttgart, Germany.
E-Mail: ane@informatik.uni-stuttgart.de, roller@informatik.uni-stuttgart.de

***Abstract*—Today, designers and engineers on collaborative design environments often work in parallel and independently using different tools distributed at separate locations. Due to unique characteristic of engineering design, interaction during product development is difficult to support. As the information and communication technologies advances, computer supported collaborative design (CSCD) becomes more promising. Nevertheless, a potential problem remains between the product design and manufacturing, which mainly lies on the geometric shape of products that exists inherent in mass-customization. Meanwhile, CAD/CAM technologies have their own authoring tools, which govern the use of independent language and format for expressing various features and geometry. This condition creates incompatibility and has significant impact to the product costs. This paper is to address the incompatibility problem by introducing an architecture of the adaptive product data communication system. The adaptive system has a capability for autonomous tracking of design changes. The tracking model supports forward and backward tracking of constraint violation during the collaborative design transactions.**

Keywords-computer supported collaborative design; product data communication; adaptive.

I. BACKGROUND

Today's industry requires massive computer-supported technologies to address the increasingly complex product development tasks and the high expectations of customers. As the information and communication technologies advances, application of collaborative engineering to product design, so-called computer supported collaborative design (CSCD), becomes more promising.

Sprow [1] defines CSCD, or so-called cooperative design, as the process of designing a product through collaboration among multidisciplinary product developers associated with the entire product life-cycle. CSCD is carried out not only among multidisciplinary product development teams within a company, but also across the boundaries of companies and time zones, with increased numbers of customers and suppliers involved in the process.

Accomplishing a design task and delivering the results to manufacturing requires huge and complex information. Meanwhile, a potential problem remains between design and manufacturing, which mainly lies on the geometric shape of products that exists inherent in mass-customization [2]. Since the CAD/CAM technologies mostly govern independent authoring tools in different languages and formats, this condition creates incompatibility and has significant impact to the product costs. Therefore, synchronization of product data along the product development life-cycle is necessary.

This paper is to address the incompatibility problem that usually occurs in a collaborative design team by introducing an architecture of the adaptive product data communication system. The adaptive system is developed based on cloud computing technology, whereby shared servers provide resources, softwares, and data to designers and engineers on remote nodes on demand. Section 2 provides the framework of CSCD, Section 3 describes the architecture of the adaptive product data communication system, Section 4 describes the system ability for tracking of design changes, and, finally, Section 5 summarizes the conclusion of the paper.

II. COMPUTER SUPPORTED COLLABORATIVE DESIGN

Many researchers consider CSCD as an application of computer supported cooperative work (CSCW) in design. The term CSCW was first used by Greif and Cashman in 1984 to describe the topic on how to support people in their work arrangements with computers [3, 4]. Design has been one of the most important applications of CSCW technologies. With the rapid advancement of Web-based technologies, CSCD has progressed dramatically. The depth and breadth of CSCD applications are far beyond the traditional definition of concurrent engineering.

Technologies like CSCW and intelligent agents have been investigated to be effective to enhance communication, cooperation, and coordination among design team as well as software tools. The CSCW tools like *groupware* are used to facilitate communication among users. Meanwhile, in CSCD an *agent* can be considered as a software system that communicates and cooperates with other software systems to solve a complex problem, which is beyond the capability of each individual software system [5].

A. Web Technology for Collaborative Design

Since its emergence in 1993, Web has been quickly applied in the development of collaborative design systems. Along with the Web, a number of associated representation technologies have been developed, such as Hyper Text Mark-up Language (HTML), eXtensible Mark-up Language (XML), and Virtual Reality Mark-up Language (VRML), to enable better cross-platform and cross-enterprise exchange of multimedia information and design models. Many early collaborative design systems were developed using the Blackboard architecture [6] and distributed-object technologies like CORBA (Common Object Request Broker Architecture) [7], COM (Component Object Model) [8], and DCOM (Distributed Component Object Model). A blackboard architecture is a distributed computing

 ISBN: 978-1-61208-117-5

architecture where distributed applications, modelled as intelligent agents, share a common data structure called the "blackboard" and a scheduling/control process.

B. Integration of Web and Agent Technologies

A CSCD system developed with the Web as a backbone will primarily provide access to catalogue and design information on components and sub-assemblies, communication amongst multimedia formats, and authenticated access to design tools, services and documents. With the development of Web services and Semantic Web technologies, Web-based infrastructure has been used in a number of collaborative design systems. A Web-based collaborative design system usually uses a client/server architecture, in which the interaction between components is predefined. This kind of approach is considered insufficient to support dynamic collaborative design, where tasks are usually involving complex and non-deterministic interactions, producing results that might be ambiguous and incomplete. Hence, integration of Web and agent technologies to support collaborative design is considered crucial.

Software agents are mostly used for supporting cooperation amongst designers, enhancing interoperability between traditional computational tools, or allowing better simulations. An agent-based collaborative design system is a loosely coupled network of problem solvers that work together to solve complex problems that are beyond their individual capabilities [9]. Software agents in such systems are communicative, collaborative, autonomous, reactive (or proactive), and intelligent.

To date, many agent applications in the Web-based collaborative design still face many challenging questions. Coping with this issue, the concept of active Web server is introduced to integrate the Web and agent technologies [10]. The active Web server has driven the emergence of Web services concept [11]. As stated by the World Wide Web Consortium (W3C) [12], a Web service is a software system designed to support interoperable machine-to-machine interaction over a network.

III. Architecture Of Adaptive Product Data Communication System

In this section, an adaptive product data communication system is being introduced. The adaptive system is designed using an integrated Web and agent-based technologies for coordination in collaborative design environment. Here the term "adaptive" represents the ability of the agent to adapt with changes in the Web-based environment that commonly source from changes of application programs, data formats and structure, in such a manner in order to improve the system's future performance.

Design collaboration requires a higher sense of working together in order to achieve a holistic creative result [13]. It is a far more demanding activity, more difficult to establish and sustain, than completing a project in cooperation or coordination. Here the architecture of the adaptive communication system is designed based on STEP [2, 14], i.e., Standard for the Exchange of Product Model Data. STEP is an ISO standard for the computer-interpretable representation and exchange of industrial product data. The system architecture contains a shared product database management system (DBMS), which is composed on a low-level language, i.e., ASCII (American Standard Code for Information Interchange), as its native format. The database consists of geometry, topology, and auxiliary information. Considering complexity of engineering objects, a "reference" of geometry and topology is built into the DBMS that consists of taxonomy and data dictionary of elements geometry.

A. Taxonomy

The taxonomy is designed to be generic that workable under a variety of CAD applications. As a reference, the taxonomy has two functions. First, it identifies and generates particular geometric shapes. Second, it classifies the geometry into specific groups of objects (e.g., crankshaft, cantilever, motor-body, etc.). Fig. 1 describes the hierarchical structure of object classes, features, faces, and geometric entities in the parent-child relationship.

B. Data Dictionary

Data dictionary is a centralized repository of information about such data like meaning, relationships to other data, origin, usage, and format [15]. Data dictionary refers to a piece of "middleware" that supplants the native format of DBMS. Software agents are implant in the middleware in HTML, XML and VRML formats as interface to the CAD/CAM (i.e., Inventor and solidworks) applications. The middleware is modelled as an active object-oriented database (OOD). The active OOD is a database that allows users to specify actions to be taken automatically given certain rules when certain conditions arise [16]. In this architecture, the data dictionary is developed as an active semantic network (ASN) and realized as an active OOD. ASN is a shared database system developed to support designers during product development [17]. The goal of ASN is to represent all knowledge relevant to the collaborative product design teams. Fig. 2 describes the data dictionary in the ASN architecture.

C. ISO/OSI Data Communication Network

Since each CAD/CAM system has different proprietary native formats, data communication in collaborative design team should be done on a neutral format, i.e., .STEP file. The terms "neutral' means that the file format is independent of different formats utilized by the various computer-aided systems. Here the data communication network is designed based on the seven-layers ISO/OSI model as depicted in Fig. 3. An ISO/OSI model is an Open System Interconnect (OSI) model developed by the International Standards Organization (ISO). Therefore, this model is considered fit to the STEP standard. The model splits the communication process into seven layers, i.e., physical, data-link, network, transport, session, presentation, and application layers.

The physical layer deals with the electrical and mechanical means of data transmission. Data-link layer frames across a single local area network (LAN) and its

 ISBN: 978-1-61208-117-5

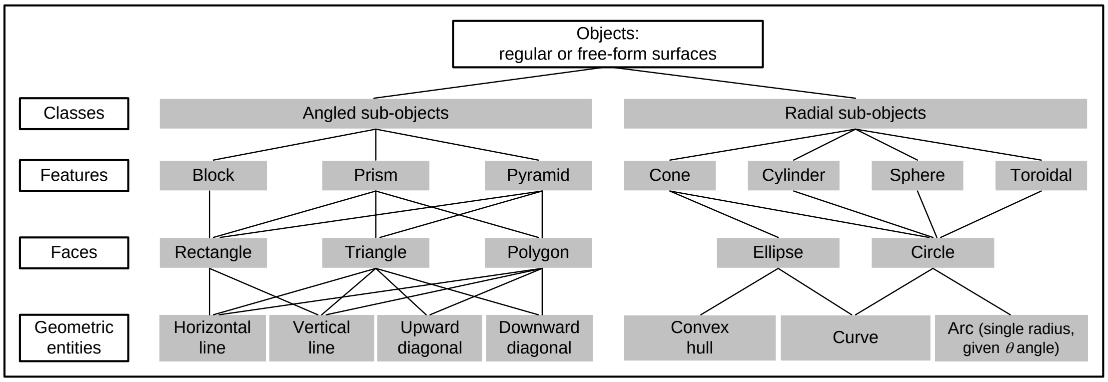


Figure 1. Taxonomy.

functions include resolution of contention for use of the shared transmission medium, delimitation and selection of frames addressed to this node, detection of noise via a frame check sequence, and any error correction or retries performed within the LAN. Network layer provides the transparent transfer of data between transport entities. Transport layer ensures that data units are delivered error-free, in sequence, with no losses or duplications. Session layer controls the dialogue between applications during a communication session. Presentation layer takes care of the syntax of the data exchanged between applications. Finally, application layer ensures that data transferred between any two applications are understood. Each layer can be developed independently and replaced without affecting the other layers.

When data is sent from workstation A to workstation B, it goes down the layers. At each layer, a control message is appended to the data. Then, the complete data is transmitted through the ISO/OSI medium to workstation B. At each layer of workstation B, the control message is stripped and proper actions are taken to convert the data into the proper format. Through an efficient data communication, conflicts and constraints can be analyzed earlier from different perspective. Hence, the collaborative design team can achieve the design objectives for an optimal product performance, at low manufacturing costs, and assurance that the product can easily and economically be serviced and maintained [18].

IV. Tracking Of Design Changes

One of the issues in collaborative design is that one must assess the impacts of a design change on other design objects and notify other parties promptly [19]. This paper adopts an approach to tracking of design changes introduced by Xie [20].

The mechanism for tracking of design changes is based on product data and their relationships. The product data contains descriptions for product specification, function decomposition structure, solution principles, layout design, assemblies, and parts. The relationships are established based on geometric constraint between two or more elements. The geometric constraint relationships define three types of constraint between parts, i.e., fit, contact, and consistent constraints. The fit constraint exists if there is a tolerance requirement between parts. The contact constraint represents a physical contact between two parts. The consistent constraint exists if two parts hold a dimensional constraint without a physical contact.

This approach supports forward tracking and backward tracking of a design change. Forward tracking identifies the impact of the change on later design stages if a design change occurs at an earlier stage. On the other hand, backward tracking identifies the impacts of changes on previous stages, if a change occurs at a later stage. The design change rules are stored in a knowledge base so that all the impacts can be retrieved through an inference engine. Therefore, designers can identify the total impacts of a proposed design change on an entire product development life-cycle.

To make the necessary design information available, product data information is extracted from design process and represented in a data model. A data model is a set of concepts that can be used to describe the structure of DBMS [21]. Here an entity-relationship model is used to describe the concepts of entities, attributes, and relationships.

In this regard, the change tracking model involves five entities, i.e., *Specification*, *Function*, *Principle*, *Design_ object*, *Assembly*, and *Part*. These entities are associated to 19 attributes, i.e., *Buy-or-make*, *Category*, *Classification*, *Cost*, *Criteria*, *Description*, *Dimension*, *High-limit*, *ID*, *Low-limit*, *Mass*, *Materials*, *Measurement*, *Quantity*, *Selected*, *Source-form*, *Tolerance*, *Type*, and *Unit*. Meanwhile, the relationships represent a set of associations amongst entities. Cardinality ratio constraints specify three common combinations for binary relationship types, i.e., one-to-one (1:1), one-to-many (1:M), and many-to-many (M:N). The relationships in this model include *Requires*, *Contains*, *Previous*, *Solution*, *Implement*, *Belongs*, and *Constraint*. Fig. 4 describes the entities and their associated attributes in the prescribed relationships.

 ISBN: 978-1-61208-117-5

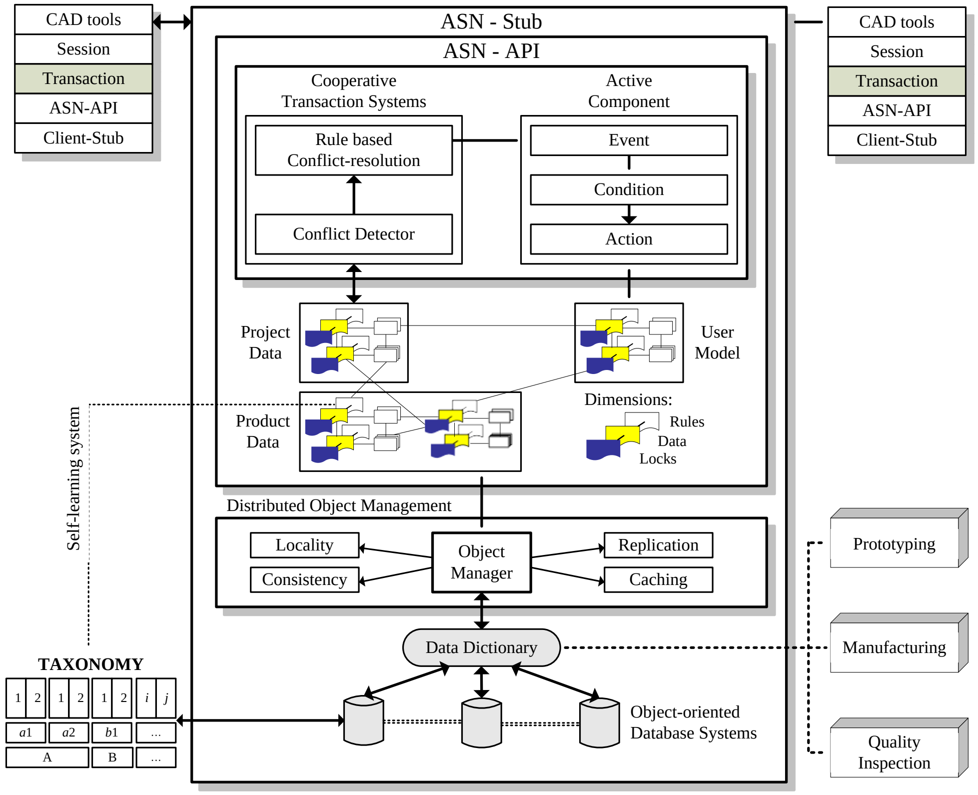


Figure 2. Data Dictionary In Active Semantic Network.

V. CASE STUDY

To verify the changes tracking model, a case study is applied on the motor-body. The initial design of motor-body has cylinder shape with 1.751 kg mass and 0.000227 m^3 volume of alloy steel (SS). The real structure sustains a distributed state of stress. The stress is represented by forces at the element joints or nodes. Correspondingly, the displacement of these points is employed in the characterization of displaced state of the element.

Generally, structural analysis problem can be treated as linear static problem under assumptions small deformation (i.e., loading pattern is not changing due to the deformed shape), elastic material (i.e., no plasticity), and static load (i.e., the load is applied to the structure in a slow of steady operation). Therefore, the force-displacement analysis is applied to the motor-body. The relationship between the joint forces and the joint displacements of finite elements should satisfy the stiffness function,

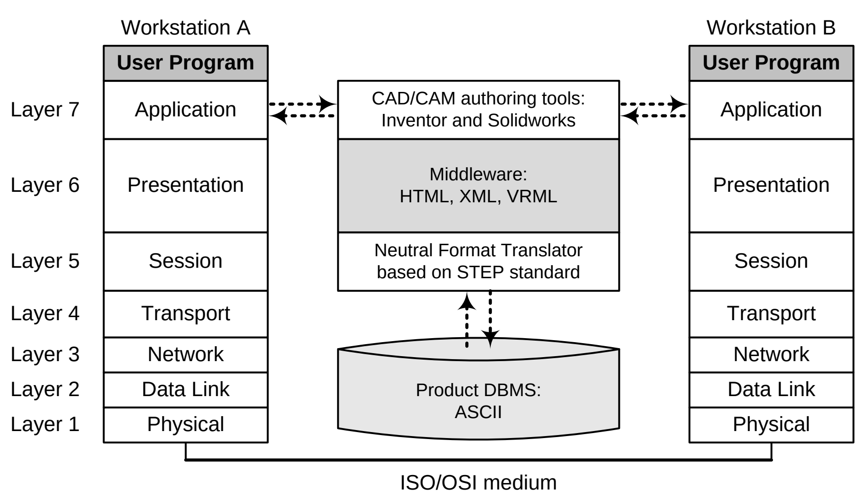


Figure 3. ISO/OSI Data Communication Network Model.

 ISBN: 978-1-61208-117-5

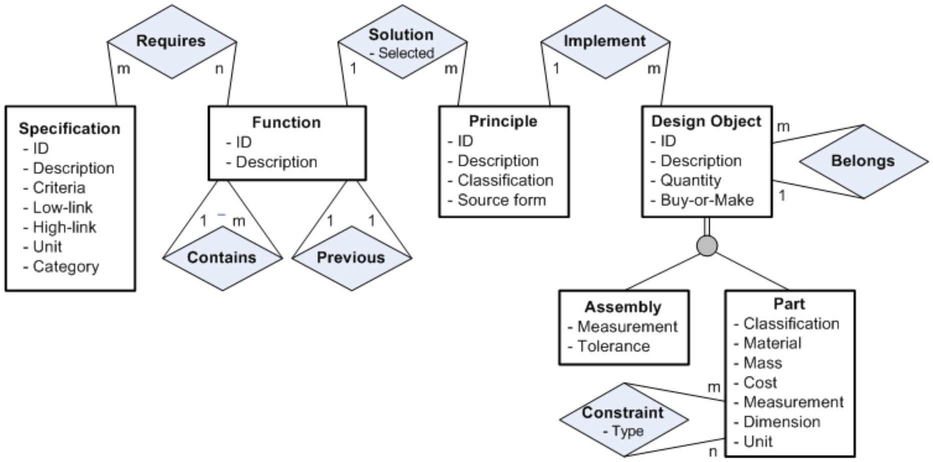


Figure 4. E-R Diagram For Change Tracking Model (Source: Xie [19]).

$$\{F\} = [k]\{\Delta\} \quad (1)$$

where {F}: element force, {Δ}: displacement vectors, and [k]: element stiffness matrix. An individual term of the [k] matrix, k_{ij}, is an element stiffness coefficient. When the displacement Δ_j is imposed at unit value and all other degree of freedom are held fixed against displacement ($\Delta_k = 0$, $k \neq j$) the force F_i is equal in value to k_{ij}.

The force-displacement analysis produces an average deformation scale at 1.72501e^{+008} and prediction of location where the most deformed mesh are possible to occur. The resultants displacement shows the minimum condition 0 mm is at location (3.969 cm, -0.499 cm, -11.000 cm) and the maximum condition 6.76612e-008 mm at location (-3.373 cm, -3.291 cm, -0.099 cm). The analysis predicts that two most possible deformed locations likely to occur at the lower-part of cab-screw holes as depicted in Fig. 5. This condition makes the cylinder shape has more possibility to slip from its position and fixtures.

Furthermore, the Von Mises stress [22] analysis is applied

$$\sigma_e = \frac{1}{\sqrt{2}}\sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2} \quad (2)$$

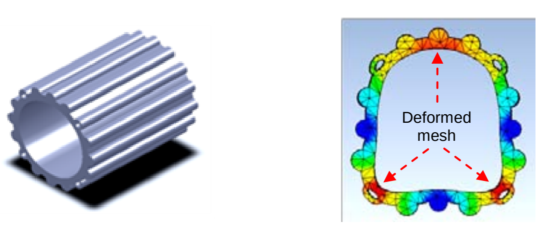
where σ_1, σ_2, and σ_3: three principle stresses at the considered point in a structure. For a ductile material, the stress (σ_e) and the yield stress of the material (σ_y) must satisfy the constraint

$$\sigma_e \leq \sigma_Y. \quad (3)$$

The results estimate minimum stress 3.04976e^{-006} N/mm^2 (MPa) at location (4.170 cm, -0.470 cm, -6.750 cm) and maximum stress 0.000572322 N/mm^2 (MPa) at location (3.524 cm, -3.385 cm, -0.350 cm). The stress is distributed from the inner cylindrical mesh boundary to the outer boundary with the highest strained locations are found at the elements adjacent to the four cab-screw holes. This condition makes the initial design has high potential failure during the assembly and product use. Therefore, it needs to be redesigned.

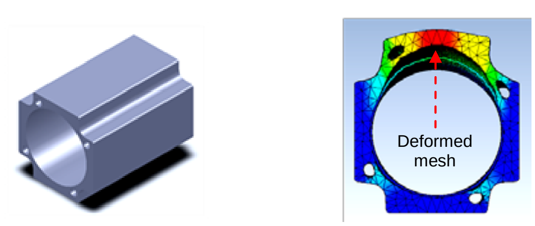
In this regard, design improvement is done based on entities, attributes, and relationships which have been defined in the E-R diagram. Change of *Design_object* from cylinder to block shape has driven change of specification, part, and assembly respectively. The progress for forward tracking of *Design_object* change is described in Table I.

As a result, a block shape of motor-body in dimension 82.5 x 82.5 x 100 millimeters with 1.859 kg mass and 0.000241 m^3 volume of alloy steel (SS) is obtained. The new design has 6.15% more weight than the initial design, but shows better performance. The force-displacement analysis of the new design produces an average deformation scale 8.59488e^{+007}, i.e., 50.18% better than the initial design. In the new design, the deformation has been localized at the upper-front to -middle of finite mesh as depicted in Fig. 6. Moreover, the von Mises stress analysis shows the results of minimum stress 3.43238e-^{007} N/mm^2 (MPa) and maximum stress 0.00139487 N/mm^2 (MPa). These structural problems are expected to be further minimized during assembly when the motor-body is joined with the motor-cover.



(*a*) cylinder shape (*b*) estimated displacement

Figure 5. Initial Design



(*a*) block shape (*b*) reduced displacement

Figure 6. Improved Design

 ISBN: 978-1-61208-117-5

TABLE I. FORWARD TRACKING OF DESIGN OBJECT CHANGE

ENTITIES	ATTRIBUTES	STATUS	DESIGN VALUES	
			INITIAL	IMPROVED
Function	ID	No change	part#1	part#1
	Description	No change	Motor-body	Motor-body
Principle	Classification	No change	Motor protection	Motor protection
	Source Form	No change	Production house	Production house
Design_Object	Quantity	No change	100 pieces	100 pieces
	Buy-or-Make	No change	make	make
Specification	Criteria	No change	mechanic – static	mechanic – static
	Low-limit	Change	3.04976e-006 N/mm^2	3.43238e-007 N/mm^2
	High-limit	Change	0.000572322 N/mm^2	0.00139487 N/mm^2
	Unit	No change	1	1
	Category	No change	automotive part	automotive part
Part	Material	No change	alloy steel (SS)	alloy steel (SS)
	Mass	No change	min 1.750 – max 1.860 kg	min 1.750 – max 1.860 kg
	Cost	No change	USD 367.82 - USD 375.00	USD 367.82 - USD 375.00
	Measurement	No change	millimeter (cm)	millimeter (cm)
	Dimension	Change	d⌀: 82.5mm, ℓ: 110 mm	w: 82.5 mm, h: 82.5 mm, ℓ: 110 mm

VI. CONCLUSION

Today, design activity is inevitable should be done as an integrated process with design optimization and manufacturing. In a collaborative design environment, the product development activities usually take place at geographically distributed locations.

This paper introduces an adaptive product data communication system that is developed by making used of the integrated Web and agent-based technologies. The architecture of the adaptive system is designed based on STEP standard. The system contains a shared product database management system (DBMS), which is composed on a low-level language as its native format. Meanwhile, the data communication network is developed based on the seven-layer ISO/OSI model. Considering high possibility of constraint violation during the collaborative design transactions, a capability for autonomous tracking of design changes is built in to the adaptive system. The tracking model supports forward and backward tracking of design changes. Therefore, it enables designers and engineers to identify the total impacts of a proposed design change on an entire product.

ACKNOWLEDGMENT

This research project is organized under the financial support of The Alexander von Humboldt Foundation and the Institute of Computer-aided Product Development Systems, Universität Stuttgart in Stuttgart, Germany.

REFERENCES

[1] E. Sprow, “Chrysler’s Concurrent Engineering Challenge,” Manufacturing Engineering, vol. 108, issue 4, 1992, pp. 35-42.

[2] B. K. Ane, “Automation of Design for Reverse Engineering in a Collaborative Design Environment: Development of STEP Neutral Format Translator for Product Data Exchange,” Research Report, Institute of Computer-aided Product Development Systems, Universität Stuttgart, Stuttgart-Germany, 2010.

[3] K. Schmidt and L. Bannon, “Taking CSCW Seriously,” Computer Supported Cooperative Work, vol. 1, issue 1/2, 1992, pp. 7-40.

[4] I. Greif, Computer-Supported Cooperative Work: A Book of Readings, San Mateo, CA: Morgan Kaufmann Publishers, 1988.

[5] W. Shen, Q. Hao, H. Yoon, and D. H. Norrie, “Applications of Agent Systems in Intelligent Manufacturing: An Update Review,” International Journal of Advanced Engineering Informatics, vol. 20, issue 4, 2006, pp. 415-431.

[6] V. Jagannathan, R. Dodhiawala, and L. S. Baum, Blackboard Architectures and Applications, San Diego, CA: Academic Press, 1989.

[7] Object Management Group, Common Object Request Broker Architecture (CORBA) Specification: Version 3.1, Needham, MA: OMG, 2008.

[8] D. Box, Essential COM, Upper Saddle River, NJ: Addison-Wesley. 1998, ISBN: 0-201-63446-5.

[9] W. Shen, D. H. Norrie, and J. P. Barthes, Multi-Agent Systems for Concurrent Intelligent Design and Manufacturing, London, UK: Taylor and Francis, 2001.

[10] W. Shen, “Web-based Infrastructure for Collaborative Product Design: An Overview,” Proc. of the 5th International Conference on Computer Supported Cooperative Work in Design, 2000, pp. 239-244.

[11] W. Shen and Q. Hao, “A Service Oriented Framework for Blow Molded Automotive Parts Design and Optimization,” Proc. of the SAE 2004 Congress, 2004, SAE 2004-01-1244.

[12] W3C Working Group,” Web Services Architecture,”World Wide Web Consortium, 2004, http://www.w3.org/TR/2004/NOTE-ws-arch-20040211, accessed on January 20, 2011.

[13] K. Huang, X. Li, S. Cao, B. Yang, and W. Pan, “Co-DARFAD - The Collaborative Mechanical Product Design System,” Proceed. of the Sixth International Conference on Computer Supported Cooperative Work In Design, 2001, pp. 163-168.

[14] International Standard Organization, Industrial Automation System and Integration – Product Data Representation and Exchange – Part 31: Conformance Testing Methodology and Framework: General Concepts, Geneva, Switzerland, ISO Central Secretariat, 1992.

[15] IBM Dictionary of Computing, “IBM Terminology,” 2008, http://www-01.ibm.com/software/globalization/terminology/d.html, accessed on January 20, 2011.

[16] D. Roller and O. Eck, “Constraint propagation using an Active Semantic Network,” in Geometric Constraints Solving and Applications, (Eds.) B. Brüderlin and D. Roller, Berlin. Springer-Verlag, 1998.

[17] B. Bernd, G. Heiko, and W. Michael, „Aktives Semantiches Konstruktions- und Zuverlässigkeitsnetz. In: Entwicklung und Erprobung Innovativer Produkte-Rapid Prototyping: Grundlagen, Rahmenbedingungen und Realisierung,” (Eds.) B. Bertsche and H.-J. Bullinger, Berlin. Springer-Verlag, 2007, pp. 130-158.

[18] J. Harrley, Concurrent Engineering, Cambridge, MA: Productivity Press, 1992.

[19] L. Wang, W. Shen, H. Xie, J. Neelamkavil, and A. Pardasani, “Collaborative Conceptual Design: A State-of-the-art Survey,” Proc. of the sixth International Conference on Computer Supported Cooperative Work in Design, 2000, pp. 204-209.

[20] H. Xie, “Tracking of Design Changes for Collaborative Product Development ,” Proc. of the Sixth International Conference on Computer Supported Cooperative Work in Design, 2001, pp. 175-180.

[21] R. Elmasri and S . B. Navather, “Fundamentals of Database Systems,” New York: Benjamin Cummings Publishing Company Inc., 1989.

[22] C. L. Liua, Z. Z. Lua, Y. L. Xub, and Z. F. Yuec, “Reliability Analysis For Low Cycle Fatigue Life of the Aeronautical Engine Turbine Disc Structure Under Random Environment,” Materials Science and Engineering, vol. A, issue 395, 2005, pp. 218-225.

Towards Implicit Enhancement of Security and User Authentication in Mobile Devices Based on Movement and Audio Analysis

Hamed Ketabdar, Mehran Roshandel, Daria Skripko
Deutsche Telekom Laboratories, TU Berlin
Ernst-Reuter-Platz 7, 10587, Berlin, Germany
{hamed.ketabdar, mehran.roshandel, daria.skripko}@telekom.de

***Abstract*—In this work, we present initial investigations towards analysing movements of a mobile device as well as ambient audio captured by the mobile device for enhancing security functionalities in the device. We present a few scenarios which can lead to security threats related to data or services on a mobile device (e.g., a phone being lost or stolen). We show how unexpected movements or ambient audio captured by the device can deliver information which can be important considering security issues, and reveal those scenarios. In addition, we present how the identity of a user can be verified (or identified) by his mobile device based on pattern of his regular physical activities such as walking. This allows for implicit and continuous re-identification of the user. The implicit process does not require active participation of user, and allows for authentication during regular daily activities. The proposed method can also be used to complement regular authentication techniques to protect for example an open account on a mobile phone. It can also help to reduce number of re-authentications by verifying that the mobile device is continuously operated by the same user since the last regular authentication. Our final goal is to come up with a correlation model describing the relationship between movements of a mobile device as well as ambient audio, and security related issues.**

Keywords-Security; Mobile Devices; Motion and Audio Analysis; Embedded Sensors; Unexpected Events; Implicit Authentication

I. INTRODUCTION

Mobile devices are one of the essentials in our daily life used for communication, storage and service access. As mobile devices technology develops, they are used more and more for storing different data, text, audio, photos, etc. Obviously, some of this data can have private or confidential content. In addition, mobile devices are also becoming a gateway to connect with many different services such as Email, E-banking, etc. Most of these services can also be related to business, confidential or private aspects of user's life. Unfortunately, there is always a risk that these confidential data or services are exposed to unauthorised people, for instance when a mobile phone is lost or stolen.

If a person left his Personal Digital Assistant (PDA) or smartphone in a cab, a power-on password would prevent anyone who found it from casually browsing content, racking up calls, and using his email account. However passwords can be shared, guessed, or stolen. Enforcing minimum password length and complexity rules can make password authentication more effective, but they do not improve usability. Compared to laptops, PDAs and smart phones are used far more frequently, for much shorter tasks, demanding near-instantaneous availability. Authentication methods that get in the way of using these frequently used shorter tasks are disabled. This is the main reason why Personal Identification Number (PIN) is unused on mobile devices. Convenience often trumps security, especially if nothing enforces policy [1]. The same argument can also be applied to other means of authentication such as finger print [2], face profile [3], voice based verification [4], etc. All these approaches are intermittent and therefore susceptible to attack, e.g., an unauthorized user can access a portable computer either by stealing a password or exploiting an open account of a user [5].

In this work, we propose a new paradigm for increasing security of data and service access on mobile devices, based on analysis of movement and audio data captured by the device. The new paradigm allows for online, implicit and continuous protection of data without the burden of involving the active attention of the user. We show that analysis of audio and physical movement data captured by a mobile phone can indicate unexpected events which can lead to having the phone being lost or stolen. In addition, we show that analysis of audio and physical movement data during regular physical activities (e.g., walking) by the user can allow for authenticating/identifying the user. The proposed method is an implicit authentication technique, i.e., it does not involve active attention of the user, and it is performed continuously as user is regularly using or carrying the device. Physical movement data can be captured by accelerometer sensors embedded in modern mobile devices. Audio data is captured by embedded microphone.

Such a paradigm can be used to increase security of data and service access on mobile devices as a stand-alone technique, or as complementary to regular authentication techniques. It can for instance protect an open account from an unauthorised user. In addition, as complement of regular authentication techniques (e.g., PIN, signature, finger print), the number of regular re-authentications can be reduced if our method detects that the same user has been continuously using the device. Moreover, the implicit security protection process can be used to implement a "Graded Security" scheme for data and service access. In this scheme, a security level score is calculated based on the outcome of audio and movement analysis. According to the calculated security level, different access policies can be established. This scheme allows protecting data and services according to their importance and security threat level of mobile device.

The paper mainly studies two cases related to audio and movement analysis for enhancing security functionalities in mobile devices. The first case is detecting unexpected events which can lead to having the mobile device being lost or stolen. This is for instance the situation that a phone falls accidentally out of the user's pocket/bag and is left unattended. The second case is using audio and movement analysis for user identification/authentication. In this case, movement and audio data captured during physical activities of the user is used as a basis for his identification. We talk about the first case in Section 3, and the second case in Section 4.

II. ANALYSIS OF MOVEMENT AND AUDIO DATA

Information about movement of the mobile device is obtained by analysis of data provided by integrated acceleration sensors. These sensors are originally used for automatic screen rotation and navigation [6, 7, 8]. Acceleration sensors integrated in a mobile device provide linear acceleration information along the x, y and z directions. The acceleration sensed by the mobile device can be due to different sources. In this work, we are mainly interested in components of acceleration caused by physical activities of user, or unexpected events such as free falls, impacts, etc. According to our experiments, these components usually appear in high frequency content of acceleration signals. Lower frequency components can be mainly due to gravity force, as well as movements of the user in a vehicle. In most of the cases, we pre-process the acceleration signals with a time derivative operation which effectively acts as a high pass filter.

Audio data is also captured using the microphone embedded in the mobile device.

In order to analyse data captured by the accelerometer or microphone, we usually extract some features from the data in certain time intervals (windows). These features are mainly based on average, variance, and rate of change of recorded signals in the interval. For instance, the average of norm of acceleration signals (along x, y and z directions) in a time interval can indicate the level of physical movement of the device in the interval.

III. UNEXPECTED EVENT DETECTION

Unexpected events experienced by a mobile device can be a sign of security threats. In this section, we review a few unexpected scenarios which can lead to security threats related to mobile devices. We further discuss how these situations can be detected based on analysis of captured motion and audio information using sensors and microphone embedded in a mobile phone.

We start with a simple but practical case. If a mobile phone accelerometer has not detected motion for a relatively long period of time, it may indicate that the phone is lost or forgotten somewhere. This may result in a security risk for data or services accessed by the phone. This situation can be easily identified by analyzing motion data obtained from the embedded acceleration sensors. In this case, the rate of change of acceleration data can be quite low over a long period of time. Upon detection of such a situation, the device

Figure 1. A mobile phone under risk of being lost or stolen.

can be locked requesting user re-authentication.

Another risky situation can be when the user is engaged in high level of physical activity (e.g., running or sports). In such a situation, attention of user to his mobile phone may be reduced, resulting in a relatively higher risk of losing the device. Detecting such a situation based on embedded accelerometer sensors allows setting security options on the mobile device to a higher level. According to our experiments (assuming that the mobile device is carried somewhere on the user body), the average and variance of acceleration show significantly high values in these cases, as compared to regular situations.

A common risky situation which can lead to having the device lost or stolen is when the mobile device falls out of user's pocket or bag, and is left unattended for a while (Fig.1). The user may not notice what has happened to the device due to distractions with other activities. We have studied detecting such a scenario based on analysing audio and acceleration information. We model this unexpected event based on three sub-events: free-fall, shock (impact with floor), and no activity (movement) after the shock. When the phone falls, it initially experiences a free fall situation. Upon hitting the floor, it experiences a shock. However, this is not enough for identifying the situation as "risky", because the user may immediately pick up the phone. Therefore, it is also necessary to check for a period of "no activity" after the shock.

We have setup user studies to evaluate our algorithm for detecting the last mentioned scenario. The experimental setup is similar to [9]. We have used iPhone 3G for the experiments. Acceleration and audio data is recorded using embedded microphone and sensors through a data collection application developed for iPhone. For the experiments, we have recorded a database of normal and risky (as defined) situations. In this database, there are 98 samples of normal situations, and 36 samples of physical shocks. In order to obtain physical shock, we let the iPhone to fall on a carpet or wooden floor form a distance of approximately 75 cm. In order to obtain normal condition samples, we let 5 users to carry the iPhone normally in their pocket, hand or bag for a

TABLE I. RESULTS FOR DETECTION OF A RISKY SITUATION WHICH CAN LEAD TO HAVING A MOBILE DEVICE BEING LOST OR STOLEN.

Algorithm	Accuracy	True Alarm	False Alarm
3 step definition	**94.4**	34	4
Only impact	86.1	31	9

period of 10 seconds. These users indulge in different day-to-day activities such as walking, jogging, taking lift and walking on stairs. We tried to have different variety of scenarios, especially those which can have similarity to a shock (due to high physical activity) such as walking on stairs and taking lift. In this way, we can make sure that our algorithm is able to distinguish between such cases and a real risky shock [9].

As mentioned earlier, the risky situation is defined as a sequence of free fall, impact (shock), and no activity period. The free fall is detected when the norm of acceleration signals (along x, y, and z directions) falls below a predefined threshold. The no-activity period is identified when the average of norm of acceleration signals in an interval after the impact (8 seconds in our case) falls below a threshold. The impact (shock) situation is detected by comparing features extracted from acceleration and audio data against a statistical model created for impact (shock). The model which is used for this work is a Multi-Layer Perceptron (MLP) trained using samples of shock (impact) and regular situation collected as previously mentioned. The features used in this study are mainly based on average and variance of acceleration components (as well as their norm), and audio signal. The MLP is then able to classify new samples of features as shock or normal (regular) situation. The “risky situation” is detected upon detection of free fall, shock, and period of no-activity in correct order. Table 1 summarizes initial results. Our studies show that defining the three steps for risky events detection can significantly reduce number of false alarms (Table 1). The first row in the table shows the results when the three step definition is used, and the second row shows results when only impact is considered as risky event.

IV. IMPLICIT IDENTITY VERIFICTION/IDENTIFICTION BASED ON AUDIO AND MOVEMENT ANALYSIS

Another possibility for using audio and movement analysis in enhancing security functionalities in mobile devices is for user authentication/identification based on regular user’s physical activities such as walking. When the mobile device is carried by the user (e.g., in his pant pocket), it can capture samples of audio and motion information, and check for a biometric patterns in them. In this way, the identity of the user can be verified in a continuous and implicit manner. The authentication is implicit, so the user does not need to actively participate in authentication process. The user only performs his regular activities and the authentication method looks for a biometric sign in his pattern of physical activities. As discussed before, this implicit authentication method can be used alone or as complementary to regular authentication methods. The device can automatically detect that it is not being carried or operated by the same user anymore, therefore switch to a higher level of required authentication. As a side advantage, this technique can reduce required number of normal authentications. If the device implicitly detects that it has been continuously used by the same person since the last authentication, it may not ask for a new authentication process for the same service. This reduces the number of repetitive re-authentication. In addition, an implicit authentication score estimated based on audio and movement analysis can be used to set up different security threat levels for the mobile device, allowing implementation of a graded security scheme.

In the following, we present our initial experiments for user identification/authentication based on data captured by a mobile device (audio and motion) during regular physical activities. We show that users can be classified with high accuracy based on captured information using a mobile phone in their pant pocket.

V. EXPERIMENTS AND RESULTS

We set initial experiments to investigate possibility of implicit user authentication/identification based on regular physical activities of user (walking in our case).

For experiments, we recorded device motion information (using embedded acceleration sensors) as well as ambient audio (using embedded microphone). The recoding is done during regular physical activities which are walking in this case. The device is carried in user’s pant pocket. We have used iPhone as mobile device and we recorded the signals using a data collection application we developed for iPhone.

We have invited 9 participants for the experiments. We captured audio at 8 KHz and acceleration at 50 Hz using embedded sensors in the iPhone. We let the iPhone to be placed regularly in the pocket, without fixing its position or orientation. The test users are asked to walk for about 2 minutes in indoor and outdoor environments. The recording for each user is repeated over 3 different days. Users were asked to come for the experiments with different sets of shoes and pants in different days, in order to take into account the effect of variability in clothing in the identification process.

Feature extraction is the first processing step. We extract two sets of features, one from acceleration signals and one from audio signal. Features are extracted over a window of 2 seconds of acceleration and audio signals. For acceleration signals, the extracted features are mainly based on average, variance and magnitude of acceleration components. Here is a list of features:

- Average field strength along x, y, and z directions.
- Variance field strength along x, y, and z directions.
- Average of Euclidian norm of filed strength along x, y, z.
- Variance of Euclidian norm of field strength along x, y, and z.
- Piecewise correlation between field strength along x and y, x and z, and y and z.

For audio signal, extracted features are mainly based on average, variance, and energy of the audio signal in each window. Variance of Fourier transform of audio signal is also used as a feature.

Extracted features are feed as input to MLP for user classification/identification. Table 2 shows classification results for different feature sets. We report results for using movement (acceleration) based features, audio based features, and a combination of movement and audio based features. As can be seen from the table, the combination of audio and movement based features provide the best user identification results (90.1%). Table 2 shows identity verification (authentication) measures for some of the users. The Receiver Operating Characteristic (ROC) measurements show a good tradeoff between true and false alarms indicating significant user authentication results.

In this experiment, we have presented initial results for user identification/authentication over a window period of 2 seconds. This means that every 2 seconds, we are able to re-authenticate the user. However, such a short interval continuous re-authentication may not be necessary in practical applications. It may be enough to have an authentication measure for instance, every minute. In such a case, short interval (2 seconds window) based authentication results can be used in a voting scheme. The identified user over a minute is the user having highest vote (recognition) in 2 second based windows. Our experiments show that user identification accuracy in this case rises to 97.5% using combination of audio and acceleration based features.

TABLE II. USER IDENTIFICATION RESULTS USING DIFFERENT FEATURE SETS(MOVEMENT, AUDIO, MOVEMENT+AUDIO)

Feature source	Accuracy
Movement	88.3
Audio	47.8
Movement + Audio	**90.1**

TABLE III. USER AUTHENTICATION MEASURES FOR SOME USERS

User ID	Precision	Recall	F-measure	ROC Area
1	0.89	0.95	0.92	0.98
2	0.92	0.87	0.90	0.96
3	0.92	0.91	0.92	0.98
4	0.92	0.92	0.92	0.97
Weighted Average	**0.91**	**0.91**	**0.91**	**0.97**

VI. DEMONSTRATOR

We have developed a demonstrator based on the proposed methodology for Apple iPhone mobile device. The demonstrator can detect an unexpected situation involving a free fall, impact and a period of no activity. Upon detection of such a risky unexpected situation, the demonstrator can automatically block access to the phone and ask for a password. It can also optionally send a message including the location of the mobile phone to a designated number.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented our initial investigations on the correlation between motion and audio information captured using embedded sensors in a mobile device for enhancing security functionalities related to the device. We have mainly investigated two cases. In the first case, we detect unexpected events, based on audio and movement analysis. We showed that an unexpected event such as a phone fallen down and left unattended can be identified with a high accuracy. For the second case, we proposed implicit user identification/authentication based on audio and movement analysis during regular physical activities (e.g., walking). We showed that a user can be identified with high accuracy on this basis. The results of such analysis can be used to arrange a graded security scheme for mobile and handheld devices based on their actual status. The proposed framework can be used as a stand-alone implicit security enhancement technique, or used as a complement to regular user authentication techniques. These results can be an initiation for a new security paradigm for enhancing security functionalities in mobile devices based on audio and movement analysis.

This work can be further developed by extending investigations for finding a general model describing the correlation between movements of a mobile device as well as ambient audio, with security risks. There are many other factors such as where the device is carried (e.g., bag, pocket, etc.) which can be also highly correlated with security of data and services on the device. Our proposed method can be also used for automatic profile management when a mobile device is shared between several users.

REFERENCES

[1] Improving Mobile Authentication http://searchmobilecomputing.techtarget.com/tip/Mobile-device-security-Improving-mobile-authentication, retrieved 12/12/2010.

[2] P. Gupta, S. Ravi, A. Raghunathan, and N. K. Jha, "Efficient Fingerprint-Based User Authentication for Embedded Systems," Design Automation Conf., pp. 244-247, June 2005.

[3] N. Aaraj, S. Ravi, A. Raghunathan, and N. K. Jha, "Architectures for Efficient Face Authentication in Embedded Systems," Proc. Design, Automation & Test in Europe, pp. 1-6, 2006.

[4] C. C. Leung, Y. S. Moon, and H. Meng, "A Pruning Approach for GMM-Based Speaker Verification in Mobile Embedded Systems," Lecture Notes in Computer Science, pp. 607-613, 2004.

[5] A. Yazji, X. Chen, R. P. Dick, and P. Scheuermann, "Implicit User Re-Authentication for Mobile Devices," Proceedings of the 6th International Conference on Ubiquitous Intelligence and Computing, pp. 325-339, Brisbane, 2009.

[6] J. Rekimoto, "Gesturewrist and gesturepad: Unobtrusive wearable interaction devices," In Proceedings of Fifth International Symposium on Wearable Computers, 2001.

[7] J. Rekimoto, "Tilting Operations for Small Screen Interfaces," UIST'96, pp. 167-168.

[8] K. Hinckley, J. Pierce, and E. Horvitz, "Sensing Techniques for Mobile Interaction," In Proceedings of ACMUIST, pp. 91-100, 2000

[9] H. Ketabdar, "Detecting physical shock by a mobile phone and its applications in security and emergency," Proceedings of MobileHCI 2009, Bonn, Germany.

An Augmented Reality Application for the Enhancement of Surgical Decisions

Lucio T. De Paolis, Giovanni Aloisio
Department of Innovation Engineering
Salento University
Lecce, Italy
lucio.depaolis@unisalento.it
giovanni.aloisio@unisalento.it

Marco Pulimeno
Engineering Faculty
Salento University
Lecce, Italy
marco.pulimeno@unisalento.it

Abstract— **The practice of Minimally Invasive Surgery is becoming more and more widespread and is being adopted as an alternative to the classical procedure. This technique has some limitations and comes at a cost to the surgeons. In particular, the lack of depth in perception and the difficulty in estimating the distance of the specific structures in laparoscopic surgery can impose limits on delicate dissection or suturing. The availability of new systems for the pre-operative planning can be of great help to the surgeon. The developed application allows the surgeon to gather information about the patient and her/his pathology, visualizing and interacting with the 3D models of the organs built from the patient's medical images, measuring the dimensions of the organs and deciding the best insertion points of the trocars in the patient's body. This choice can be visualized on the real patient using the Augmented Reality technology.**

Keywords - user interface; Augmented Reality; medical image processing

I. INTRODUCTION

One trend in surgery is the transition from open procedures to minimally invasive laparoscopic operations where visual feedback to the surgeon is only available through the laparoscope camera and direct palpation of organs is not possible.

Minimally Invasive Surgery (MIS) has become very important and the research in this field is ever more widely accepted because these techniques provide surgeons with less invasive means of reaching the patient's internal anatomy and allow entire procedures to be performed with only minimal trauma to the patient.

The diseased area is reached by means of small incisions inside the body; specific instruments and a camera are inserted in the body and what is happening inside the body is shown in a monitor. The surgeon does not have a direct vision of the organs and he is thus guided by the camera images. This surgical approach is very different from the one of open surgery, where the organ can be fully visualized and handled.

As a promising technique, the practice of MIS is becoming more and more widespread and is being adopted as an alternative to the classical procedure. The advantages of using this surgical method are evident for the patients because the possible trauma is reduced, the postoperative recovery is nearly always faster and scarring is reduced.

Despite the improvement in outcomes, these techniques have their limitations and come at a cost to the surgeons. In particular, the lack of depth in perception and the difficulty in estimating the distance of the specific structures in laparoscopic surgery can impose limits on delicate dissection or suturing.

Due to the great deal of difficulties involved in MIS related to perceptual disadvantages, many research groups, motivated by the benefits MIS can bring to patients, are now focusing on the development of surgical assistance systems.

On the other hand, advances in technology are making possible to develop systems that can help surgeons to perform their tasks in ways that are both faster and safer.

Latest developments in medical imaging acquisition and computer systems make possible the reconstruction of 3D models of the organs providing anatomical information barely detectable by CT and MRI slices or ultrasound scan as well as the safe guidance of instruments through the body without the direct sight of the physician.

The emerging Augmented Reality (AR) technology has the potential to bring the advantage of direct visualization in open surgery back to minimally invasive surgery and can increase the physician's view of his/her surroundings with information gathered from patient medical images.

In contrast with Virtual Reality technology where the user is completely immersed in a synthetic environment and cannot see the real world around him, AR technology, which consists of the addition of extra information to the real scene, allows the user to see virtual objects in addition to the real world. The user is under the impression that the virtual and real objects coexist in the same space.

In medicine, Augmented Reality technology makes it possible to overlay virtual medical images onto the patient, allowing surgeons to have a sort of "X-ray" vision of the body and providing them with a view of the patient's anatomy. The patient becomes transparent and this virtual transparency will therefore make it possible to find tumors or vessels not by locating them by touch, but simply by visualizing them through Augmented Reality.

The virtual information could be directly displayed on the patient's body or visualized on an AR surgical interface, showing where the operation should be performed.

 ISBN: 978-1-61208-117-5

This paper presents an advanced platform for the visualization and the interaction with the 3D patient models of the organs built from CT images.

The availability of a system for the pre-operative planning can be of great help to the surgeon and this support is even more important in pediatric laparoscopic surgery where a good understanding is needed of the exact conditions of the patient's organs and the precise location of the operational site.

In addition, the developed application allows the surgeon to choose the points for the insertion of the trocars on the virtual model and to overlap them on the real patient body using the Augmented Reality technology.

This work is part of the ARPED Project (Augmented Reality Application in Pediatric Minimally Invasive Surgery) funded by the Fondazione Cassa di Risparmio di Puglia. The aim of the ARPED project is the design and development of an Augmented Reality system that can support the surgeon through the visualization of anatomical structures of interest during a laparoscopic surgical procedure.

II. Previous Works

In general, AR technology in minimally invasive surgery may be used for training purposes, pre-operative planning and advanced visualization during the real procedure. Several research groups are exploring the use of AR in surgery and many image-guided surgery systems have been developed.

Devernay et al. propose the use of an endoscopic AR system for robotically assisted minimally invasive cardiac surgery [1].

Samset et al. present tools based on novel concepts in visualization, robotics and haptics providing tailored solutions for a range of clinical applications [2].

Bichlmeier et al. focus on handling the problem of misleading perception of depth and spatial layout in medical AR and present a new method for medical in-situ visualization [3].

Navab et al. introduce the concept of a laparoscopic virtual mirror: a virtual reflection plane within the live laparoscopic video, which is able to visualize a reflected side view of the organ and its interior [4], [5].

Kalkofen et al. carefully overlay synthetic data on top of the real world imagery by taking into account the information that is about to be occluded by augmentations as well as the visual complexity of the computer-generated augmentations added to the view [6].

De Paolis et al. present an Augmented Reality system that can guide the surgeon in the operating phase in order to prevent erroneous disruption of some organs during surgical procedures [7].

Soler et al. present the results of their research into the application of AR technology in laparoscopic. They have developed two kinds of AR software tools (Interactive Augmented Reality and Fully Automatic Augmented Reality) taking into account a predictive deformation of organs and tissues during the breathing cycle of the patient [8].

The collaboration between the MIT Artificial Intelligence Lab and the Surgical Planning Laboratory of Brigham led to the development of solutions that support the preoperative surgical planning and the intraoperative surgical guidance [9].

Papademetris et al. describe the integration of image analysis methods with a commercial image-guided navigation system for neurosurgery (the BrainLAB VectorVision Cranial System [10].

III. The 3D Models of Patient's Organs

In MIS, the use of images registered to the patient is a prerequisite for both the planning and guidance of such operations. From the medical image of a patient (MRI or CT), an efficient 3D reconstruction of his anatomy can be provided in order to improve the standard slice view by the visualization of the 3D models of the organs; colors associated to the different organs replace the grey levels in the medical images.

In our case study the 3D models of the patient's organs have been reconstructed using segmentation and classification algorithms provided by ITK-SNAP [11].

ITK-SNAP provides semi-automatic segmentation using active contour methods, as well as manual delineation and image navigation; it also fills a specific set of biomedical research needs.

In our case study, the slice thickness equal to 3 mm has caused some aliasing effects on the reconstructed 3D models that could lead to inaccuracies.

Therefore we have paid special attention during the smoothing of the reconstructed models in order to maintain a good correspondence with the real organs.

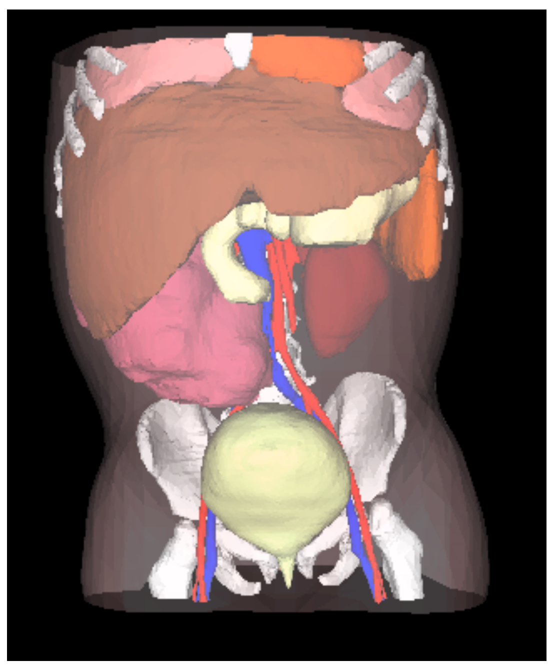


Figure 1. 3D model of the patient's organs.

 ISBN: 978-1-61208-117-5

By means of the user interface it is possible to display all the organs of the abdominal region or just some of these using the show/hide functionality; it is also possible to change the transparency of each organ.

The clinical case is a two-year-old child with a benign tumor of the right kidney.

Figure 1 shows the result of the image processing using ITK-SNAP; the skin and the muscles of the abdominal region are displayed in transparency and the tumor is shown in magenta.

I. The Developed Application

The developed application is supplied with a specific user interface that allows the user to take advantage of the feature offered by the software.

Starting from the models of the patient's organs, the surgeon can notice data about the patient, collect information about the pathology and the diagnosis, choose the most appropriate positions for the insertion of the trocars and overlap these points on the patient's body using the Augmented Reality technology.

In this way it is possible to use this platform for the pre-operative surgical planning and during the real surgical procedure too.

In addition, it could be used in order to describe the pathology, the surgical procedure and the associated risks to the child's parents, with the aim of obtaining informed consent for the surgical procedure [12].

In the developed application, as shown in Figure 2, all the patient's information (personal details, diseases, specific pathologies, diagnosis, medical images, 3D models of the organs, notes of the surgeon, etc.) are structured in a XML file associated to each patient.

A specific section for the pre-operative planning includes the visualization of the virtual organs and the physician can get some measurements on the organ or pathology and measure the distances.

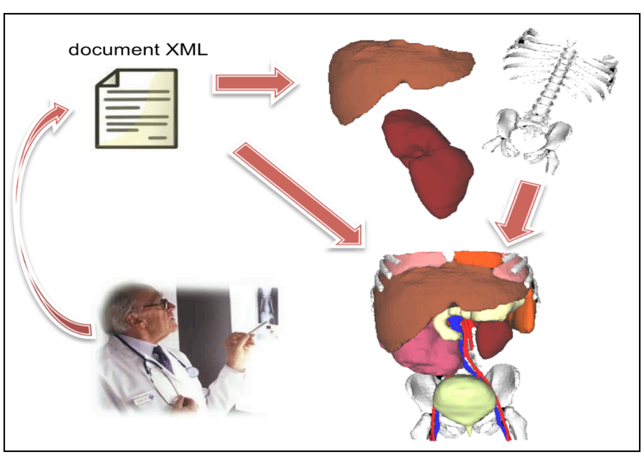


Figure 2. Patient's data collected in a XML file.

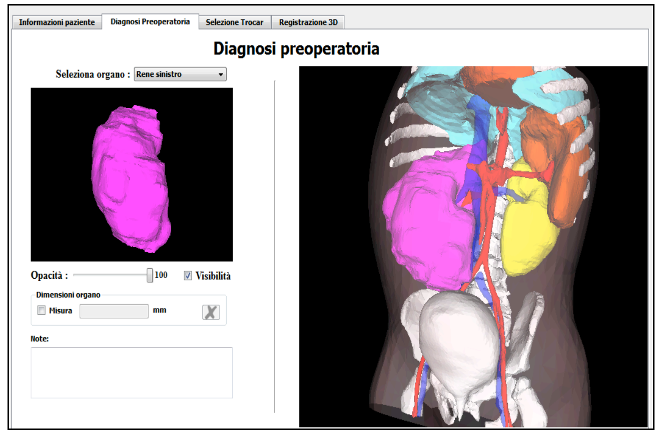


Figure 3. Section for the interaction with the organs.

Figure 3 shows the specific section of the user interface for the interaction with the 3D models of the patient's organs.

By means of a detailed view of the 3D model, the surgeon can choose the trocar entry points and check if, with this choice, the organs involved in the surgical procedure can be reached and if this is the correct choice in order to carry out the procedure in the best way.

Complications associated with initial abdominal entry are a prime concern for laparoscopic surgeons. In order to minimize first access-related complications in laparoscopy, several techniques and technologies have been introduced in the past years. The problem of blind access is that it may imply vascular injuries caused by the blind entry of instruments in the abdominal cavity.

A possibility to solve this problem may be the direct visualization of under-layer viscera and vessels.

Our application, by means of an Augmented Reality module, supports the placement of the trocars on the real patient during the surgery procedure and simulates the insertion of the trocars in the patient body in order to verify the correctness of the chosen insertion sites.

The Augmented Reality surgery guidance aims to combine a real view of the patient on the operating table with virtual renderings of structures that are not visible to the surgeon. In this application we use the AR technology in order to visualize on the patient's body the precise location of selected points on the virtual model of the patient.

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 fiducial points and an optical tracker is used. The tracker system consists of 2 IR cameras and uses a position sensor to detect infrared-emitting or retro-reflective markers affixed to a tool or object; based on the information received from the markers, the sensor is able to determine position and orientation of tools within a specific measurement volume.

 ISBN: 978-1-61208-117-5

Usually an optical tracker is already in the modern operating rooms and provides an important help to enhance the performance during the real surgical procedures.

Figure 4 shows the section for the accurate choice of the trocar insertion points.

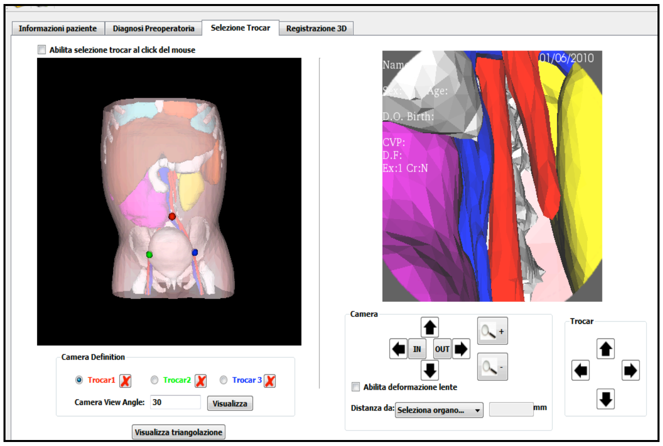


Figure 4. Section for the choice of the trocar insertion.

Using the augmented visualization, the chosen entry points for the trocars can be visualized on the patient's body through the Augmented Reality technique in order to support the physician in the real trocar insertion phase.

Figure 5 shows the augmented visualization with the trocar entry points overlapped on the patient's body.

Figure 5. The augmented visualization.

II. USABILITY TESTS

In order to evaluate the validity and the usability of the developed application and to receive possible suggestions from the users, some tests have been carried out. The test phase has been realized in order to allow the users to check all the functionalities of the application.

After a short period of training (5 minutes), the users have been tried to carry out different procedures and, subsequently, they have reported the impressions on a specific questionnaire. 15 subjects have been tested the application for an average time of 7 minutes and 43 seconds.

The obtained results can be considered satisfactory and some annotations to improve the user interface and the usability of the application have been considered.

III. CONCLUTIONS AND FUTURE WORK

The developed application offers a tool to visualize the 3D reconstructions of the patient's organs, obtained by the segmentation of a CT scan, and to simulate the placement of the trocars in order to verify the correctness of the insertion sites.

Furthermore the system retains patient and pathology information that the surgeon can insert and includes an Augmented Reality module that supports the placement of the trocars on the real patient during the surgery procedure.

An accurate integration of the virtual organs in the real scene is obtained by means of an appropriate registration phase based on fiducial points fixed onto the patient. In addition, a complete user interface allows a simple and efficient utilization of the developed application.

The platform can support the physician in the diagnosis step and in the preoperative planning when a laparoscopic approach will be followed. In addition, this support could lead to a better communication between physicians and patient's parents in order to obtain their informed consent.

The building of a complete Augmented Reality system that could help the surgeon during the other phases of the surgical procedure has been planned as future work; the acquisition in real time of a patient's video and the dynamically overlapping of the virtual organs to the real patient's body will be developed taking into account the surgeon point of view and the location of medical instrument.

An accurate AR visualization modality will be developed in order to provide a realistic depth sensation of the virtual organs in the real body.

Accuracy and usability tests will be also carried out.

REFERENCES

[1] F. Devernay, F. Mourgues, and E. Coste-Manière, "Towards Endoscopic Augmented Reality for Robotically Assisted Minimally Invasive Cardiac Surgery", IEEE International Workshop on Medical Imaging and Augmented Reality, 2001, pp. 16-20.

[2] 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.

[3] 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", Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07), 2007, pp. 129-138.

[4] N. Navab, M. Feuerstein, and C. Bichlmeier, "Laparoscopic Virtual Mirror - New Interaction Paradigm for Monitor Based Augmented

ISBN: 978-1-61208-117-5

Reality”, IEEE Virtual Reality Conference 2007 (VR 2007), Charlotte, North Carolina, USA, 2007, pp. 10-14.

[5] C. Bichlmeier, S.M. Heining, M. Rustaee, and N. Navab, “Laparoscopic Virtual Mirror for Understanding Vessel Structure: Evaluation Study by Twelve Surgeons”, 6th IEEE International Symposium on Mixed and Augmented Reality (ISMAR'07), Nara, Japan, 2007.

[6] D. Kalkofen, E. Mendez, and D. Schmalstieg, “Interactive Focus and Context Visualization in Augmented Reality”, 6th IEEE International Symposium on Mixed and Augmented Reality (ISMAR'07), Nara, Japan, 2007, pp. 191-200.

[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”, Joint Virtual Reality Conference of EGVE - ICAT - EuroVR, Lyon, France, 2009.

[8] L. Soler, S. Nicolau, J-B. Fasquel, V. Agnus, A. Charnoz, A. Hostettler, J. Moreau, C. Forest, D. Mutter, and J. Marescaux, “Virtual Reality and Augmented Reality Applied to Laparoscopic and NOTES Procedures”, IEEE 5th International Symposium on Biomedical Imaging: from Nano to Macro, 2008, pp. 1399-1402.

[9] W.E.L. Grimson, T. Lozano-Perez, W.M. Wells, G.J. Ettinger, S.J. White, and R. Kikinis, "An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization" Transactions on Medical Imaging, 1996.

[10] X. Papademetris, K.P. Vives, M. Di Stasio, L.H. Staib, M. Neff, S. Flossman, N. Frielinghaus, H. Zaveri, E.J. Novotny, H. Blumenfeld, R.T. Constable, H.P. Hetherington, R.B. Duckrow, S.S. Spencer, D.D. Spencer, and J.S. Duncan, “Development of a research interface for image guided intervention: Initial application to epilepsy neurosurgery”, International Symposium on Biomedical Imaging ISBI, 2006, pp. 490-493.

[11] P.A. Yushkevich, J. Piven, H. Cody, S. Ho, J.C. Gee and G. Gerig, “User-Guided Level Set Segmentation of Anatomical Structures with ITK-SNAP”, Insight Journal, Special Issue on ISC/NA-MIC/MICCAI Workshop on Open-Source Software, Nov 2005.

[12] E. Bollschweiler, J. Apitzsch., R. Obliers, A. Koerfer, S. P. Mönig, R. Metzger, and A. H. Hölscher, “Improving informed consent of surgical patients using a multimedia-based program? Results of a prospective randomized multicenter study of patients before cholecystectomy”, Ann Surg. 248, 2, Aug. 2008, pp. 205-211.

Interactive System for Medical Interventions Based on Magnetic Resonance Targeting

Sylvain Martel and Manuel Vonthron
NanoRobotics Laboratory, Dept, of Computer and Software Engineering, and institute of Biomedical Engineering
École Polytechnique de Montréal (EPM), Campus of the Université de Montréal
Montréal, Canada
e-mail: sylvain.martel@polymtl.ca

***Abstract*—Magnetic Resonance Targeting (MRT) is a new approach that aims at navigating untethered therapeutic or diagnostic micro-entities through the vascular network until a specific targeted location is reached. The platform used for such intervention is referred to as a Magnetic Resonance Navigation (MRN) system and it typically takes the form of a clinical Magnetic Resonance Imaging (MRI) scanner upgraded with special software and hardware modules to allow such platform to perform MRN in an effective manner. Considering the complexity of MRN operations and the speed at which real-time operations are being performed, an interactive system capable of providing a proper interface to allow an interventional radiologist or the like to properly interact and use such platform becomes an essential, yet a critical component for the success of this new medical interventional approach. Here, this paper presents for the first time an overview of a new interactive system for MRN operations suitable to conduct preliminary interventions.**

Keywords-magnetic; Magnetic resonance imaging; targeted interventions; cancer therapy; user interface

I. INTRODUCTION

Magnetic Nanoparticles (MNP) are often used to disrupt the high intensity magnetic field of a clinical Magnetic Resonance Imaging (MRI) scanner by creating a net field inhomogeneity that can be picked up and then processed using MRI techniques. Such MNP-based contrast agents are then typically used for imaging purposes such as to produce MR-images of blood vessels. Similarly, the same MNP when embedded in an untethered microscale entity can be used to track the displacement of the same micro-entity in the human body using a clinical MRI scanner. The field inhomogeneity created by these embedded MNP can then be used as a tracking signal that can then be fed back to an external computer as positioning data to be processed to perform closed-loop navigation or trajectory control to navigate such micro-entity along a path previously plotted in a portion of the vascular network that link the injection site to a pre-determined targeted location. Propelling and steering such magnetic micro-entities can be done by a technique referred to as Magnetic Resonance Propulsion (MRP) where magnetic gradients are generated to induce propelling force on the embedded MNP fully saturated by the high intensity homogeneous field of the MRI scanner, to propel and/or steer the micro-entities along a planned trajectory. Unlike the use of an external magnet, the technique is efficient even in deep locations in the body since it is independent of the distance between the source of magnetization and the object being navigated. Hence, MRN allows targeting efficacy at any depths in the human body and not just when operating close to the skin surface as it is the case when using a simple magnet positioned outside the patient.

MRN has been demonstrated and validated experimentally for the first time by navigating a 1.5 mm ferromagnetic bead in the carotid artery of a living swine (Martel *et al.*, 2007). As a first target medical application of MRN, cancer therapy is particularly suited. As such, special MR-navigable entities capable of carrying therapeutic agents for tumor chemoembolization and referred to as Therapeutic Magnetic Micro Carriers (TMMC) have been synthesized successfully (Pouponneau *et al.*, 2009). By navigating and targeting TMMC allows a significant reduction of the presence of highly toxic therapeutic agents in the systemic blood circulation. Such decrease in secondary toxicity reduces the negative health related effects for the patients while improving therapeutic efficacy with a lower dosage of drug. But for reaching a suitable embolization site deep enough in the vasculature to release the therapeutics sufficiently close to the tumor for improved therapeutic efficacy, the overall size of these TMMC must be reduced significantly to allow efficient MRN in smaller diameter vessels. Since the MR-based propelling force reduces significantly with a reduction of the effective quantity of MNP in the TMMC, the magnetic gradients required for adequate MRN operations must be increased beyond what is possible using the orthogonal imaging coils typically available in clinical MRI scanners. As an example, the initial version of TMMC has an overall diameter of approximately 40 micrometers (μm), requiring special propulsion coils capable of generating up to approximately 400 mT/m of 3D gradients compared to approximately 40 mT/m from the imaging coils of conventional clinical MRI scanners (Martel *et al.*, 2009). Hence, not only physiological data, but the characteristics of the MR-navigable micro-entities and the specifications (capabilities) of the MRN system must be taken into account when attempting to perform medical interventions requiring Magnetic Resonance Targeting (MRT), i.e. targeting using MRN. Hence, due to the level of complexity involved, an interactive system providing an interface between this technology and the medical specialists becomes highly desirable.

In Section II, our interactive display and an overview of the principles behind MRN operations will be described prior to describe the fundamental requirement for such interactive system. To better understand the problematic, a typical interactive flow for MRN operations will be given in Section IV. In Section V, the new interactive system still under development is described in more details before to conclude.

II. ACTUAL INTERACTIVE DISPLAY

An example of the actual computer interface dedicated to MRN operations is depicted in Fig. 1. First, an image of the blood vessel that will be used during MRN is first acquired using contrast agents. Then this image is filtered to remove any artifacts that could confuse MRN operations and it is then displayed as shown on the upper left corner of Fig. 1 showing in this particular case, the carotid artery of a living swine. Following this initial step, the interactive display allows a medical specialist to indicate to the computer system where the navigable micro-entities must pass to reach the desired target from the catheter release site. As such, the interactive display allows the user to plot waypoints over the image of the blood vessels to indicate the desired trajectory.

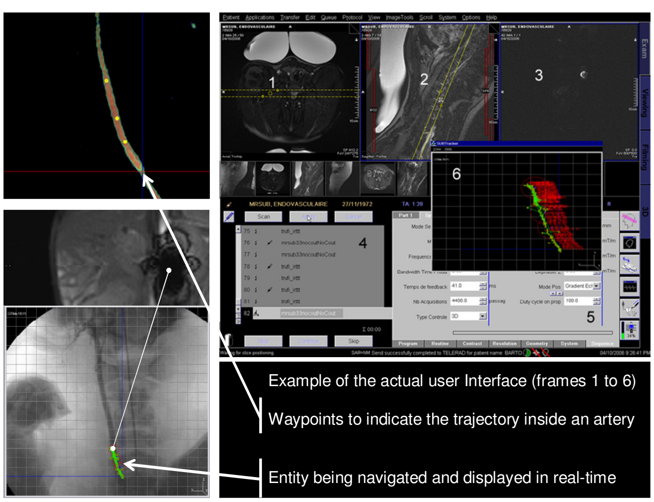


Fig. 1 – Preliminary computer interface used for MRN operations.

But there are many unknowns not indicated in this interactive system that shall be considered in order to guarantee safe and reliable MRN operations. The propulsive force induced to the magnetic micro-entity and generally described in Eq. 1 is one example.

$$\vec{F}_M = R \cdot V \left(\vec{M} \cdot \nabla\right) \vec{B} \cdot \quad (1)$$

Eq. 1 states that the induced magnetic force (N) produced by the magnetic gradients (T m^{-1}) generated by the MRN system and acting on a ferromagnetic core or an ensemble of MNP with an effective volume V (m^3), depends on the duty cycle R, i.e. the percentage of time per MRN closed-loop cycle being dedicated to propulsion (dimensionless), which can be reduced due to longer MRI-tracking acquisition time and/or in some particular cases to allow time for cooling of the coils used for MRP, and M being the volume magnetization of the core material (A m^{-1}). When such magnetization reaches a saturation level (M_{SAT}), the highest induced propulsion force per unit volume can be reached for a given soft magnetic material (being typically the case when placed in a high intensity magnetic field such as in the bore or tunnel of a standard 1.5T or higher field clinical MRI scanner).

As indicated in Eq. 1, the propulsive force induced on the embedded MNP and required to navigate the micro-entities taking into account physiological conditions, depends not only on the characteristics of the objects being navigated, but also on the specifications of the MRN system. As such, the interface depicted in Fig. 1 is not adequate to conduct MRN operations without a deep knowledge of all required physiological, navigable entities, and MRN system specifications. Frames 1 to 3 in Fig. 1 for instance are MR-images of the physiological environment where various measurements (e.g. angles of the blood vessels, width, etc.) were taken manually by engineers to evaluate the requirements and to adjust the MRN parameters for successful navigation. It becomes then obvious that the interactive system must ease the use of such technology by medical practitioners without the need for the participation of a well trained and knowledgeable engineering team. Indeed, in this example, engineers entered from the various measurements and a prior knowledge of the platform, the adequate parameters for safe and reliable MRN operations (as depicted in frame 6 and in the lower left corner in Fig. 1) in special windows such as the ones depicted in frames 4 and 5 in Fig. 1.

III. FUNDAMENTAL REQUIREMENTS FOR AN ADEQUATE MRN INTERACTIVE SYSTEM

A simplified diagram showing the fundamental architecture of a MRN interactive system is depicted in Fig. 2. The diagram shows three main categories of data required by the MRN core to perform adequate MRN operations.

The first category is the physiological data. These data describe the static and dynamic physiological information pertinent to MRN operations. These include mainly data information about the blood vessels (e.g. diameter, length, branches or bifurcations, angles, etc.) that will be used to reach the targeted location. Static data are extracted from the image of the blood vessels gathered and filtered as

 ISBN: 978-1-61208-117-5

described earlier and shown in the upper left corner of Fig. 1. Dynamic physiological data include blood flows and artifacts created by the movement of the patient such as during respiration. Indeed, MRN becomes more sensitive to the amplitude of body-related movements when operating in smaller diameter vessels where the distance between successive bifurcations becomes typically smaller than the amplitude of the movement of the patient. Hence, real-time image registration becomes essential. Since the acquisition of an MR-image is too long to be executed during real-time MRN operations, pre-acquired images taken at different physiological conditions that can be encountered during the intervention must be available to the MRN core for proper retrieval. Feedback information of the position of the micro-entities from known physiological features can be superposed on a selected pre-acquired multi-modality image and displayed by the interactive system. Other variables include heart rate that causes pulsatile flow that must also be taken into account. As such, pre-measured blood flow and real-time synchronization with the heart rate becomes additional data that when fed to the MRN core may improve MRT efficacy.

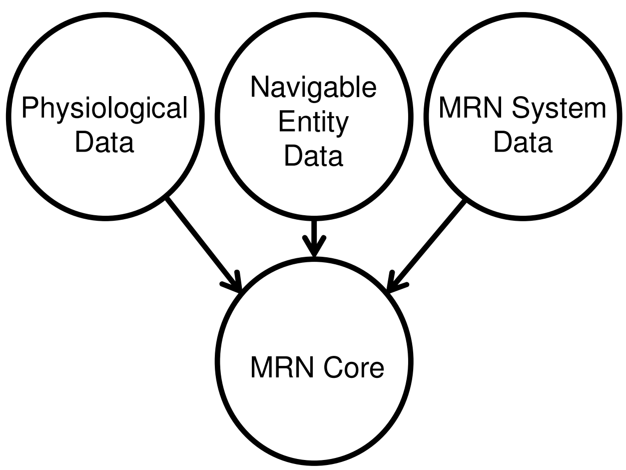


Fig. 2 – Simplified diagram of the architecture of the MRN interactive system

Data on the navigable entities are also critical for proper MRN operations. For instance, each type of navigable micro-entities may have different size, weight, and/or different volume of magnetic material with also different magnetization saturation that will lead to different displacement responses from the same applied directional gradients. The reasons are that synthesis with a different drug may pose constraints on the quantity of MNP being embedded, and/or one may opt for a larger dose of therapeutics per TMMC by lowering the quantity of MNP, etc. This is to say that the same therapeutic agent could be encapsulated in micro-carriers having different magnetophoretic velocities when subject to the same magnetic gradients. Hence, each type of TMMC in this particular case must be characterized and the data influencing MRN operations, stored in a database accessible to the MRN core.

Finally, several technical data of the MRN system and especially the ones related to MR-imaging/tracking and propulsion gradients that could impact MRN operations must be available to the MRN core.

IV. TYPICAL INTERACTIVE FLOW

The typical interactive flow required between the MRN system and the user is divided into three phases, namely the planning or MRN pre-operative phase, the MRN operative phase, and the MRN post-operative phase.

The first part of the pre-operative phase consists in gathering the required physiological data and to import it in the MRN interactive system. Then after some image and data processing, an image representation of the blood vessel network is displayed on the computer screen with the relevant physiological data. At this time, the MRN Interactive System (MRNIS) allows the user to perform several functions such as zooming, shifting, rotating, etc., to have a better perspective of the routing network available for MRT. Then on another window, the user is then allowed to enter MRT-related information starting by identifying the targeted region and the amount and type of therapeutic agents to be delivered if chemoembolization has been selected from the menu dedicated to the type of intervention to be performed. Then based on these information entered and the measured physiological data, the MRN core (Fig. 2) identifies one type of TMMC having the most appropriate characteristics from the ones available in the database. Then, the MRNIS displays the identification code of the selected TMMC with the dosage which would typically differ and be greater than the therapeutic dosage entered by the medical specialist. Once the respective TMMC dose has been put in a special automatic injection system, the MRNIS highlights the possible regions for catheter-release of the TMMC. Once the surgeon has selected the catheter-release site by clicking on a location within the region being highlighted, the MRN core calculates the best path that appears as a proposed trajectory being plotted along the image of the blood vessel being displayed. At this time, waypoints determined by the MRN core identifying directional changes required to achieve the requested MRT also appear on the computer display. Once the surgeon has approved the trajectory, the MRNIS is ready to proceed to the operative phase.

At this moment, once the operative phase has been initiated, the tip of the catheter is placed by the surgeon with the help of the MRNIS. Once the tip is positioned to the pre-approved release site, the injection phase is ready to be initiated. Prior to the injection, the MRNIS displays the number of injections with the respective doses and the time between successive injections for approval by the surgeon. In many cases, more than one injection may be required to

 ISBN: 978-1-61208-117-5

achieve the therapeutic dosage specified in the MRNIS by the medical specialist. This injection procedure is determined by the MRN core based on many parameters including but not limited to the size and the orientation of the long axis of the aggregates. More specifically, several TMMC will form an aggregate due to dipole-dipole interactive forces between neighbored TMMC. Each aggregate has a needle-like shape with the longer axes oriented toward z-axis, i.e. along the longitudinal axis of the MRI or MRN bore or tunnel where the patient has been placed. Depending on the orientation of the blood vessel's segment being navigated and the orientation of the next bifurcation with regards to the z-axis, a too large quantity of TMMC may cause problem such as pre-embolization or others, especially when the length of the aggregate is close to the diameter of the blood vessel being navigated and perpendicular or near perpendicular to the z-axis. In such case, the MRN core will tend to recommend several smaller injections instead of one or a few larger injections.

After each injection, the MRN core controls the generation of gradients and the whole MRN operations being performed in real-time. At this time, there is typically no interventions by the surgeons since the required rate of making decisions based on several parameters is beyond human's capabilities. The position of the TMMC is plotted and displayed in real-time by the MRNIS to be monitored by the medical specialist. Once the dose of therapeutic agents has reached the targeted chemoembolization site, the post-operative phase is initiated. In this phase, typical MRI interface is used by the interventional radiologist to image and to assess the success of the intervention.

V. New MRNIS Implementation

The interface shown in Fig. 1 has been used initially without any functionality aimed at assisting the user. During the preliminary tests with this interface, the user needed to have a high technical knowledge about not only the MRN platforms but also be able to make the right decisions for successful MRN operations. As such, the interventional radiologist was initially working with several engineers well familiarized with MRN-related issues. Although future work aimed at improving the look of the user interface to enhance and facilitate the interactions with the medical staffs, the actual interface takes the form of windows that can be displayed and moved to obtain the same interface shown in Fig. 1 with additional windows dedicated to assist the user in setting up the MRN platform for proper operations. The ultimate goal for this MRNIS is to ensure that the users will not need to enter complex technical data or specifications that require a deep understanding of MRN operations.

An important part of the software architecture concerns the "surrounding software environment". This includes all software components not embedded in the MRI system nor imaging, control or propulsion units: roadmap planning, 3D visualization tools or live feedback. In order to build such a user environment with a great extensibility and maintainability, we use well-established open-source tools and protocols. The proposed approach is to embed algorithms developed earlier inside the 3DSlicer software (Pieper *et al.*, 2006). This medical imaging tool is already used by other interventional imaging projects. It offers a great stability as well as excellent way to embed our own algorithm while providing numerous ready-to-use modules such as fiducials, OpenIGTLink (Tokuda *et al.,* 2009), or vessel segmentation support. Because of the modularity of this solution, we are able to add without modifications to the basis of the architecture, vessel segmentation or haptic interface to our system.

Another important step is the roadmap planning which is managed by the the fiducials module of the 3DSlicer software. The user can insert such points along the vessels during a similar procedure. In addition to the manual path planning protocol, vessel segmentation-assisted path planning is a very promising procedure. Indeed, 3DSlicer provides segmentation modules such as the Vascular Modeling Toolkit. This allows us to implement a multi-objective best path discovery algorithm that would allow an automatic computation of the best roadmap from the injection point to the drug release target. Such an algorithm takes into account parameters like vessel diameter, blood pressure, intersection complexity, steering power of the magnetic gradient with respect to the net magnetization of the device.

Another important aspect of the new MRNIS is live vizualization where the user is given real-time feedback of the operations being conducted. Indeed, the tracking algorithm can send the estimated position of the device to the host software, also through an OpenIGTLink connection. Our preliminary tests show no noticeable delay between the actual move of a device and the display on a 3DSlicer scene.

VI. Conclusion

A short description of an interactive system dedicated to MRN operations has been provided to show to the readers the importance of providing an adequate interface to allow the medical specialist to not only be able to use such system but also to be able to mange the complexity of this new technique. Indeed, the interactive system must be designed to allow the surgeon to enter specifications related to the medical protocol while being able to assess the progress of the intervention. Here, a basic interactive system have been described since more advanced operations making use of a combination of different types of navigable micro-entity for enhanced therapeutic and/or diagnostic target interventions could also be envisioned. Nonetheless, it is still difficult at this point to determine the best format for such interactive system since our group is the pioneer of MRN and the first one to attempt such interactive software. It is by observing practitioners and taking notes of theirs feedbacks that such interactive system coped with a proper user interface will be optimized.

ACKNOWLEDGMENT

This project is supported in part by the Canada Research Chair (CRC) in Micro/Nanosystem Development, Fabrication and Validation and grants from the National Sciences and Engineering Research Council of Canada (NSERC), the Province of Québec, and the Canada Foundation for Innovation (CFI).

REFERENCES

[1] S. Martel, J.-B. Mathieu, O. Felfoul, A. Chanu, É. Aboussouan., S. Tamaz, P. Pouponneau, G. Beaudoin, G. Soulez L'H. Yahia, and M. Mankiewicz, "Automatic navigation of an untethered device in the artery of a living animal using a conventional clinical magnetic resonance imaging system," Applied Physics Letters, vol. 90, 114105, March 12, 2007

[2] S. Martel, O. Felfoul, J.B. Mathieu, A. Chanu, S. Tamaz, M. Mohammadi, M. Mankiewicz, and N. Tabatabaei, "MRI-based nanorobotic platform for the control of magnetic nanoparticles and flagellated bacteria for target interventions in human capillaries," *International Journal of Robotics Research (IJRR)*, *Special Issue on Medical Robotics*, vol. 28, no. 9, pp. 1169-1182, Aug. 2009

[3] P. Pouponneau, J-C. Leroux, and S. Martel, "Magnetic nanoparticles encapsulated into biodegradable microparticles steered with an upgraded magnetic resonance imaging system for tumor chemoembolization," *Biomaterials*, 30(31), pp. 6327-6332, 2009

[4] S. Pieper, B. Lorensen, W. Schroeder, and R. Kikinis, "The na-mic kit: Itk, vtk, pipelines, grids and 3DSlicer as an open platform for the medical image computing community," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*, pp. 698–701, 2006

[5] J. Tokuda, G. Fischer, X. Papademetris, Z. Yaniv, L. Ibanez,. Cheng, H. Liu, J. Blevins, J. Arata, A. Golby, *et al.*, "OpenIGTLink: an open network protocol for image-guided therapy environment," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 5, no. 4, pp. 423–434, 2009

A Framework for Computer Based Training to In Vitro Fertilization (IVF) Techniques

A. F. Abate, M. Nappi, S. Ricciardi
Department of Mathematics and Computer Science
University of Salerno
Fisciano (SA) -Italy
{abate, mnappi, sricciardi}@unisa.it

***Abstract* - This paper presents a visual-haptic framework for the simulated training to some key procedures of the In Vitro Fertilization techniques which are become very popular to address several infertility conditions. Two of the most crucial procedures typically involved in the fertilization process, the Intra Cytoplasmic Sperm Injection (ICSI) and the Embryo Transfer (ET) are integrated in the system proposed. The aim is simulating them both at the visual and kinesthetic level by means of a specifically developed virtual environment. This environment includes the human egg, the selected sperm and the micro needles required during the ICSI as well as the catheter, the womb and the embryo involved in ET. The proposed approach exploits a two hand-based haptic devices mimicking the force feedback of the actual manipulation gear and a visual-haptic engine simulating the shape and the dynamic behavior of the main components involved in the two aforementioned stages of the artificial fertilization process.**

Keywords: visual-haptic interface; 3D object manipulation; virtual training

I. INTRODUCTION

Today, haptic devices providing realistic force feedback to the manipulation of virtual objects [1] allow the users of virtual simulators not only to practice at a visual level but also to develop the haptic-knowledge required to perform hand-based tasks [2]. Medical/surgical training applications [3] may particularly benefit from a visual-haptic approach, since they are inherently dependent on physical interaction [4] [5]. In this study the aforementioned interaction paradigm is exploited for the simulated training of two key techniques commonly required for In-Vitro Fertilization (IVF): the (Intra Cytoplasmic Sperm Injection) typically known as ICSI and the Embryo Transfer (ET) which are briefly explained in the following lines.

The term ICSI refers to the injection of a sperm directly into the cytoplasm of the egg. This procedure by-passes all the natural barriers that the sperm has to encounter. The ICSI procedure begins by drawing out the previously immobilized sperm into a tiny micro-needle, carefully maintaining it at its tip. The micro-injection needle is manipulated using a micro-manipulator which has extremely fine control capabilities. The egg itself is held onto another micro-tool by gentle suction to keep it firmly positioned. The micro-needle containing the sperm is pushed gently up against the outer shell (pellucida zone) and carefully injected through the shell, through the outer membrane of the egg and directly into the centre of the egg itself, i.e., the egg's cytoplasm (Figure 1a). At the end of the injection procedure the micro-injection needle is carefully withdrawn and suction on the egg is released.

After subsequent culture procedures, in case of fertilization, the Embryo Transfer (ET) procedure is performed by placing the embryos back in the uterus by means of a specific flexible catheter (Figure 1b), where they will hopefully implant and develop to result in a live birth. The ET procedure is a critically important procedure, and the physician can ruin everything with a carelessly performed embryo transfer. The entire IVF cycle depends on delicate placement of the embryos at the proper location near the middle of the endometrial cavity with as little trauma and manipulation as possible. To our best knowledge, this is the first proposal of an integrated ICSI/ET virtual training system, while there are only very few works addressing the ICSI simulation through virtual/augmented reality techniques. Banerjee et alii [6] propose a cellular micro-manipulation simulator based on the *Immersive Toucht*^{TM} VR system including a high-resolution display coupled with a haptic device providing force feedback during the simulated cell injection procedure, while the main limit reported about this approach is the lack of hand-eye coordination. Mizokami et al. [7] suggest a system to simulate the ICSI procedure by means of a Sensable's Phantom stylus-based haptic device, which is however limited to simulate only the interaction with the micro-needle manipulator. According to the embryologists involved in this research a useful training system should realistically simulate procedures which often involves both hands, therefore we decided to implement a two-hand based interaction approach to perform the tasks required.

II. SYSTEM'S ARCHITECTURE

Though the VR-related issues may seem to be prevailing in this proposal, the main challenges are represented by the two-hand interaction and by the realism of the visual-haptic perceptions to be provided during the simulated manipulation procedure. Indeed, for such a virtual training system to be effective and useful, the perceptual level of the simulation is more important than the exact agreement with the underlying physic laws. According to objectives mentioned in the introduction, a couple of CyberForce® hand-based force-feedback devices by Cyberglove Systems have been adopted in order to provide the user with haptic sensations while performing simulated ICSI and ET. The CyberForce device is made up by an articulated exoskeleton anchored to the back of the user's hand which is devoted to the recovery of grounded forces to the user's arm-hand-fingers system within its operative volume. The overall architecture of the Virtual-ICSI simulator is schematically shown in Figure 2, with its main components.

- *Human-Machine Interface*: receives as input the positional/rotational user's info from the haptic sub-system and outputs the user's activity data to the Physics Engine. The HMI enables to explicitly (by vocal commands) or implicitly (by hand activity) modify the simulation evolution.

- *Visual Rendering Engine*: integrates the tracking data and the current state of the physical simulation, transforming the polygonal geometry according to the camera viewpoint and rasterizing the scene in frames to be sent to the Head Mounted Display. It also checks for collision arising between the interacting objects, outputting a vector representation of any collision event, which is employed by the Haptic Rendering Engine which simulates contact forces. The VE is built on the Quest3D graphics programming environment [8] based on the DirectX API.

- *Haptic Rendering Engine*: is responsible for reproducing the haptic behaviour of all the objects involved in the interaction by means of specific haptic models. HRE depends on the Visual Rendering Engine and directly controls the Haptic Sub-System to exert contact and feedback forces. It outputs force-relevant data consisting in one force value for each finger plus the three-component force vector to be actuated by the hand back force transmission arm by applying a non-linear transfer function. The penetration of objects into other objects is therefore prevented by the combined action of the collision detection and the resulting force actuation by the force feedback device.

Contact forces are simulated by measuring them in terms of the depth of penetration of the virtual hand model into the grasped object.

- *Haptic Sub-System*: is made by left and right exoskeletons and it translates the output of the haptic rendering in terms of force feedback, also acting as an alternative input interface to select and activate the available functions.

- *Tracking System*: captures the user's hands position and orientation to enable coherent visualization by the visual engine.

- *Physical Engine*: simulates the dynamic behaviour of any object involved in the virtual simulation and represented in the 3D Dataset. The representation complies with an approximation of a subset of the physics laws appropriate to the simulation, by means of a set of physical parameters as mass, static/dynamic friction, and stiffness/elasticity. Rigid body dynamics is accomplished by means of the Newton Dynamics API [9] which is based on a deterministic solver instead of a more common solver based on linear complementary problem (LCP) or iterative methods, resulting in a more accurate and stable solutions. Soft body dynamics, which is required to realistically simulate the effect of the egg-needle and egg-pipette interaction, would be very compelling to render in real-time on a purely physical base, therefore it has been approached by pre-calculated 3D morphing.

- *Auxiliary Vocal Interface*: allows the user to control the system by (context dependent) vocal commands together to the haptic sub-system which represents the main user interface. This additional interface level is required since often during operations the user typically has both hands engaged in the manipulation.

- *3D Dataset*: is the source of every virtual content which is represented with polygonal geometry, textures, shadiness, physical properties and processed by the Visual Rendering Engine and the Physical Engine.

- *Head Mounted Display*: allows the user to experience an immersive simulation from a viewpoint resembling the microscope ocular.

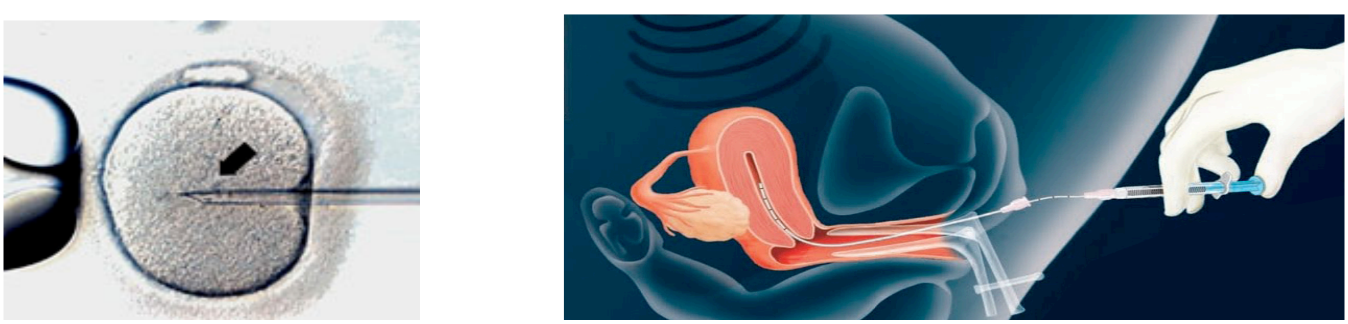


Figure 1. (a) The micro-needle penetrates the egg's zone-pellucida and reaches cytoplasm during ICSI. (b) A pictorial view of the embryo transfer procedure. The embryologist carefully insert the catheter through the cervix until the target site is reached and the fertilized embryos are gently released.

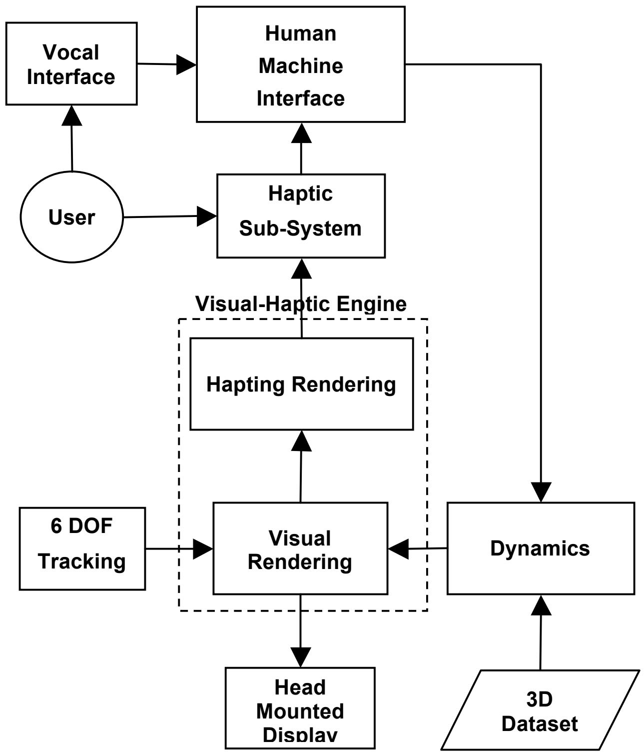


Figure 2. Schematic view of the system's architecture.

The egg model is based on concentric geodetic spheroids replicating the different cell's membranes, and its topology allows an ideal shape deformation when in touch with the micro-needle. The flexible catheter controlled by the user during the ET is approximated as a cinematic chain where each link's rotational values affect the previous links according to the distance in the chain and to a parametric decay function. The approach to render the contact between the catheter and soft organic tissues, like the cervix or the endometrium, exploits deformability/stiffness mapping. By means of this technique, texture mapping (typically simulating visual properties such as color, transparency, roughness, shininess, etc.) can be used to associate local deformability data to 3D geometry instead of relying on object-level properties.

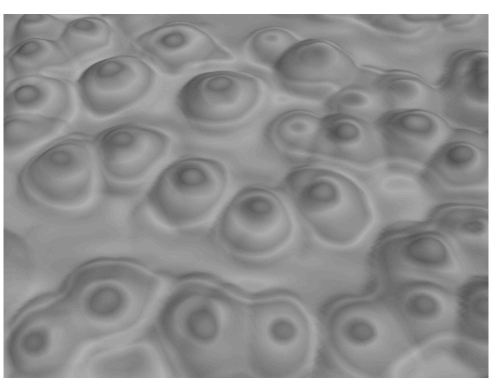


Figure 3. An example of deformability map representing the local stiffness of the simulated endometrium respectively by means of an 8 bit texture.

The deformability map is associated to mesh vertices through mapping coordinates in the form (u, v), previously projected onto the surface. The additional info can be represented through each pixel's RGB channels in a color texture or even in a grayscale bitmap, according to different arrangements offering a great flexibility of use (Figure 3). In its simplest form an 8 or 16 bit grayscale image may encode the local stiffness parameters required to compute the reaction force at a texel level, thus providing a range of 256 or 65536 stiffness levels with a spatial granularity only depending by image's resolution. A specific pixel shader processes the frames to provide a "ultrasound like" appearance to the rendered images reproducing the look of the diagnostic imagery to enhance the realism of the simulated intervention.

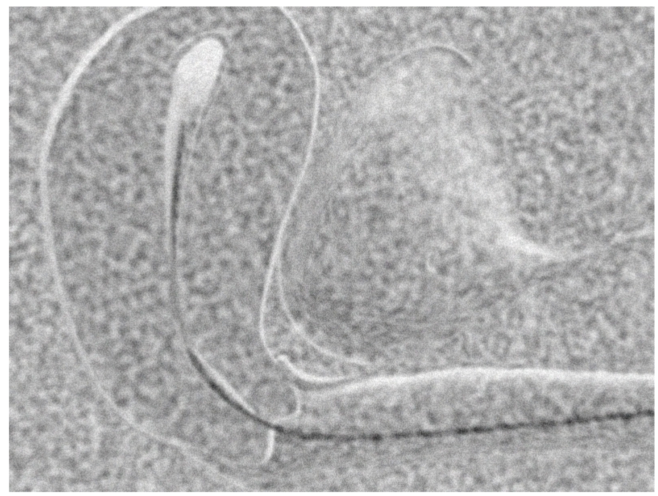


Figure 4. A rendering showing the simulated Embryo Transfer

III. FIRST EXPERIENCES WITH THE SYSTEM

We performed some preliminary experiments on the framework described above, to verify the subjective response of the expert and trainees embryologists to the virtual training. The test bed hardware included a dual quad-core Intel Xeon processor based on a Mac Pro workstation from Apple Inc., equipped with 8 Gigabytes of RAM and an Nvidia Quadro 5600 graphics board with 1,5 Gigabytes of VRAM. Five embryologists have been involved in the experimental sessions after a brief training on the usage of the HMD (a Cybermind Visette Pro SXGA) and of the haptic devices (Figure. 5). Each operator participated to 6 different sessions (3 for ICSI and 3 for ET) for a total of 30 sessions. After each session, each operator had to fill a questionnaire, assigning a vote in the integer range 1-10 (the higher the better) to seven subjective aspects of the simulated intervention and precisely: A. Realism of Visual Simulation; B. Realism of Haptic Perceptions; C. Accuracy of Simulated Manipulation; D. Visual-Haptic Coherence; E. HMD Alignment; F. Haptic System Fatigue; G. Simulator Usefulness.

 ISBN: 978-1-61208-117-5

A. Measures the overall visual realism of the simulation in terms of its training efficacy. This value is therefore affected not only by the graphics quality delivered by the system (the level of detail in the 3D anatomy, the frame rate, etc.) but also by how credible are the visual aspects of the dynamic simulation.

B. Measures the realism of the haptic sensations provided during the virtual experience. This value is therefore influenced by the limits of the haptic device in terms of force intensity and degree of freedom (for instance the Cyberforce cannot convey torque on wrist and arm joints) but also by the quality of the haptic-rendering algorithms adopted to simulate the contact forces during collision between solid and deformable bodies.

C. Measures the effectiveness of the visual-haptic manipulation, including grasping, releasing and exertion of forces on the virtual objects during the manual intervention.

E. Measures the subjective perception of spatial and temporal coherence between visual and haptic stimula provided by the system during the simulation. Therefore this value measures the quality of the perceptual illusion generated during simulation.

F. Measures comfort level experienced by users wearing the Head Mounted Display. The values reported in this experiment are clearly dependent on the particular HMD solution adopted, and on the device's technical specs, particularly the field of view, the resolution and the display's corner-to-corner sharpness, so they may considerably change if other devices are chosen. Negative issues related to the immersive stereoscopic visualization may have an impact on this value as well.

G. Measures the accretion of abilities and, consequently, the reduction of human stress related to a tricky medical scenario.

While these Figures are subjective and the number of users involved in these first trials is small, the overall evaluation has been positive so far.

Figure 5. Two exoskeletons during a simulated manipulation.

TABLE 1. Subjective evalutions resulting from the resume form

Features	Min	Avg.	Max
(A) Realism of Visual Simulation	6	7.1	9
(B) Realism of Haptic Perceptions	5	6.3	8
(C) Accuracy of Sim. Manipulation	6	7.2	8
(D) Visual-Haptic Coherence	6	6.8	7
(E) HMD Sickness	4	5.4	7
(F) Haptic System Fatigue	4	5.8	6
(G) Simulator Usefulness	6	7.3	8

IV. CONCLUSIONS

A framework for the visual-haptic simulation of In Vitro Fertilization procedures has been presented in this paper. The main visual and haptic aspects of the simulated procedure have been positively evaluated during our preliminary tests, while the main concern is related to the overall fatigue involved in wearing the articulated exoskeletons and the head mounted display. Anyway our work is at an early stage and we are in the process to set up and perform more polished and accurate experiments to measure the advantages and the limits of this framework for IVF practice.

REFERENCES

[1] M. A. Srinivasan and C. Basdogan, Haptics in virtual environments: Taxonomy, research status, and challenges, Computer and Graphics, Vol 21, Issue 4, , pp. 393-404.

[2] C. Krapichler, M. Haubner, A. Lösch, and K. Englmeier, (1997) "Human-Machine Interface for Medical Image Analysis and Visualization in Virtual Environments", IEEE conference on Acoustics, Speech and Signal Processing, ICASSP-97. Vol 4, pp. 21-24.

[3] K. F. Kaltenborn and O.Rienhoff, Virtual Reality in Medicine. Methods of information in medicine. Vol. 32, N 5, 1993, pp.407-417

[4] C. Basdogan, C. Ho, and M. A. Srinivasan, Virtual Environments for Medical Training: Graphical and Haptic Simulation of Laparoscopic Common Bile Duct Exploration. In IEEE/Asme Transactions On Mechatronics, Vol. 6, No. 3, September 2001, pp. 269-284.

[5] O. Körner and R. Männer, Implementation of a Haptic Interface for a Virtual Reality Simulator for Flexible Endoscopy. In Proceedings of 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'03), 2003, pp. 278-285.

[6] P. Banerjee, S. Rizzi, and C. Luciano, Virtual Reality and Haptic Interface for Cellular Injection Simulation. In Proceedings of 14th Medicine Meets Virtual Reality, JD Westwood et al, YOS Press, 2007, pp. 37-39.

[7] N. Abe, R. Mizokami, Y.Kinoshita, and S.He, Artificial Reality and Telexistence, 17th IEEE International Conference on, 28-30 Nov. 2007, pp 143-148.

[8] Quest3D visual development software: http://quest3d.com/

[9] Newton Game Dynamics API, (www.newtondynamics.com).

Real Time Drunkness Analysis Through Games Using Artificial Neural Networks

Audrey ROBINEL
LAMIA, Université Antilles-Guyane
Pointe-à-Pitre, France
audrey.robinel@univ-ag.fr

Didier PUZENAT
LAMIA, Université Antilles-Guyane
Pointe-à-Pitre, France
didier.puzenat@univ-ag.fr

***Abstract*—In this paper, we describe a blood alcohol content estimation prototype based on a comportment analysis performed by artificial neural networks. We asked to subjects that had drunk alcohol to play a video-game after having measured their blood alcohol content with a breathalyser. A racing game was modified so that it could provide various data related to the use of the controls by the player. Using the collected data, we trained our neural network in order to be able to determine whether or not the subject had exceeded a blood alcohol content threshold. We also succeeded in estimating this blood alcohol content with a mean error of 0.1g/l. We could perform those estimations independently of the track played among the two ones used. It was also performed in "real time", e.g., using only the data collected within the last 10 seconds of playing.**

***Keywords*-User interfaces; Games; Neural network applications; Cognitive sciences; Psychology; Human factors.**

I. INTRODUCTION

Driving under influence is not only forbidden, but also dangerous, being a major cause of accidents. Devices for measuring the blood alcohol concentration of a driver have thus been developed. The current and most common approach for a driver is to use a "breathalyser" [3]. Such a device measures the amount of alcohol contained in the exhausted air blown by the subject into the device by using a chemical reaction. The blood alcohol content is computed from this value. Law enforcement class devices are very reliable but expensive. Most of the low cost devices sold to regular citizen are unreliable and provide erroneous measures, and require regular calibration or parts replacement (we had a hard time finding a good device, which came at a higher price). Furthermore, if ever a good device is used, the driver must not forget to test himself before driving, which is very likely to happen for a drunk person.

But what if the car could estimate whether or not the driver is in condition for driving? Alcohol affects the driver and thus creates dangerous behavior on the road. This means that if we monitor the "low level" characteristics of the car controls used, we should be able to correlate the subject's blood alcohol content to his driving ability [9]. Our final goal is to create a natural interface embedded in the vehicle that would use the data collected by the car's computer to analyze the driver's behavior and actions to provide a diagnostic to the driver and warn him before he becomes really dangerous. For now, we present a prototype in order to demonstrate the feasibility of the concept. We will not use a real car but a video game, and gather data from the use of the game controls. In order to fulfill the objectives (detailed in the Section II), we had to instrument a game (Section III-A). We chose SuperTuxKart, a racing game for the reasons presented in the Section III-C. Following the measuring protocol described in Section III-E, we collected low level data (Section III-D) of 120 game runs. Using all those examples, and a K-fold cross-validation derived algorithm (Section IV-A), we measured the ability of the network to detect whether or not a subject's blood alcohol content exceed a threshold in Section V. We thereafter used the system to estimate the value of the blood alcohol content, in Section VI. We then showed how the system could perform this task independently of the track played, in Section VII. Subsequently, after defining what we call a "real time estimation" in Section VIII, we demonstrate the ability of our prototype to perform such an estimation in Section VIII. To conclude, we present in Section IX a global interpretation of the results and the perspectives opened.

II. OBJECTIVES

Our main goal is to demonstrate that using the data collected from an instrumented game played by a subject, a trained artificial neural network is able to determine whether or not a subject's blood alcohol content exceeds a fixed value, and even estimate the blood alcohol content value. Our secondary goal is to determine if the neural network is able to provide those estimations independently of the track or if it must be used on predefined tracks. Our last goal for the prototype is to perform in real time (we have a specific definition, presented in Section VIII), at any moment of the game run.

III. OUR SYSTEM

A. *"Instrumenting" a game*

We instrumented a video game that provided many low level measurements. By "Instrumenting" we mean taking measurements of player actions on input devices such as a joystick or a steering wheel. This devices response being progressive rather than binary, we can collect the state of the device and obtain a large set of "low level" data about the intensity of the subject's actions on the controls. We

 ISBN: 978-1-61208-117-5

gather these data continuously in time, which enables us to compute the evolution of some parameters over time.

B. Artificial neural networks

Playing a video game implies mobilizing many skills. Depending on the game, it may be more or less necessary to think, to act quickly, to act with precision, etc. Thus, instrumenting a game should give valuable information on the player. However, the point is to know what to measure and how to go from the measurements to a conclusion. The prototype presented in this paper uses artificial neural networks to select and combine relevant measures to assess the player, and is based on previous works [9]. Once trained on a representative population, a dedicated neural network will be able to evaluate a new player. Our approach eliminates the need to explicit the significance of measures for a given goal, therefore it becomes possible to make "low level measurements", e.g., the frequency of corrections on the steering wheel during a racing car simulation.

Among all artificial neural network models, for all our present developments, we have chosen a classical Multi-Layer Perceptron (MLP) with a back-propagation learning algorithm [2]. The neural network has been implemented with the free open source *Fast Artificial Neural Network Library* (FANN) [8]. FANN is a C library facilitating the use of the developed neural network within the instrumented video game (or interface, embedded computer, etc.), e.g., to perform a real-time evaluation.

C. The chosen game

We chose a racing game because of the dynamic aspect of such games. Since races are played on tracks, it implies a predefined path. The subject has then a restricted freedom, which turns the runs (played games) into reproducible and comparable tasks, enabling us to create an uniform examples set. Furthermore, it can be controlled with a steering wheel [7], which is an intuitive device for controlling a car as presented in Figure 1.

We selected SuperTuxKart [5] because it is extremely easy to handle. Anyone who could drive a car, even if unfamiliar with computers can perform the tests. It has very simple tracks, on which it is impossible to loose one's way (every subject were able to finish the races). Tanks to the open source license (GNU GPL), we could edit the code to include our data collecting library.

D. The collected data

For each full lap of the circuit, a vector of 5 components is preserved and will feed the artificial neural network. Those components are (i) the average number of steps per rotation (the steering wheel is analogical), (ii) the number of accidents (collisions and falls), (iii) the total number of actions of the steering wheel, and (iv) the number of changes in direction (e.g., the driver was turning left and is now turning right). We define a "run" as a session played by a subject on a full track lap. When a subject plays a run, the system generates XML files containing all the race measurements. Each run then provides us an example (in the real time context, each run will provides us with multiple examples, more on that in Section VIII) that can be used for our problematics. Each subject does multiple runs, and the combination of all runs is our full examples set. For each run, we also measured the subject's blood alcohol content with a breathalyser.

Figure 1. The prototype during a run

E. The measuring and data collection protocol

In order to obtain a coherent example set, we used the same protocol for each subject. We had selected two tracks ("skyline" and "snowmountain", see Figure 2) for our data collection campaign, and every subject played on both tracks. Each subject played being sober on skyline, then on snowmountain, so that he could get used to the game controls. We then repeated it, still with a 0 g/l blood alcohol content, which provided us sober playing examples. Then the subject drank some alcohol, and we waited 20 minutes. After this delay, the subject's blood alcohol content were measured with a breathalyser (model AL7000 from Alcopass [1]). Two other games were then played (same tracks). The operation were repeated 15 to 20 minutes later, and for some subjects another time 15 to 20 minutes later again. This provided us with 8 to 10 examples per subject, two with a blood alcohol content of 0 g/l, and the other ones with various values greater than 0 g/l and lesser than 1 g/l.

At last, we obtained a 120 examples set containing game statistics distributed between 14 different subjects (in order to avoid subject-specific results) with various blood alcohol content, between 0 and 1 g/l. We ensured that, after the experiment, no subject drove a real car while having a blood alcohol content superior to 0 g/l.

Figure 2. The skyline track on the left, snowmountain on the right.

IV. Experimental protocol

Before presenting our results, we will explain the methodology used to obtain the numbers that will be showed later. Neural networks are trained on a learning set and then used in generalization, on examples that are not in the learning set. Generalization can be done either in production, or in order to test the network as presented in Section IV-A. We used a method derived from K-fold cross validation [6] to evaluate the generalization performance of the network. This provided the success rate and the average error of the network for each problematic.

A. The K-fold cross-validation derived algorithm

Our algorithm was the following : at start, we take an example base (in this example, it has 100 examples). We initialize the algorithm by setting a loop index `i`, variable `s` (successes) and variable `f` (failures) to 0. We then split our example base in two: examples 0, 1 and 2, constitute the generalization set and examples from 3 to 99 form the learning set. We train the network on the learning set, and test it on the generalization set. We evaluate the network's answer and increment `s` or `f` in case of success or failure. We then increment `i`, and start again the same with generalization set containing examples 1, 2 and 3, and learning set containing examples 4 to 99 and example 0. We keep doing that until `i` reaches 99 (in that case, the generalization set is constitute of examples 99, 0 and 1, while learning set is the rest). Thus, on each pass, we trained the network on a subset of examples, and tested it on another disjoint subset. Therefore, the network was *never* tested in generalization on any example used in the learning set. We obtain in the end the total number of successes and failures. We can then compute the success rate of the neural network as accurately as possible.

B. Determination of right and wrong network answers

We will now explain how we decide when the neural network response is right or wrong. When the network returns a value for an example, we can compare it to the real value, e.g., the estimated blood alcohol content and the measured value. We define ϵ as the distance between the expected value v and the neural network output o: $\epsilon = \|v - o\|$. We define the total cumulated error as the sum of all ϵ obtained for each example. We then obtain the average error by dividing this sum by the total number of examples tested.

We will set a maximal tolerated value (ϵ_{max}) for ϵ: if ϵ exceed ϵ_{max}, the network answer will be considered wrong (a failure). It will be right (a success) otherwise. A value of ϵ_{max} is defined for each experiment.

From now on, when we write that the success rate in generalization is x% for a value of epsilon, it means that for x% of the tested examples, the error of the network were less than ϵ_{max}. We always give the success rate in generalization, and never the learning rate.

	(i) skyline	(ii) snow	(iii) both tracks
successes	129	121	233
failures	3	5	25
success rate (%)	97.73	96.03	90.31
mean error (g/l)	0.0587	0.0972	0.1873

Table I
THRESHOLD EXCEEDING DETERMINATION

V. Determination of threshold exceeding

We first tried to determine if the neural network is able to detect when a subject's blood alcohol content exceeds a fixed threshold of 0.4 g/l. When the subject's measured blood alcohol content was over 0.4 g/l, the network was expected to return a value of 1 on the output, and a value of 0 otherwise.

We present in Table I the success rate and mean error of the network when using an example set of runs played on (i) skyline, (ii) snowmountain and (iii) both tracks.

This first experiment is the easiest for the network, and offers the best success rates, over 95% of good determinations. The results are similar for skyline and snowmountain, while they slightly decrease when we use both tracks in the same base (it will be studied further in section VII).

VI. Estimation of blood alcohol content

This experiment was conducted in order to see if the neural network could give an estimation of the subject's blood alcohol content. The network is expected to return (as output value) the subject's measured blood alcohol content, between 0 g/l and 1 g/l.

We note in Tables II and III lower success rates than in the previous experiments (especially for lower values of ϵ). However, this was expected. As we try to return an accurate value, we define a lower tolerated error (ϵ_{max}). This indeed decreases the success rate.

 ISBN: 978-1-61208-117-5

ϵ	successes	failures	success rate
0.10	85	47	64.39%
0.12	102	30	77.27%
0.14	108	24	81.82%
0.15	112	20	84.85%
0.16	114	16	87.88%
0.18	119	13	90.15%
0.20	122	10	92.42%

Table II
BLOOD ALCOHOL CONTENT ESTIMATION ON SKYLINE, IN FUNCTION OF ϵ. THE AVERAGE ERROR IS 0.083860 G/L.

ϵ	successes	failures	success rate
0.10	70	53	56.91%
0.12	71	52	57.72%
0.14	78	45	63.41%
0.15	81	42	65.85%
0.16	84	39	68.29%
0.18	94	29	76.42%
0.20	98	25	79.67%

Table III
BLOOD ALCOHOL CONTENT ESTIMATION ON SNOWMOUNTAIN, IN FUNCTION OF ϵ. THE AVERAGE ERROR IS 0.1158 G/L.

VII. TRACK INDEPENDENT ESTIMATIONS

In order to demonstrate that the system is able to estimate the blood alcohol content of the subject independently of the race track played, we conducted two other experiments. The first one, in Section VII-A, focuses on a network trained on a base containing examples from both tracks. The second one, presented in Section VII-B, shows how the network behaves when trained on a learning base containing examples from one track and is tested on a generalization base made of examples from another track.

A. Testing the network with examples from both circuits

We will start with the mixed tracks data sets. The example set includes data gathered from game runs played on the two tracks, skyline and snowmountain. Using K-fold cross validation, we measured the ability of the network do estimate the blood alcohol content of the subjects after being trained on such a base.

The results presented in the Table IV does not show a significant changes in neither the success rate nor the mean error of the network. This tends to indicate that the prototype is able to estimate the blood alcohol content without the need of a specific portion of road, implying that we could construct examples set from various road sections or tracks.

variable	value
successes	77
failures	22
success rate (%)	77.78
mean error (g/l)	0.0587

Table IV
RESULTS OF THE MIXED EXAMPLE SET FOR ϵ=0.20

values of epsilon	A : sky → snow	B : snow → sky
Success rate for ϵ=0.15	26.47%	40.62%
Success rate for ϵ=0.20	82.35%	62.62%
Success rate for ϵ=0.25	88.24%	71.88%
Average error (g/l)	0.185407	0.226358

Table V
TRACK INDEPENDENCE TEST RESULTS.

B. Training the network with examples from one track, and testing it with examples from another track

After this, we trained the network on the examples from the first track, and then checked it's generalization results on the examples of the second track. This configuration ensures that the second set of examples (games played on the second track) is unknown to the network.

This time, we created two examples sets instead of one: (i) the "sky set" contains all the game runs that were played on the "skyline" track; (ii) the "snow set" contains all the game runs that were played on the track "snowmountain". We will try two configurations: in the first one (A) we will use the sky set as training set and snow set as generalization set. In the second one (B), the the snow set as training and sky set as the generalization set. The results obtained for different values of ϵ are presented in Table V. On contrary of other experiments, we did not use the k-validation algorithm, since we had two distinct sets.

If we consider low values of ϵ, the success rate is quite low. We had to increase ϵ in order to maintain a success rate comparable to the previous ones. The obtained ϵ reaches really important max values, with in the worst case 0.25 g/l. But despite the fact that the mean error also increased, the network could still perform valuable estimations. This tends to indicate the ability of the prototype to be used to estimate the blood alcohol content independently of the track. Furthermore, this shows how our neural network is able to cope with data gathered on an unknown portion of track. This avoids the need to either create an exhaustive learning set containing all possible situations or to use the system in the same environment that were used for training. Of course it is far from being perfect as, in these conditions, the prototype is less accurate with much lower success rates in some cases (experiment B, for example) and significantly higher average errors (reaching 0.22 g/l for experiment B). But the prototype tends to demonstrate that it is able to generalize from one track to another.

Of course, we could have more meaningful results with more tracks, but the game did not provided enough tracks that met our requirements, detailed in section III-C.

 ISBN: 978-1-61208-117-5

VIII. Real time estimations

We will here demonstrate the ability of the system to perform "real time" estimations of the measured blood alcohol content.

A. Definition of "real time estimation"

We define as "real time" an estimation done using the data gathered within the last elapsed seconds. We define n as the length in seconds of the interval of play used. An n seconds real time estimation means that the estimation is performed using the data of n seconds of race instead of using the full length of the race. The interval of play used for this determination will be called a "window" and n is defined as the size of the window.

In order to obtain comparable data, we normalize the variables by dividing their values by n. By this mean we also ensure that the trained network can estimate the blood alcohol content for various values of n, e.g., it can be used indifferently to provide a real time estimation based on 10 or 20 seconds.

We kept using the examples from both tracks for our real time experiments.

B. Protocol

We checked here if the network is able to perform a real time determination based on 10, 15, 20, 30, 40 and 50 s windows. In order to study the impact of n on the results, we kept all other parameters equal. In order to do that, when cutting the runs in order to obtain n seconds examples, we only kept the first example, so that the number of generated examples remained constant. We thus used data from the interval $[0, n]$ in order to perform the n seconds real time estimation.

C. Impact of n on the success rate and the mean error

We present in the Figure 3 the evolution of the success rate (as always in generalization) and in the Figure 4 the variation of the mean error in function of the value of n. The success rate increases with the length of the interval, while the mean error decreases. Increasing the window size improves the quality of the examples, which indeed enhances the results.

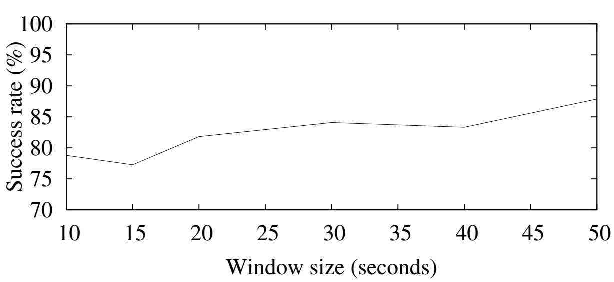


Figure 3. Evolution of the success rate in function of the window size

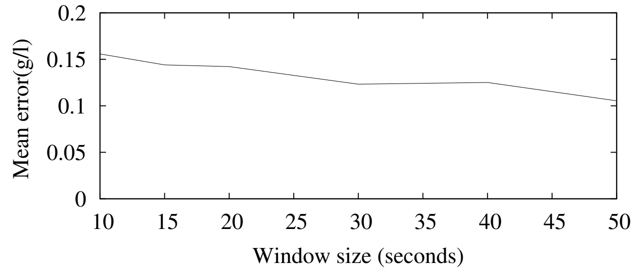


Figure 4. Network mean error as a function of the window size

	successes	failures	success rate	mean error
10s	104	28	78.79	0.155820
15s	102	30	77.27	0.143969
20s	108	24	81.82	0.142155
30s	111	21	84.09	0.123291
40s	110	22	83.33	0.125081
50s	116	16	87.88	0.105408

Table VI
Real time blood alcohol content estimation on a mixed set (containing both tracks), for a value of ϵ of 0.2 for different values of the window size (n).

D. Other real time results using all the available windows

The previously presented results focus on the impact of the window size on the success rate and mean error. Furthermore, we obtained similar results when using all the available windows. Indeed, when we split our 60 seconds long runs in windows of 20 seconds, we obtain 3 examples. In the previous experiments we kept only the first one. Keeping all the obtained examples gave similar results. This also confirms the track independent estimations capability of the network, since the windows [0-20], [20-40] and [40-60], as an example, are 3 different sections of the track. Indeed, only one lap is performed. Thus, the subject never drives two times on the same section during a run.
We also obtained comparable results for the threshold exceeding detection (with better success rates), which we do not present here as they are quite similar (slightly better) than the ones detailed earlier in Section VIII-C.

IX. Conclusion

We have demonstrated that it is possible not only to detect threshold exceeding but also to estimate the blood alcohol content of a subject. These estimations were achieved independently of the track. Considering that it is hardly feasible to create an exhaustive learning base containing all possible situations, it makes possible to only fill the learning base with the most common situation patterns. We then rely on the generalization capability of the neural network to provide estimations on new patterns. We also managed to obtain a real time estimation, and still independently of the track played. The prototype is able to provide an estimation using a 10 seconds window (e.g., after only 10 seconds of playing), and to improve the reliability and the accuracy of the estimation by using a larger window. It should thus be possible to conceive an embedded interface that would be able to provide estimations on the driver after only a short while, and enhance these estimations after more time. The fact that the estimations are not bound to a fixed length of time allows many configurations.

Furthermore, such a system could re-evaluate the driver continuously, and adapt the estimations to a changing situation. Indeed, someone who just drunk some alcohol may feel perfectly able to drive, but as alcohol is absorbed in blood, the side effects will alter his driving skills up to the point where driving becomes dangerous. An embedded device based on our concept could detect the evolution of the driver and warn him before he becomes really dangerous.

X. Limits and drawbacks

While we could solve the problematics, we had to degrade the accuracy of the system (ϵ) in some case in order to maintain acceptable success rates. As for track independence, we note in some cases a significant drop on success rate and an increase of the average error. But this experiment was performed to see how the network would cope with unknown tracks. Considering that we only had two tracks, only one remained for training. As shown, if more accuracy is required, we can use several tracks. Furthermore, if this system were to be used in a production environment, we would try to create a base containing a representative set of the situations that can be encountered.

We also noted the much higher average error for real time estimations when using small windows such as 10 s, but it was expected. It indeed was intended to show the neural network behavior in an extreme case, setting the lower bound of the window size.

In some cases we reached an accuracy of 0.11 g/l, which is close to the 0.1 g/l of the breathalyser. Owing to the fact that we calibrated our system with this breathalyser, improving the accuracy can only be achieved by using a better breathalyser. On some cases though, the accuracy were quite far from the breathalyser's. As the present prototype was created as a proof of concept, it was not intended to be on par the breathalyser in every experiment.

XI. Perspectives

We demonstrated the ability of the prototype to perform meaningful estimations even in a simple environment. Those results will be the base of our next works: we plan to improve and test the system on a more realistic environment in so that it might be embedded in a vehicle later on. The prototype demonstrated the viability of the concept in a simplified environment, we will now take it to a more complex virtual world before trying to use it in the real world. For now, much work is left in order to achieve this goal: we will have to find what data to gather from a more realistic system, or among the numerous data monitored by modern cars, and also a way to accurately measure those data. Considering the amount of completely different situations in a real driving environment (the "game play" of SuperTuxKart is very simplified), we will have to use either much more inputs, or other sorts of neural networks such as mixture of neural experts [4] to cut down the problematics in smaller and more simple ones, considering how different it is to drive on the highway and in congested urban areas.

No matter what neural network model we will use, it will anyway imply much more work to select meaningful inputs for the network, and probably require the use of parallel computing in order to explore configurations until we obtain the optimal ones for a given problem. Hopefully, generalization will always be possible on a low cost, low power consumption, and low computing power system, making the software easily usable in embedded devices.

Furthermore, we foresee that increasing the complexity of the environment and the amount of measured characteristics might enable us to perform not only more accurate estimations, but also many other estimations, such as detection of tiredness, attention drop, or the use of driving impairing drugs or medications. Those very similar problematics may use the same system with just another examples set.

References

[1] Alcopass. Alcopass AL7000 breathalyzer documentation., December 2010. http://www.alcopass.com/.

[2] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.

[3] E. Bogen. Drunkenness, a quantitative study of acute alcoholic intoxication. *California and Western Medecine*, XXVI(6):778–783, 1927.

[4] P. Estvez, H. Paugam-Moisy, D. Puzenat, and M. Ugarte. A scalable parallel algorithm for training a hierarchical mixture of neural experts. *Parallel Computing*, 28(6):861–891, 2002.

[5] J. Henrichs, M. Gagnon, and C. Pelikan. SuperTuxKart, Free 3D kart racing game., December 2010. http://supertuxkart.sourceforge.net.

[6] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection, 1995.

[7] Logitech. Formula Vibration Feedback Wheel., December 2010. http://www.logitech.com/en-in/441/298.

[8] S. Nissen. Implementation of a fast artificial neural network library (FANN). Technical report, Department of Computer Science University of Copenhagen, October 2003.

[9] D. Puzenat and I. Verlut. Behavior analysis through games using artificial neural networks. In *Third International Conferences on Advances in Computer-Human Interactions*, pages 134–138, February 2010.

Tactile Sensing for Safe Physical Human-Robot Interaction

Norbert Elkmann, Markus Fritzsche, Erik Schulenburg
Business Unit Robotic Systems
Fraunhofer Institute for Factory Operation and Automation
Magdeburg, Germany
e-mail: norbert.elkmann@iff.fraunhofer.de
markus.fritzsche@iff.fraunhofer.de
erik.schulenburg@iff.fraunhofer.de

***Abstract*— Human-robot interaction in a shared workspace permits and often even requires physical contact between humans and robots. A key technology for a safe physical human-robot interaction is the monitoring of contact forces by providing the robot with a tactile sensor as an artificial skin. This paper presents a pressure-sensitive skin for the mobile assistant robot LiSA (Life Science Assistant). It can be adapted to complex geometries and it can reliably measure contact on the entire robot body. The sensitive skin is equipped with integrated cushioning elements reducing the risk of dangerous injuries in physical human-robot interaction. Besides its safety function, the sensitive skin offers touch-based robot motion control that simplifies human-robot interaction. In the paper we describe the sensor setup and the hardware implementation on the mobile assistant robot LiSA and explain the strategies for a safe human-robot interaction. Beyond that we describe the algorithms enabling direct touch-based robot control and present some fundamental results from our evaluation experiments.**

Keywords- artificial skin; human-robot interaction; mobile robot; tactile sensor

I. INTRODUCTION

Long banished behind fences and safeguards, robots are now increasingly capturing new fields of application as service robots or assistance systems. New strategies in human-robot collaboration (HRC) erase the boundaries between workspaces. Safety and protection of humans are of utmost importance when humans and robots collaborate directly or come into contact in human-robot coworker scenarios.

In such scenarios, the assurance of safe physical human-robot interaction will rely on the results of a detailed risk assessment, identifying the various hazards and their association with physical quantities of robot motion. In general, it is essential that we can reliably limit the robot's position and speed as well as its static force and its potential impact force. In case of an impact, we must also consider the post-impact behavior.

State-of-the-art robot control systems can safely limit a robot's speed and position, whereas safety-rated limitation of forces is currently not a characteristic of industrial robots. Thus, we propose to integrate an artificial skin into the overall safety concept for collaborative robots in order to provide safety-rated information on contact and on static forces.

During the last 30 years various sensitive skins based on tactile sensors and using diverse approaches have been reported [1-3]. Most of them form an array of individual pressure sensors. Covering large parts of a robot body poses several significant engineering challenges.

The tactile skin must fulfill certain design engineering requirements: It must be pliable to adapt to curved robot bodies, tough and dependable to withstand a significant number of contact cycles, energy-absorbing to soften collisions and, finally, easy to manufacture. In addition, there are also fundamental requirements for the sensor system's reliability, since the tactile skin represents the only barrier between the robot's forces and a human in direct physical human-robot interaction.

Various attempts have been made to tackle these issues and to develop robot skin systems in recent years. Iwata et al. [4] used rigid covers on parts of their humanoid robot WENDY. The covers are mechanically secured at a single point on a multi-axis force sensor. This allows accurate measurement of forces and torques acting on the cover. The system has limited "multi-touch capability". When several forces act on a cover at the same time, the associated multi-axis force sensor only measures one force generated by all the force vectors applied.

The wiring topology is one of the most challenging problems when implementing a distributed tactile sensor system. The larger the number of sensing elements, the thicker the wire bundle and the larger grows the amount of data. To solve this problem, piezo-resistive sensor patches with embedded data processing electronics were effectively implemented in the ARMAR-III robot [5]. Embedded electronics process local tactile data in order to limit the bandwidth requirements. The sensor patches are custom-designed to cover the respective parts of the robotic arms.

Ohmura et al. [6] developed a solution to adapt tactile sensor systems to the curved surfaces of robots. Their approach also applies a networked architecture to connect a large number of individual controllers, each scanning a limited number of taxels. The electronics and transducers are embedded in a tree-shaped flex/semi-flex PCB support, which simplifies mechanical adaptation to curved surfaces.

 ISBN: 978-1-61208-117-5

Canata et al. [7] presented an alternative approach to cover complex geometries. Inspired by the principle of triangulation used in computer graphics, they applied a mesh of triangular shaped sensor modules to cover three-dimensional surfaces. Each sensor module is supported by a flexible substrate, thus allowing the sensor to conform to smooth curved surfaces. Three communication ports placed along the sides of each sensor module allow interconnecting adjacent modules.

The aforementioned solutions have already been implemented to solve a multitude of problems. However, a tactile sensor system implemented in direct human-robot interaction must not only be able to detect contact but also to cushion it if necessary.

Furthermore, the sensor system must be able to absorb the stopping distance when the robot stops without applying high forces to the collision partners.

The tactile sensors we analyzed prior to our research and development work are only conditionally suited for this application, since they lack the necessary softness and compliance.

To solve these problems we implemented an energy-absorbing layer into the tactile sensor solution applied to our mobile robot LiSA. With this additional layer we are not only able to measure interaction forces and represent them spatially and quantitatively resolved, moreover we are able to cushion collisions. The risk of dangerous injuries in human-robot interaction can be significantly decreased by this measure.

The setup and implementation of our tactile sensor solution on the LiSA robot will be described in the following sections.

II. Sensor Technology

This section deals with the setup of our tactile transducer. We will discuss adaptations to complex geometries and challenging ambient conditions. Finally, the data processing unit will be described.

A. Tactile Transducer Technology

The heart of our tactile transducer technology is a flexible measuring sensor (Figure 1), which is about 2 mm thick and entirely made of textile to obtain maximum mechanical reliability.

Instead of classic cables, textile conductive paths span a sensor matrix which consists of flexible sensor cells. The individual sensor cells are based on variable, pressure-dependent resistors and have a defined value in their unloaded state. Deviations from this value are a measure of the force acting on the sensor. Weiss et al. [8] describes the working principle of such sensor cells.

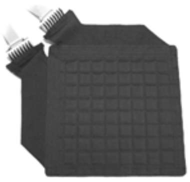


Figure 1. Prototype 8x8 tactile transducer

Since each sensor cell provides an analyzable signal even in the unloaded state, the functional capability of the individual sensor cells can be monitored. Thus it is possible to detect any failure of the individual sensor cells when the textile conductive paths short-circuit or are cut (Figure 2). The thusly achievable “intrinsic safety” of the sensor system is an essential basis for using the sensor system as a safety sensor.

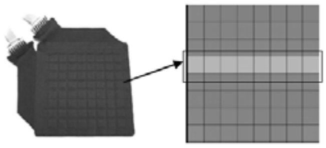


Figure 2. Sensor with related sensor data, showing a malfunction on a single sensor line (light gray) and normal sensor data (dark gray)

For an optimal interaction between the robotic system’s geometry and the artificial skin, the tactile transducers should be adapted to the individual case of application, e.g., by customizing one or more of the following options:

- shape and size of the tactile transducers
- shape and size of the individual sensor cells
- force measurement range
- thickness
- shell material

The structure of the textile sensor system allows for integrating application-specific cushioning zones (Figure 3). They consist of special energy-absorbing materials and therefore enable controlled deceleration and stopping of the robot system in case of an unintentional contact without exposing the collision partner to high load peaks. These cushioning zones include significant design factors such as the robot system’s speed and geometry and the safety circuit’s response time.

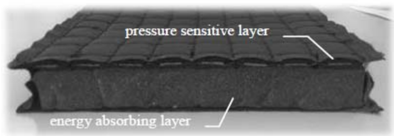


Figure 3. Cross section of a tactile transducer, showing pressure-sensitive layer and energy-absorbing layer

The tactile transducers may be constructed for a wide range of ambient conditions by selecting an appropriate shell material from waterproof, breathable or particularly rugged designs (Figure 4). Depending on its properties, the shell material is sewn, welded or bonded.

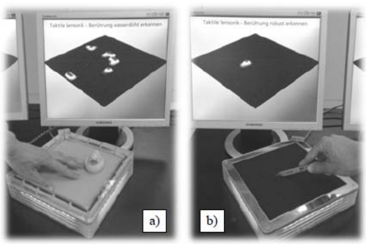


Figure 4. Tactile transducer with related sensor data: a) water proof b) cut-resistant

The tactile transducers may have any geometric shape and can even be adapted to two-dimensional curved free-forms. The number, shape and size of the individual sensor cells can be customized.

A demonstration of this geometry-adapted sensor system's functionality can be found in a study in which the KUKA lightweight robot with extremely complex geometry served as the target system (Figure 5).

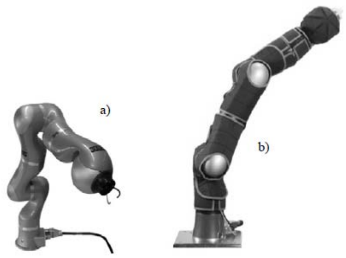


Figure 5. KUKA lightweight robot
(a) without and (b) with the tactile sensor system

B. Sensor Controller

Our tactile sensing system is equipped with an intelligent sensor controller as front end. It contains microcontroller-based circuits scanning and sampling the connected tactile transducers.

We use two separate multiplexers to address the individual sensor cells of our matrix-based tactile transducers (Figure 6). The number of channels needed is identical with the number of rows and columns. In order to eliminate the problem of crosstalk in matrix-based tactile sensor systems, we implemented a signal conditioning unit which is based on the zero potential method proposed by Shimojo et al. [9].

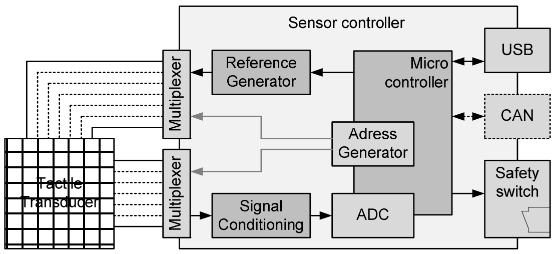


Figure 6. Block diagram of the sensor controller

The resistance change of the selected sensor cells is measured by an integrated ADC. The ADC has a resolution of 10 bits at sampling frequencies up to 20 kHz.

Rapid data communication interfaces such as CAN or USB are employed to make the acquired sensor data available. These data may thus be integrated into the robot control for further processing or visualization.

Integrated preprocessing algorithms provide low-level safety functions. If the load applied to the tactile transducer exceeds a predefined threshold, a usually closed safety switch is opened. This safety switch is integrated into the robot system's safety circuit and may be used to stop the robot's movement.

C. Sensor Topology

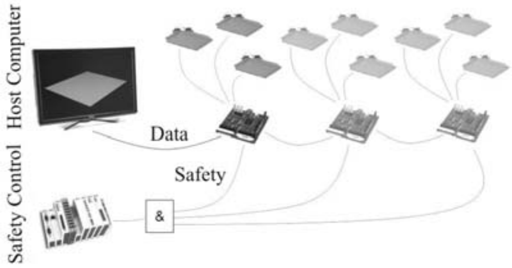


Figure 7. Sensor topology

As shown in Figure 7, a CAN or USB bus can be used to connect the multiple sensor controllers to a superordinate control. In turn, each sensor controller supports multiple tactile transducers.

By default the tactile transducers are connected to the sensor controller via ZIF connectors. Custom connection boards (Figure 8) can be mounted to an extension header available on the sensor controller.

Figure 8. Sensor controller with housing
and custom connection board

Using a logical AND link, the safety switch of each sensor controller can be integrated in the robot system's safety circuit. Thus, if one controller fails or detects a dangerous contact, the robot system will enter a safe state.

For pure safety applications, the sensor controller may be used stand-alone, i.e., the data interfaces merely serve parameterization.

III. THE LiSA ROBOT

Figure 9. The LiSA robot interacting with a human

The Life Science Assistant (LiSA) [10] was built and developed within the LiSA project supported by the German Federal Ministry of Education and Research as part of its program “Key Innovations: Service Robots”.

The objective of this project was to develop a mobile service robot suitable for everyday routine tasks in lab environments of biotechnology companies (Figure 9). One key aspect of development was an overall system design aiming for safe human-robot interaction within shared workspaces.

The LiSA assistant robot mainly consists of a mobile robot base with a robotic arm mounted atop. The robotic arm was realized by using a classic SCARA setup. Safe human-robot cooperation is ensured by an extensive safety and sensor system on board the robot. Besides proven safety technologies, such as laser scanners and bumpers, that have already been implemented in automatic guided vehicle systems many times before, novel safety components and strategies such as the artificial skin were tested within the project.

IV. An Artificial Skin for LiSA

The LiSA robot has been equipped with fourteen geometry-adapted skin patches (Figure 10). Five of them are located on the robot base, four on the lower segment of the robotic arm and five on the upper segment.

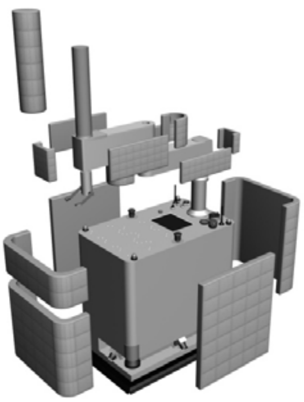


Figure 10. Exploded view of the LiSA robot showing the individual skin patches

The skin patches were designed in a way that they entirely protect the robot in its directions of movement. The size and number of sensor cells and thus the attainable spatial resolution varies across the robot, depending on the positions of the skin patches.

The biggest sensor cells with about 10 cm x 10 cm are located on the mobile robot base, and smaller sensor cells with about 3 cm x 3 cm cover the robotic arm. Altogether the robot’s surface is covered by 375 sensor cells.

Initially, the placing of sensor cells was inspired by the human tactile sensor system: We have a low spatial resolution on the body (mobile robot base) and a higher spatial resolution on the extremities (the robotic arm). The extremities are mainly used for interaction and therefore require a precise determination of the contact position. Mounted on the mobile platform, the artificial skin just represents an additional feature complementing the existing bumpers and laser scanners. Thus even the low spatial resolution provides us with useful information.

As shown in
Figure 11 LiSA’s skin patches are set up in three layers.

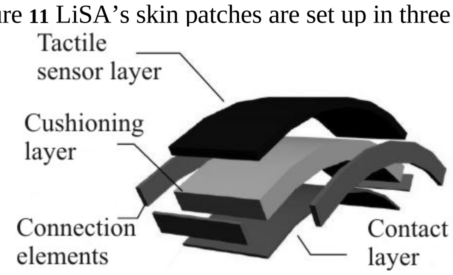


Figure 11. Exploded view of sensor patch

The top layer forms the tactile sensor layer, the middle is a cushioning layer and the back layer can be best described as a contact layer, providing mechanical and electrical contact to the robot.

To give an optimal fit to the sensor patches, all three layers are sewn together by using precast connection elements. These elements also guide the electrical signals from the top layer to the bottom layer.

The cushioning layer fulfills two contradictory tasks. On the one hand it should be soft to cushion collisions. On the other hand it should be strong and durable to resist a large number of collisions and to give mechanical support to the tactile sensor layer placed on top of the cushioning layer.

In accordance to the bumpers at the robot base we integrated a 40 mm thick cushioning layer into the skin patches placed at the mobile robot base. The skin patches at the robotic arm have been equipped with a 20 mm thick cushioning layer.

V. Implementation of Force-Guided Motion Control

Due to the artificial skin we are able to detect forces applied to the robot’s surface. By interpreting these forces as motion commands, it is possible to implement touch-based motion control algorithms. Our current development stage includes implementation of such control algorithms on LiSA’s robotic arm.

The robotic arm is a classic SCARA setup with four degrees of freedom (Figure 12). To apply force-guided motion control to all axes, we implemented three different strategies of motion control.

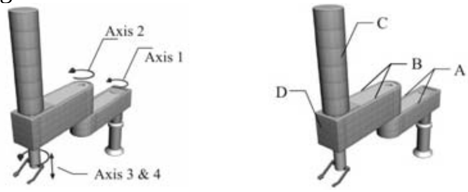


Figure 12. The robotic arm of LiSA with control functions related to the skin patches, A / B – axis-wise motion control, C – Cartesian control, D – control functions for axis 3 / 4 and the gripper

 ISBN: 978-1-61208-117-5

The skin patches A and B (as depicted in Figure 12) are used for axis-wise motion control of axis 1 and axis 2, respectively. Skin patch C can be used for simultaneous control of axis 1 and axis 2 in Cartesian space. Skin patch D is used to control axis 3, axis 4 and the gripper.

A. *Axis-wise motion control*

For axis 1 and axis 2, axis-wise motion control algorithms are implemented. The control algorithms are mainly based upon a torque compensation approach. In accordance with Figure 12 these algorithms are activated if forces are applied to the skin patches A or B.

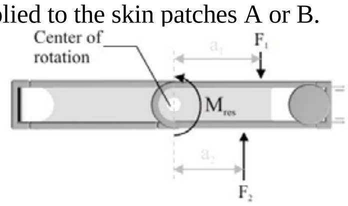


Figure 13. Top view of the robotic arm, with forces applied to axis 2

As shown in Figure 13 forces F_i applied to the skin patches result in a torque M_{res} which the control algorithms try to compensate. Thus the affected axis is moving in the direction of the resulting torque with the speed v_{rot} as calculated in (1).

$$v_{rot} = v_{max} * \frac{M_{res}}{M_{max}} = v_{max} * \frac{\sum_{i=1}^{n} F_i * a_i}{F_{max} * a_{max}} \qquad (1)$$

M_{max} - *maximal torque*
F_{max} - *maximal force measureable by the tactile sensor*
a_{max} - *maximal distance between center of rotation and force contact point*

B. *Cartesian motion control*

The cylindrical casing of the lifting spindle is covered by skin patch C (cf. Figure 12). Pushing this cylinder causes a change of the gripper's position in a horizontal plane.

To achieve this movement we need to detect the individual forces F_i applied to the cylinder and sum up the appropriate force vectors. The resulting vector F_{res} indicates a direction and velocity of movement (Figure 14). It can be used to calculate the necessary rotational speeds for the axes 1 and 2 to move the manipulator as desired. This approach is an easy way to teach horizontal Cartesian positions.

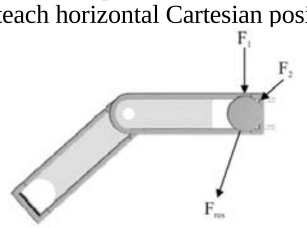


Figure 14. Top view of the robotic arm, with forces applied to skin patch C

C. *Special Motion Control*

Axis 3 and axis 4 cannot be controlled directly since they are not covered by skin patches. Nevertheless we can provide touch-based motion control by implementing virtual buttons to the skin patch D (cf. Figure 12).

We implemented two buttons for each axis and the gripper, as depicted in Figure 15. Unlike standard push buttons only providing an on/off signal, these tactile sensor-based buttons give us force-based motion control: The higher the force applied to the button the faster the selected axis moves. The motion speed of the selected axis is calculated according to equation (2).

$$v = v_{max} * \frac{F}{F_{max}} \qquad (2)$$

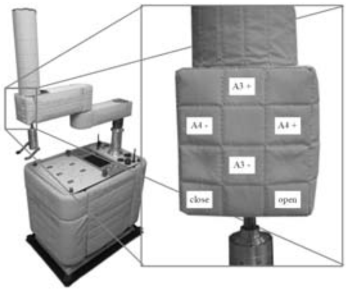


Figure 15. Skin patch D with special motion control functions
A3+ / A3- / A4+ / A4- : motion control for axis 3 and axis 4
open / close : motion control for the gripper

VI. Experimental Evaluation of the Safety Function

As mentioned before, the cushioning layer integrated into the sensor setup is used to cushion collisions and to provide mechanical support to the tactile sensor layer. Thus the objective of our evaluation process was to show that our sensor setup can fulfill both contradictory tasks.

In a first experiment we investigated the ability of the cushioning layer to give mechanical support to the tactile sensor layer. To do so, we integrated a tactile transducer with a 20 mm cushioning layer into a force measurement stand and recorded the sensor resistance and the applied pressure according to the compression of the tactile transducer.

As illustrated in Figure 16, we have a noticeable change in resistance within the first two mm. Within this sensing range, the internal preprocessing algorithms of our sensor controller can generate a reliable stop signal. After signal generation about 15 mm can be used to stop the robot without loading the collision partner with high load peaks.

Within this experiment we were also able to identify the minimal pressure needed to generate a reliable stop signal, which is about 0.25 N/cm².

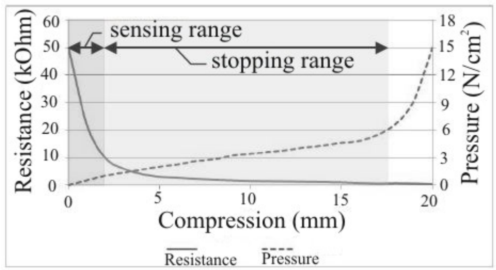


Figure 16. Results of first experiment

In a second series of experiments we investigated the sensor system's ability to cushion collisions. With the aid of a high dynamic force measurement stand we measured forces and impact energy. The main part of the force measurement stand is a three-component force measurement plate (Figure 17). The collisions were simulated by using a pendulum, whose deflection and thus kinetic energy were identical for all collision experiments.

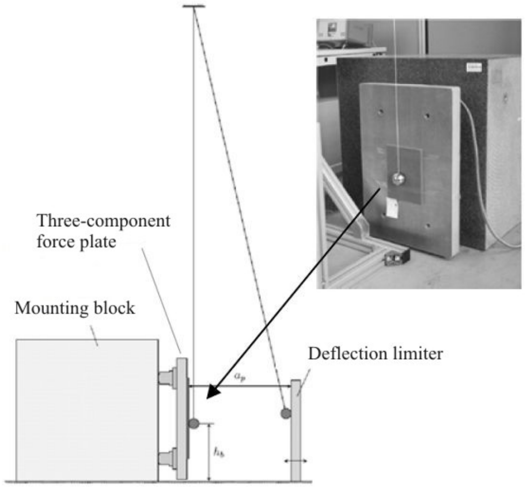


Figure 17. High dynamic force measurement stand

We investigated two different collision scenarios. In a first experiment we applied one of our standard tactile transducers without cushioning layer to the measurement plate. In the second experiment we employed a tactile transducer with a 20 mm cushioning layer to cushion the collision. As shown in Figure 18 the use of a tactile transducer with cushioning layer can significantly reduce the load peak. In the given example the load peak could be reduced from 820 N to about 24 N, representing a reduction factor of 34.

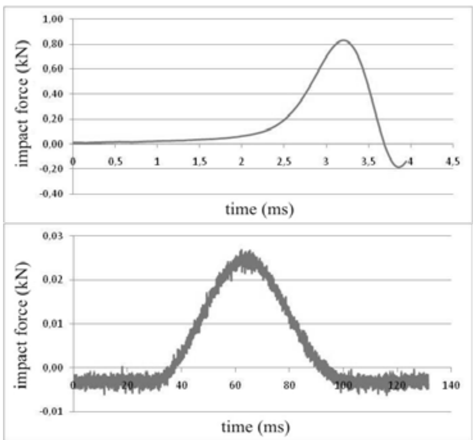


Figure 18. Results of collision experiments
a) with standard tactile transducer
b) with cushioned tactile transducer

VII. Conclusion and Further Work

In this paper, we described the setup and system integration of a pressure-sensitive skin for a mobile robot. The use of mainly textile components enabled us to create a versatile and flexible skin solution.

The textile setup makes the tactile transducers insensitive to mechanical stress like bending. This enables us to shim the tactile transducer by a cushioning layer. Since the cushioning layer provides the robot with a soft surface capable of absorbing collision energy, it also ensures an enhanced safety in human-robot interaction.

The innovative mounting technology using snap fasteners for mechanical and electrical connection enables us to ensure that the sensor patches are mounted properly.

The sensor technology primarily aims at covering large robotic structures with low to medium spatial resolution. It is hardly qualified for high spatial resolution applications such as the coverage of finger tips, etc.

At the time of this publication our robot LiSA is entirely equipped with the sensitive skin. The sensor system has been successfully integrated into the emergency stop system and into the motion control system, as well. Thus we are able to provide a safe touch-based motion control interface for human-robot interaction scenarios.

References

[1] J.G. Webster, "Tactile sensors for robotics and medicine". New York: Wiley, 1988.

[2] H.R. Nicholls, "Advanced tactile sensing for robotics". Singapore: World Scientific, 1992.

[3] V. Lumelsky, "Sensitive skin". Singapore: World Scientific, 2000.

[4] H. Iwata and S. Sugano, "Whole-body covering tactile interface for human robot coordination", in Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on, 2002, pp. 3818-3824 vol.4.

[5] T. Asfour, K. Regenstein, P. Azad, J. Schröder, and R. Dillmann, "ARMAR-III: A Humanoid Platform for Perception-Action Integration", Proceedings of 2nd International Workshop on Human-Centered Robotic Systems, 2006

[6] Y. Ohmura, Y. Kuniyoshi, and A. Nagakubo, "Conformable and scalable tactile sensor skin for curved surfaces", in Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on, 2006, pp. 1348–1353.

[7] G. Cannata, M. Maggiali, G. Metta, and G. Sandini, "An embedded artificial skin for humanoid robots", in Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on, 2008, pp. 434–438.

[8] K. Weiss and H. Woern, "The working principle of resistive tactile sensor cells", Proc. of the IEEE International Conference on Mechatronics and Automation, Niagara Falls, Canada, July 2005

[9] M. Shimojo, A. Namiki, M. Ishikawa, R. Makino, and K. Mabuchi, "A tactile sensor sheet using pressure conductive rubber with electrical-wires stitched method", Sensors Journal, IEEE, vol. 4, no. 5, pp. 589–596.

[10] E. Schulenburg, N. Elkmann, M. Fritzsche, J. Hertzberg, and S. Stiene, "LiSA: Auf dem Weg zur sicheren Assistenzrobotik", KI - Künstliche Intelligenz, vol. 24, no. 1, pp. 69–73, 2010

[11] M. Fritzsche, N. Elkmann, "An Artificial Skin for Safe Human-Robot-Interaction". In: Humanoids 09. Workshop on Tactile Sensing in Humanoids – Tactile Sensors & beyond. (Paris 7. December 2009) - Proceedings, pp. 42-43.

A Walking Aid Integrated in a Semi-Autonomous Robot Shopping Cart

Hermann Kaindl, Bernhard Putz, Dominik Ertl
Vienna University of Technology
Institute of Computer Technology
Vienna, Austria
(kaindl, putz, ertl)@ict.tuwien.ac.at

Helge Hüttenrauch, Cristian Bogdan
Royal Institute of Technology
School of Computer Science
Stockholm, Sweden
(hehu, cristi)@csc.kth.se

***Abstract*—Challenged and/or elderly people experiencing limited mobility impairment may want to get support for walking. In public, e.g., in a supermarket, they may want to get this support without it being immediately visible. Therefore, we integrated walking aid functionality into a robot shopping cart. It can support a customer to lean on the cart while the walking pace is controlled to follow a user-determined setting. More precisely, the user of the cart can get walking assistance by holding specifically designed handle bars supporting both arms. This construction is fully integrated in a prototypical robot designed as a shopping cart.**

***Keywords*—Walking aid; semi-autonomous robot shopping cart.**

I. Introduction

Especially in an aging society, more and more (partial) challenges arise. Primarily elderly people often experience challenges with walking. As long as they can walk, however, they typically prefer some kind of walking assistance over a wheel chair. In public, they may even want to hide their challenge and to get such support without it being immediately visible. The latter requirement is hard to fulfill with a dedicated walking aid.

We developed a (prototypical) robot serving as a shopping cart and, from the very outset, we planned to integrate a walking aid. When someone would use such a robot shopping cart, this very robot can serve both the usual shopping purposes and as a walking aid. Its motor power lets it drive to desired goods and help move loaded goods, and at the same time provide walking assistance. In addition to the mode with the walking aid, the robot cart has a steering mode, an autonomous mode and semi-autonomous modes for guiding a user to certain products and for following the user. Unless the robot is in the walking aid mode, its handle bars are in a locked position.

This shopping cart is a companion robot designed to perform services in a complex and cluttered dynamic environment, a supermarket, shared by humans and other robots. To share such an environment with humans and to cooperate with them requires certain communication abilities from the robot. Even our prototypical implementation for research purposes can handle multimodal input through a touch screen, speech input and a barcode reader [3], as well as output through a graphical user interface and speech output [2]. It can even reinforce multimodal output through a motion cue [6]. The primary services that we implemented in the prototype are management of a shopping list and guiding a user to the products from the shopping list.

The remainder of this paper is organized in the following manner. First, we explain the design and implementation of this walking aid. Then we present an exploratory study with it. Finally, we discuss related work.

II. Walking Aid Design and Implementation

We designed the cart to provide walking assistance for users with special needs. In particular, it provides physical support for walking, via an add-on ergonomically designed for such users. Another main goal was to particularly support those who want to hide their challenge while shopping in a supermarket.

Taking existing designs for non-robot walking aids into consideration, the idea was to use a construction supporting both arms with handle bars positioned at waist-height and equipped with buttons for user control. However, no adjustments in width and height were intended, in order to keep the design and the first-time usage simple. The buttons can be pressed easily while holding the handle bars for being physically supported. Due to safety reasons, two (yellow) bumper rings enclose the outer cart shape for detecting collisions. These bumpers limit the leg-moving space, i.e., a walking aid add-on inside the bumpers has safety-related advantages but is ergonomically inapplicable.

Therefore, the handle bars of our walking aid are swivel-mounted on the cart (see a design sketch in Figure 1 with body proportions). Figure 2 shows the implemented walking aid in its locked position. Whenever the walking aid is to be used, the handle bars can be swiveled out of their vertical spring-locked position. Figure 5 shows this position of the implemented walking aid when in use for steering the robot. The swivel-out event is detected by two micro switches to enable an automatic mode change of the cart.

Each of the two bars ends is equipped with two pushbuttons for controlling speed and orientation of the cart. Two of these buttons can be seen in Figure 2. The walking aid control through all four buttons is illustrated in Figure 3.

 ISBN: 978-1-61208-117-5

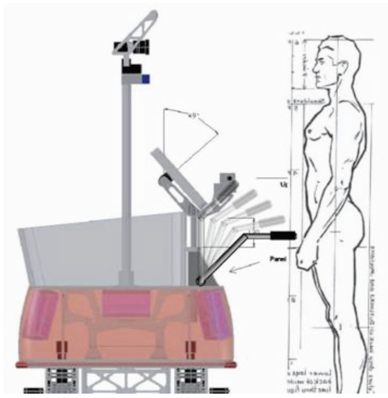


Figure 1. Swivel-mounted handle bars for our cart walking aid related to body proportions.

Figure 2. Implementation of the walking aid — locked position.

However, there is no step-less speed control, and compared to the autonomous modes of the cart, maximum speed and acceleration are reduced for better controllability.

III. Exploratory Study

We have explored how a robotic shopping cart apart from its main function of carrying goods and guiding to locations in the store, could have a second function as a walking aid. In order to conduct a formative evaluation of the walking aid, we carried out a small user study where we let two users drive the robot (as shown in Figure 5), who where untrained on the walking aid. As the prototype is motorized and in an early stage, we did not involve people with actual walking impairments as users of the system for safety reasons.

A. Setup and task

The participants were introduced to the notion of a walking aid. They were also shown how to operate the robot using the buttons on the handles, both with the illustration shown in Figure 3 and by letting them try it out on their own. Using a think-aloud protocol, the participants reported which one of the functionalities they attempted to use.

After that, they had to perform the tasks depicted in Figure 4. The first task was to drive *forward* in a slalom course to a point located in the map (see Point 1). Once at Point 1, they were instructed to drive *backward* to Point 2. This slalom course contains the typical maneuvers one would perform with such a shopping cart. The grey boxes shown in the figure are empty cartons with a dimension of about $60 \times 60 \times 60$cm. The participants were instructed that they should not touch the boxes with the robot.

After they had performed these tasks, the participants received a subjective questionnaire and filled it out.

B. Results

Both participants managed to learn driving the robot using the given interface, and they also succeeded with the driving task. Further studies, however, are necessary to figure out if this electro-mechanical interface is helpful and easy to learn for the elderly and (partially) challenged people as well. Since only two people took part in the study, it makes little sense to provide quantitative data. There were some open questions where the participant expressed their opinions:

- *"[I] would like the direction button on the other hand"*
- *"a bigger turning radius would be better for handicapped people"*
- *"The global behavior is easy to understand even if a few tries are needed to learn the commands"*
- *"[It's] easier to learn by using than by reading instructions."*

The construction was considered stable and pleasant to grasp, and its handling and use easy to learn. Interestingly, just trying it out was considered more effective than the instructions through the illustration in Figure 3. Still, the placement of buttons should be studied further.

IV. Related Work

Most of the related research activities focus solely on the walking function. They implemented their support for walking as a primary, single function, either by assisting with the driving using the handle as input (see [1], [11]) or by assisting visually impaired users to a location using a physical interface modeled after the leashes of guide dogs. The user grabs the handle and follows the robot (see [8], [9]).

Graf et al. [5] designed and implemented a semi-autonomous walking aid platform for elderly and handicapped people. These authors have people in mind that use

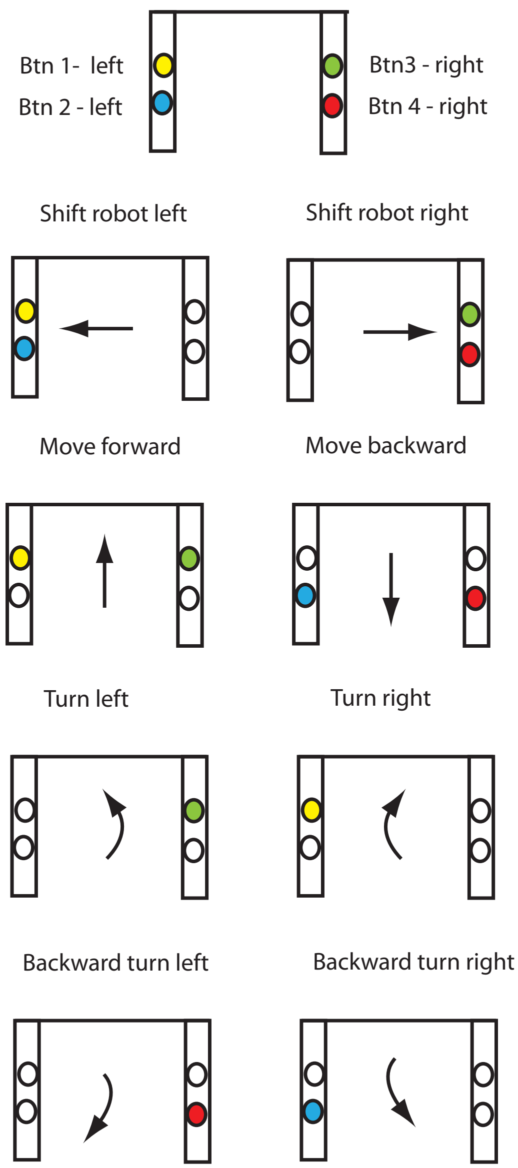


Figure 3. Walking aid control through the four buttons.

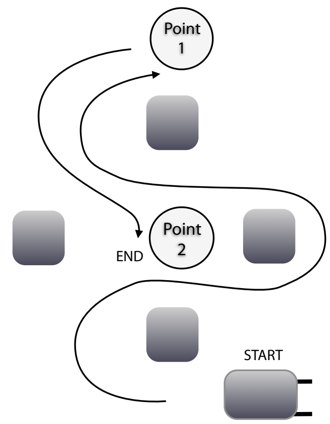


Figure 4. The map showing the task that was used for the pilot study.

this platform day-by-day, thus getting used to it. While this walking aid is used mainly at home, we focus on a populated domain like a supermarket, where people typically do not have much time to learn a user interface. Moreover, we focus on slightly impaired users who are, in principle, able to walk on their own, but might require some help when maneuvering a shopping cart filled with a lot of goods, and would like to hide their walking problems.

Addressing the same problem of keeping the elderly and handicapped independent, previous research by Lee et al. [10] and Annicchiarico et al. [1] presents personal assistive mobility devices for elderly people with walking impairments. However, their focus is on users requiring a "pure" walking support. Our work does not fit well for those that who have more severe problems when using a shopping cart in a supermarket.

Kanda et al. [7] presented an affective guide robot in a shopping environment. This robot supports people to navigate through a shopping mall and provides information about products via a speech interface for input and output. Even though our service robot shares the same domain with this service robot, their work does not have any functionality to help people with a walking impairment.

Glas et al. [4] introduce a guiding and helping robot called RoboPal. It is designed to operate in the domain of everyday life, acting as a guide or helping people with daily errands in real-world environments. The authors studied nonverbal cues of RoboPal associated with leading and following behavior of the robot. However, due to its morphology, RoboPal is a communication robot only and not intended to physically support the elderly or disabled in the course of maneuvering goods in a shopping cart.

V. Conclusion

We designed and built a walking aid that is prototypically integrated in a semi-autonomous robot shopping cart. Through building it and a small exploratory study, we gained first insights and showed the principal feasibility of such an approach. After the safety issues will have been solved satisfactorily, studies and experiments with people having partial challenges with walking can be a next step for further development of such an integrated walking aid.

Acknowledgments

This research has been carried out in the CommRob project (http://www.commrob.eu), partially funded by the EU (contract number IST-045441 under the 6th framework programme). The idea of a walking aid integrated in a robot shopping cart originally came from Vasilios Spais in the course of preparing the project proposal. Special thanks

Figure 5. The walking aid in use.

are due to the Institute of Mechanics and Mechatronics of the Vienna University of Technology and, in particular, Peter Unterkreuter for providing help in the manufacture and assembly of all mechanical walking aid components.

REFERENCES

[1] R. Annicchiarico, C. Barrué, T. Benedico, F. Campana, U. Cortés, and A. Martínez-Velasco. The i-Walker: An intelligent pedestrian mobility aid. In *Proceedings of the 2008 European Conference on Artificial Intelligence (ECAI 2008)*, pages 708–712. IOS Press: Amsterdam, The Netherlands, 2008.

[2] D. Ertl, J. Falb, and H. Kaindl. Semi-automatically configured fission for multimodal user interfaces. In *Proceedings of the Third International Conference on Advances in Computer-Human Interactions (ACHI 2010)*, pages 85–90. IEEE Computer Society Press: Piscataway, NJ, USA, 2010.

[3] D. Ertl, S. Kavaldjian, H. Kaindl, and J. Falb. Semi-automatically generated high-level fusion for multimodal user interfaces. In *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences (HICSS-43)*. IEEE Computer Society Press: Piscataway, NJ, USA, 2010.

[4] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita. RoboPal: Modeling role transitions in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2007)*, pages 2130–2137, April 2007.

[5] B. Graf, M. Hans, and R. D. Schraft. Care-o-bot ii — development of a next generation robotic home assistant. *Auton. Robots*, 16(2):193–205, 2004.

[6] H. Hüttenrauch, C. Bogdan, A. Green, K. S. Eklundh, D. Ertl, J. Falb, H. Kaindl, and M. Göller. Evaluation of Robot Body Movements Supporting Communication. In *Proceedings of New Frontiers in Human-Robot Interaction Symposium of the Convention Artificial Intelligence and Simulation of Behaviour (AISB)*, Leicester, UK, 2010.

[7] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*, pages 173–180. ACM: New York, NY, USA, 2009.

[8] V. Kulyukin, J. Nicholson, and D. Coster. Shoptalk: Toward independent shopping by people with visual impairments. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '08)*, pages 241–242. ACM: New York, NY, USA, 2008.

[9] V. A. Kulyukin and C. Gharpure. Ergonomics-for-one in a robotic shopping cart for the blind. In *HRI '06: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, pages 142–149, New York, NY, USA, 2006. ACM.

[10] G. Lee, T. Ohnuma, and N. Chong. Design and control of JAIST active robotic walker. *Intelligent Service Robotics*, 3:125–135, 2010. 10.1007/s11370-010-0064-5.

[11] T. Röfer, T. Laue, and B. Gersdorf. iWalker — An Intelligent Walker providing Services for the Elderly. In *Proceedings of the European Conference on Technically Assisted Rehabilitation (TAR '09)*, Berlin, Germany, 2009. VDe/VDI.

Towards Automated Human-Robot Mutual Gaze

Frank Broz*, Hatice Kose-Bagci†, Chrystopher L. Nehaniv* and Kerstin Dautenhahn*
* *University of Hertfordshire*
Department of Computer Science
Hatfield, UK
{*f.broz, c.l.nehaniv, k.dautenhahn*}*@herts.ac.uk*
† *Istanbul Technical University*
Computer Engineering Department
Istanbul, Turkey
hatice.kose@itu.edu.tr

***Abstract*—The role of gaze in interaction has been an area of increasing interest to the field of human-robot interaction. Mutual gaze, the pattern of behavior that arises when humans look directly at each other's faces, sends important social cues communicating attention and personality traits and helping to regulate conversational turn-taking. In preparation for learning a computational model of mutual gaze that can be used as a controller for a robot, data from human-human pairs in a conversational task was collected using a gaze-tracking system and face-tracking algorithm. The overall amount of mutual gaze observed between pairs agreed with predictions from the psychology literature. But the duration of mutual gaze was shorter than predicted, and the amount of direct eye contact detected was, surprisingly, almost nonexistent. The results presented show the potential of this automated method to capture detailed information about human gaze behavior, and future applications for interaction-based robot language learning are discussed. The analysis of human-human mutual gaze using automated tracking allows further testing and extension of past results that relied on hand-coding and can provide both a method of data collection and input for control of interactive robots.**

***Keywords*-mutual gaze; human-robot interaction; psychology; Markov model**

I. INTRODUCTION

Mutual gaze is an ongoing process between two interactors jointly regulating their eye contact, rather than an atomic action by either person [1]. This social behavior is important from an early developmental stage; even young infants are responsive to being the object of a caretaker's gaze [2]. Mutual gaze behavior is the basis of and precursor to more complex task-oriented gaze behaviors such as visual joint attention [3].

Mutual gaze is also an important part of face-to-face communication. It is a component of turn-taking "proto-conversations" between infants and caretakers that set the stage for language learning [4] and is known to play a role in regulating conversational turn-taking in adults [5]. There is evidence that children learn this coordination of gaze with conversational turns during early language acquisition, shifting towards an adult-like pattern as they gain more language skills [6].

Rather than measuring actual eye contact, "mutual gaze" is typically defined as subjects looking at one another's faces [1]. This is primarily due to the measurement limitations of having gaze direction coded by a human observer. There is limited evidence that people themselves may have trouble distinguishing gazes to different features within the face [7]. But the gaze patterns used in these experiments were highly unnatural, and people may access direction more accurately from natural conversational gaze.

Recently, the field of human-robot interaction has become increasingly interested in the role of gaze in a variety of conversational tasks, and robots have been programmed to produce natural-appearing mutual gaze behavior. But these robots base their behavior on models that typically focus on one important aspect of mutual gaze, such as reactivity or timing, while ignoring others. In work by Yoshikawa and colleagues, the robot responds to human gaze but do not take any action to regulate the duration of mutual gaze itself [8]. In the story-telling robot study by Mutlu, Forlizzi and Hodgins, a robot produces human-directed gaze behavior based on a model with realistic timings that is not responsive to real-time gaze information [9].

The characteristics of human gaze when interacting with robots is also an active area of research. Yu and colleagues performed a temporal analysis of human gaze and speech behavior from a human-robot interaction word teaching task with a robot that autonomously performed a simple form of joint attention [10]. While this study provides insight into patterns of human gaze at a robot, the simplicity of the robot's controller makes it unlikely that humans found the gaze interaction to be natural or its dynamics to be similar to gaze between two humans. Also, gaze was analyzed by looking at the entire robot rather than examining whether people fixated on the robot's face or particular facial features as they would with a human partner. In another human-robot study, Vollmer and colleagues found a significant decrease in gaze directed at the learner in a tutoring task when the

 ISBN: 978-1-61208-117-5

learner was a childlike virtual robot with a simple salience-based attention model rather than a human child [11]. They cite differences in the robot's visual feedback as a likely reason for the differences in tutor behavior. Gaze behavior in dyads is an interaction, and the robot's gaze policy will have an impact on the human. In order to support natural and effective gaze interaction, it is worthwhile to first look at gaze behavior in human-human dyads. By examining human gaze interactions, we can gain insight into how to build better gaze policies for robots that interact with people.

For a robot to successfully negotiate humanlike mutual gaze, it must both be responsive to the human's immediate gaze behavior and possess an internal model of mutual gaze based on time and other significant factors. Robotic systems designed to learn language through interaction by exploiting the structure of child-directed speech (e.g., [12]) could especially benefit from a gaze model that supports social engagement. Building models by using data collected from human-human pairs is likely to improve the quality of interaction with these systems.

There has been some previous research into using human-human gaze data to produce agent gaze. Raidt and colleagues conducted a study into face-to-face real time communication and gaze direction [13]. However, people interacted through a pair of video displays. While this is appropriate to their computer-agent model, it unnaturally constrains people's options for movement (as opposed to co-located face-to-face conversation). Also the speech task involved was one of repetition and memorization rather than natural conversation. Given these constraints on user behavior, it is unclear whether the data collected is representative of human conversational gaze behavior.

The rest of this paper describes a conversational gaze interaction experiment and its results. In Section II, the implementation details of the system used for data collection is described. Section III presents the design and setup of the experiment itself. Experimental results are discussed in Section IV, including overall amounts of mutual gaze and gaze duration, differences in behavior of individual pairings and gaze at specific areas of the face. In Section V a Markov model of the gaze data is discussed, both in terms of what it reveals about the data set and in terms of how such a model could be used for robot control.

II. System Overview

The automated detection of mutual gaze requires a number of signal-processing tasks to be carried out in real time and their separate data output streams to be combined for further processing. Note that if the goal of this work were solely to study mutual gaze in humans rather than to provide input for a robot control system, there would be no requirement for real-time operation. The video could be collected and then analyzed later offline. The system is a mixture of off-the-shelf programs and custom-written software combining and processing their output. The interprocess communication was implemented using YARP [14].

Figure 1. A pair of subjects engaging in conversation during an experiment.

ASL MobileEye gaze tracking systems were used to collect the gaze direction data [15]. The output of the scene camera of each system was input into face-tracking software based on the faceAPI library [16]. Each participant also wore a microphone which was used to record a simple sound level (though this data was collected, it is not included in the experimental analysis in this paper). Timestamped data of gaze direction (in x,y image pixel coordinates) and the location of the partner's facial features (in pixel coordinates) were recorded at a rate of 30 hertz. In order to synchronize time across machines to maintain timestamp accuracy, a Network Time Protocol (NTP) server/client setup was used. This setup is typically able to maintain clock accuracy among machines within a millisecond or less over a local area network [17].

III. Experiment

Experiment participants were recruited in pairs from the university campus. A requirement for participation was that the members of each pair know one another. This requirement was chosen to limit one possible source of variability in the data, as strangers have been shown to exhibit less mutual gaze than people who are familiar with one another.

The pairs were seated approximately six feet apart with a desk between them. This distance was chosen to allow comparisons to be made to earlier studies. They were informed that they would engage in an unconstrained conversation for ten minutes while data was recorded. The participants were asked to avoid discussing potentially upsetting topics (so that extreme emotional reactions would not effect their behavior) and given a list of suggestions should they need one, which included: hobbies, a recent vacation, restaurants, television shows, or movies. After filling out a consent form and writing down their demographic information, each participant was led through the procedure to calibrate the gaze tracking system by the experimenter before the trial began.

After the trial, participants were asked to complete a short questionnaire. This questionnaire's purpose was to collect additional data about individual traits that may have an influence on gaze behavior. Participants were asked which country they'd lived in for most of their life and how and how well they knew their partner for the experiment. They also filled out a ten item personality inventory (TIPI), a short personality test which assesses people on big five personality traits [18].

IV. Results

Ten pairs of people participated in the study. Of these pairs, five experienced errors during data collection that resulted in their data being discarded from the study. The nature of these errors were: loss of gaze tracker calibration due to the glasses with the camera mount slipping or being moved by the participant, failure of the face tracker to acquire and track the face of a participant, and failure of the firewire connection that was used to transmit the video data to the computers for analysis. These failures reflect the difficulty of deploying a real-time system for mutual gaze tracking due to the complexity of the necessary hardware and software components. This experiment was the first of its kind conducted by this group, and quickly and reliably calibrating the system for each different individual was a process that required practice. Participants were promised that the experiment would last no longer than thirty minutes. In the case of difficulties with calibration or hardware failure, we continued the experiment and collected incomplete data from the working portions of the system so as not to inconvenience the participants. The five remaining pairs of participants for whom complete face and gaze tracking data were available were used for data analysis. They ranged in age from 23 to 69. Of the pairs, two were male-male, two were male-female, and one was female-female.

For each pair, the contiguous two minute period of their data with the lowest number of tracking errors was selected for analysis. The gaze behavior was divided into a set of high-level gaze states. In all pairs observed, one participant looked at their partner noticeably more than the other. The participant with the high face-directed gaze level will be referred to as the "high" participant and the partner with the lower level of face-directed gaze will be referred to as "low". The gaze states and their descriptions are given below:

- Mutual - mutual gaze, as defined as both participants' looking at one another's face area
- At Low - the high gaze level partner looks at the face of the low level partner while they look elsewhere
- At High - the low gaze level partner looks at the face of the high level partner while they look elsewhere
- Away - both partners look somewhere other than their partner's face

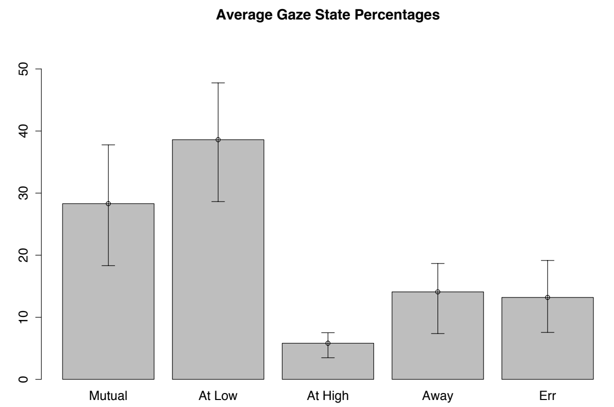


Figure 2. The average percentage of time spent in each gaze state by all pairs, with 95% bootstrap confidence intervals.

- Err - gaze state could not be classified due to missing gaze direction or face location readings

The average percentage of time spent by all pairs in each gaze state is shown in Figure 2. The mutual gaze results are in agreement with results from earlier studies which predict conversational mutual gaze at this distance to be around 30% [19]. It can also be seen that there is a marked asymmetry in the partners' gaze behavior, with one partner looking at the other's face far more often.

While the percentage of timesteps that cannot be classified due to tracking error is relatively high, we do not believe that it has a dramatic impact on the mutual gaze estimate. The reason for this is the source of the errors in the tracking system. Transient errors in the system occur for one of two reasons, lost gaze tracking readings or lost face tracking readings. Face tracking is usually lost when either of the participants move their head very quickly or when the partner's face is outside of the head-mounted camera's field of view (which happens when a person's head is directed away from their partner). Gaze tracking is lost when the system cannot find the participant's pupil, which typically happens when they are looking at a point at the periphery of their vision. Therefore, tracking errors occur most frequently when one or both participants have their head and/or gaze directed away from the other's face rather than towards it. This hypothesis will be supported by an examination of the pattern of transitions between gaze states (to be discussed in Section V).

A. Results for individual pairings

Figure 3 shows the gaze state percentages for each individual pair of participants. It can be seen that there are noticeable differences in gaze behavior between pairs. In particular, pairs three and five exhibit far more mutual gaze

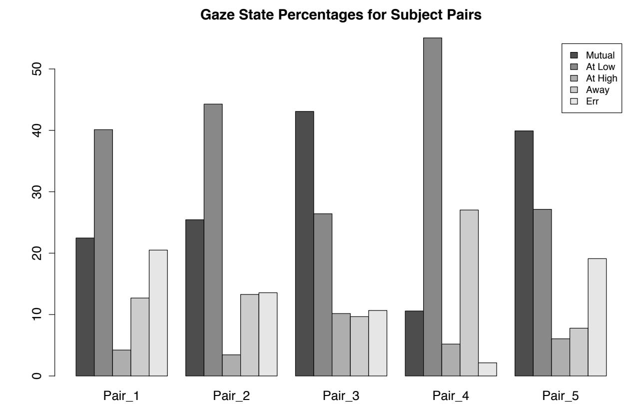


Figure 3. The average percentage of time spent in each gaze state by each pair.

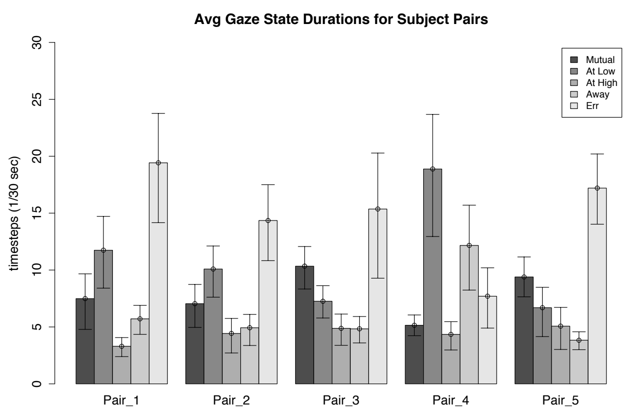


Figure 4. The average duration of time spent in each gaze state by each pair, with 95% bootstrap confidence intervals.

than the other pairs (these were a male-female and a male-male pair, respectively). We examined characteristics that might explain this difference, such as personality traits of extroversion and social dominance, how well the partners knew one another, and gender. The personality traits were measured through either single responses or a combination of responses (in the case of social dominance) from the short personality test administered. However, we failed to find a characteristic that could explain the difference in this small sample. In future studies with a larger number of participants we will focus on identifying characteristics that lead to different gaze dynamics and how knowledge about such characteristics might be incorporated into a gaze model for a robot to support more natural-seeming behavior.

One surprising result of this study is that the duration of each mutual gaze ranges from around one third to one sixth of a second (see Figure 4). These measurements are far shorter than a previously reported average figure of 1.18 seconds [19]. This may be due to the fact that automated techniques are more capable of measuring gaze at a fine temporal resolution than a human experimenter making judgements during observation. There is also a possibility that transient tracking error may cause the system to underestimate gaze durations. This will be investigated in future experiments using offline video analysis.

B. Gaze at significant facial features

Both the mouth and the eyes are features of great visual interest during communication because of the information they transmit through gaze and speech. The face tracking software used in this system allows the tracking of these features. In order to better understand what aspects of the face people attend to during conversation, the timesteps during which the pairs were in mutual gaze (looking at each other's faces) were classified according to what features were attended to. The regions of interest examined were:

- Mouth - this area is defined by the outside of the lip contour
- Eye - this area contains two separate regions for both the left and right eye
- Face - this area is defined as all of the face other than the mouth or eyes

The resulting percentages of mutual gaze time are shown in Figure 5. The vast majority of mutual gaze was made up of the pairs looking at somewhere other than the significant features on the face (the Face-Face state). The next most common state was ”Mouth-Face”, where one person looked at the other's mouth. The only other state that occurred with regularity for all of the pairs was ”Eye-Face”, though this was far less common than looking at the mouth.

Perhaps the most surprising result from this data is the lack of eye contact. Of all the participants, only Pair 1 exhibited any eye contact at all (and it was momentary). Because the measurement limitations we discussed earlier have prevented the study of direct eye contact in psychology, we have no way of knowing whether this result is out of line with normal human behavior, though it violates our naive expectations about gaze. There are a few possible explanations for this result. One is that it is possible, given that the eyes are a small target in the scene camera's image, that minor calibration errors in the gaze tracker may prevent at-eye gaze from being registered as such. The fact that eye-directed gaze was registered (in both the Mouth-Eye and Eye-Face states) proves that it can be detected by the system, though it could still be underestimated. Another possible explanation is that people actually don't frequently look into the center of other's eyes, preferring instead to glance at the area surrounding them. Given previous research [7] suggesting that people may have problems correctly assessing

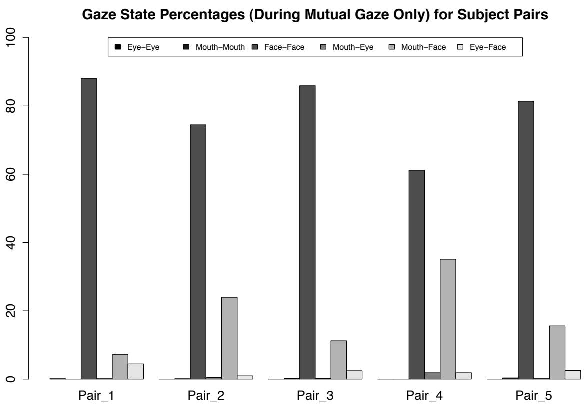


Figure 5. The average percentage of time spent by each pair attending to different facial features while in mutual gaze.

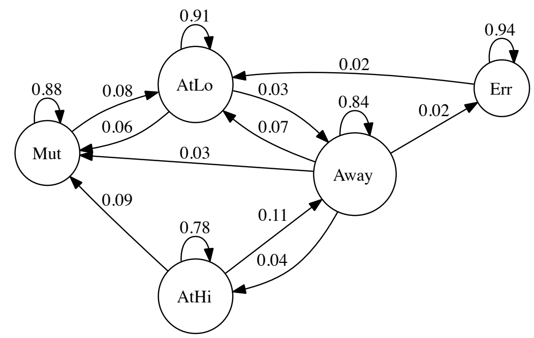


Figure 6. Markov model of the gaze state transitions for all pairs. For clarity, transitions of less than two percent probability are not displayed.

gaze direction even when it is directed at themselves, they may experience this as direct eye gaze. It is also possible that people typically exhibit more eye-directed gaze, and that the glasses used in the gaze tracking system cause people to fixate on the eyes of their partner less often than they would if their partner were not wearing the glasses. Artificially doubling the size of the eyes during analysis did not lead to a significant increase in the amount of eye contact detected (though it at least quadrupled the percentage of gaze classified as Eye-Face for each pair). So if near-eye gaze occurs frequently in the data either because of minor gaze tracker miscalibration or natural human behavior, it still does not explain the lack of eye contact registered. Future experiments will seek to further explore this unexpected result and verify that it is a real phenomenon and not a system limitation. Raw video will be recorded and analyzed offline to access the real-time system's accuracy.

Given that this was a conversational task, it is not surprising that the relative amount of mouth directed gaze was high. But it is unclear what is happening in the broadly defined Face-Face state. For controller design, the exact properties of gaze at seemingly non-significant parts of the face may make a difference between realistic and non-realistic gaze behavior. Further analysis will explore what areas of the face are being attended to and what the typical length and patterns of gaze fixations on them are.

V. Towards a model of gaze behavior

As a method of further analysis and as a first step towards using this data to implement a gaze controller for a robot, we created a Markov model of the interaction using data from all five pairs. A Markov model (or Markov chain) is a graphical probabilistic model that describes the state transitions of a system or process [20]. Data from the contiguous two minute period with the lowest error rate for each pair was combined to construct a model of their average behavior. This model is shown in Figure 6. Each gaze state of the interaction is a node in the model. The chance of reaching any other state from a given state at the next timestep is given by the probabilities on the outgoing edges from that state. The probability of staying in the same state at the next timestep is the probability of the state's edge that points back to itself. These self-transitions cause the time spent in each state to follow a geometric distribution, which agrees well with the form of the data observed. In order to improve the readability of the model and emphasize its major dynamics, transitions of less that 0.02 probability are not shown. From this model, it can be seen that gaze rarely alternates between being mutual and having both partners look away. Typically, one partner holds the face-directed gaze longer than the other. Also, note that the only state with a transition into the Err state with a large enough probability to be displayed is the Away state, which supports our hypothesis that tracking errors are non-uniformly distributed and most frequently occur when gaze is directed away from a partner's face.

VI. Conclusion

In this paper, a system for the real-time automated detection of mutual gaze was described, and results were presented from natural conversational interactions between human pairs. The overall level of mutual gaze observed was in line with predictions from the psychology literature on mutual gaze. But the durations of the mutual gaze episodes was far shorter. Additionally, an analysis of gaze at specific facial features found virtually no evidence of simultaneous direct eye contact between the conversational partners. These results highlight the potential for obtaining different, possibly more accurate measures of behavior using automated methods.

This real-time system is designed not purely for analysis,

 ISBN: 978-1-61208-117-5

but to provide gaze information as input to a controller for a humanoid robot in the future. As a demonstration of how we intend to use this human-human gaze data to produce a robotic gaze controller, we created a Markov model from the data collected and discussed how it captures the gaze behavior dynamics of the humans.

There are numerous opportunities for future work that could improve the sophistication and realism of conversational gaze controllers. This data has not yet been analyzed based on participants' conversational role (speaker or listener). We plan to do so in the near future, comparing the results we obtain with what is predicted by the psychology literature. Our resulting robotic gaze controller will use conversational role as an input in order to better support natural conversational interaction.

Another opportunity for analysis is to further examine people's patterns of gaze at facial features such as the eyes and mouth, as well as which non-significant areas of the face people gaze at during conversation. This could improve the life-likeness of the gaze actions taken by the robot. The human-human results for such an analysis is interesting in and of itself. There is a lack of results of this sort because of the difficulty of accurately measuring gaze in this manner without supporting technologies. Automated analysis could provide new insights into how humans interact through gaze as well as how they could interact with robots.

Acknowledgment

This research was conducted within the EU Integrated Project ITALK (Integration and Transfer of Action and Language in Robots) funded by the European Commission under contract number FP7-214668.

References

[1] M. Argyle, *Bodily communication*, 2nd ed. Routledge, 1988.

[2] S. M. Hains and D. W. Muir, "Infant sensitivity to adult eye direction." *Child development*, vol. 67, no. 5, pp. 1940–1951, October 1996.

[3] T. Farroni, "Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis," *Journal of Experimental Child Psychology*, vol. 85, no. 3, pp. 199–212, July 2003.

[4] C. Trevarthen and K. J. Aitken, "Infant intersubjectivity: Research, theory, and clinical applications," *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 42, no. 01, pp. 3–48, 2001.

[5] C. Kleinke, "Gaze and eye contact: A research review." *Psychological Bulletin*, vol. 100, no. 1, pp. 78–100, 1986.

[6] L. D'Odorico, R. Cassibba, and N. Salerni, "Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication," *Journal of Psycholinguistic Research*, vol. 26, no. 5, pp. 539–556, September 1997.

[7] M. v. Cranach and J. H. Ellgring, "Problems in the recognition of gaze direction," in *Social communication and movement : studies of interaction and expression in man and chimpanzee*, M. von Cranach, Ed. London: Academic Press, 1973, pp. 419–443.

[8] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "The effects of responsive eye movement and blinking behavior in a communication robot," in *IROS*, 2006, pp. 4564–4569.

[9] B. Mutlu, J. Forlizzi, and J. Hodgins, "A storytelling robot: Modeling and evaluation of human-like gaze behavior," in *Humanoids*, 2006, pp. 518–523.

[10] C. Yu, M. Scheutz, and P. Schermerhorn, "Investigating multimodal real-time patterns of joint attention in an hri word learning task," in *HRI '10: 5th ACM/IEEE international conference on Human-robot interaction*. New York, NY, USA: ACM, 2010, pp. 309–316.

[11] A.-L. Vollmer, K. S. Lohan, K. Fischer, Y. Nagai, K. Pitsch, J. Fritsch, K. J. Rohlfing, and B. Wredek, "People modify their tutoring behavior in robot-directed interaction for action learning," in *DEVLRN '09: Proceedings of the 2009 IEEE 8th International Conference on Development and Learning*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 1–6.

[12] J. Saunders, C. L. Nehaniv, and C. Lyon, "Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors," in *2nd Intl Symp. on New Frontiers in HRI, AISB*, 2010.

[13] S. Raidt, G. Bailly, and F. Elisei, "Analyzing and modeling gaze during face-to-face interaction," in *7th International Conference on Intelligent Virtual Agents, IVA'2007 7th International Conference on Intelligent Virtual Agents, IVA'2007*, ser. 17-19 September 2007, Paris, France, Paris France, 09 2007, pp. 100–101.

[14] G. Metta, P. Fitzpatrick, and L. Natale, "Yarp: Yet another robot platform," *International Journal of Advanced Robotics Systems, special issue on Software Development and Integration in Robotics*, vol. 3, no. 1, 2006.

[15] Applied Science Laboratories, "Mobile Eye gaze tracking system." [Online]. Available: http://asleyetracking.com/

[16] Seeing Machines, Inc., "faceAPI." [Online]. Available: http://seeingmachines.com/

[17] D. L. Mills, "Improved algorithms for synchronizing computer network clocks," *SIGCOMM Computer Communication Review*, vol. 24, pp. 317–327, October 1994. [Online]. Available: http://doi.acm.org/10.1145/190809.190343

[18] S. D. Gosling, P. J. Rentfrow, and J. Swann, W. B., "A very brief measure of the big five personality domains," *Journal of Research in Personality*, vol. 37, pp. 504–528, 2003.

[19] M. Argyle and R. Ingham, "Gaze, mutual gaze, and proximity," *Semiotica*, vol. 6, no. 1, pp. 32–49, 1972.

[20] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

Towards a General Communication Concept for Human Supervision of Autonomous Robot Teams

Karen Petersen and Oskar von Stryk
Department of Computer Science
Simulation, Systems Optimization and Robotics Group
Technische Universität Darmstadt, Darmstadt, Germany
{petersen|stryk}@sim.tu-darmstadt.de

***Abstract*—Towards a general concept for human supervision of autonomous robot teams supporting the specific strengths of humans and robots, a communication concept between robots and a human supervisor is presented in this paper. The communication goal is to let the supervisor control a robot team with high-level commands, e. g., by adapting mission details and influencing task allocation in a manner that is applicable to different task allocation methods in general. For this purpose, the supervisor needs a high-level overview of the current state of mission and robots, which can be obtained with the presented approach. Relevant, important events are detected by the robots using complex event processing, and are labeled by topic and priority. A policy system controls the amount of messages that are sent to the supervisor. Notifications are used to inform the supervisor about the mission progress, unexpected events and errors. Queries are used to transfer decisions to the supervisor, to make use of implicit knowledge and experience in critical situations. The robots' level of autonomy can be adapted using policies, that require decisions to be either taken autonomously by the robots, or with support by the supervisor, using different query modes. Example scenarios from different applications including urban search and rescue will be used for validating the proposed concept.**

***Keywords*-human-robot team interaction; supervisory control; complex event processing; policies**

I. Introduction

Teams of autonomous robots have the potential to solve complex missions such as Urban Search And Rescue (USAR), but are not yet sufficiently reliable and powerful to operate without any human supervision. However, humans and robots have many complementary capabilities, which can contribute significantly to a successful and efficient mission achievement if utilized properly. For example, humans are very good at cognitive tasks such as visual perception, coping with unfamiliar or unexpected situations, and prediction of future states of the world based on incomplete knowledge of the current state. Robots, in contrast, have their strengths, for example, in fast execution of well-defined or repetitive tasks, evaluation and storage of large amounts of data, or operation in areas inaccessible to humans, like narrow spaces, in the air, or contaminated areas. This complementarity has already been observed when comparing humans and machines almost 60 years ago (c. f. Section III-A), which leads to the assumption that this situation will not change significantly in the near future. Therefore, in human supervision of autonomous robots efficient use should be made of these specific strengths.

A. Abilities and Scenarios for Interaction

The proposed concept addresses scenarios, where a team of autonomous robots can be supported by a human supervisor with high-level instructions. The main goal of the robots (called *mission*) can be subdivided into tasks, that may be known prior to the mission start or can emerge during the mission, where each can be fulfilled by a single robot. A task allocation method is used to decide which robot works on which task. It should be noted that the proposed approach can be applied to very different task allocation methods. This can be either a centralized planner for the whole team, or a distributed algorithm, where the robots negotiate the tasks among each other. The choice of an appropriate task allocation algorithm depends on the concrete mission setup and environmental conditions.

Common teleoperation interfaces require one operator per robot, which implies that having a team of robots also requires a team of operators. If a single human supervisor shall be enabled to control a whole robot team, a fundamentally different approach is needed.

The main difference between a supervisor and an operator as defined in [1] is, that the supervisor usually interacts by specifying goals and intentions, while the operator interacts at the action level. High-level commands from the supervisor in a USAR mission may be used, e. g., to confirm or discard a robot's object hypotheses (e. g., victims or fire), to classify terrain trafficability, or to specify regions that are to be searched first or for a second time by a specific robot. In a robot soccer scenario, supervisor interactions may include changing or adapting a team's tactic, or allocating specific roles to individual robots. Common for all applications is, that the supervisor should be enabled to modify the tasks' parameters and the allocation of tasks to robots, and to act as decision support for the robots, e. g., in case they are not granted sufficient authority or do not have sufficient information for good autonomous decisions.

 ISBN: 978-1-61208-117-5

B. General Concept

The goal of the presented concept is on the one hand to enable the supervisor to modify the mission's and tasks' details (including task allocation), and on the other hand to allow robots to transfer decisions to the supervisor, if the robots are not allowed or not able to decide autonomously. Mission and task allocation adaptations can be realized by introducing a layer, that modifies the cost calculated for executing a task, which is independent of the underlying task allocation algorithm. However, to enable the supervisor to take such decisions, he or she needs to be aware of the team's progress towards mission achievement and the current state of the world and the robots.

In research on teleoperation, the required knowledge about the robot's state and the environment is called Situation Awareness (SA), which is adopted from pilot situation awareness and measured with the same tools [2]. SA includes many details needed for teleoperation, that are not needed for high-level commands of a supervisor like the detailed pose of a robot, its detailed numerical distance to obstacles, full video streams of the cameras or live map-data. Therefore, we introduce the term *Situation Overview (SO)*, which includes more general knowledge about the world and the robots' status, e. g., the health of each robot (battery status, sensor and actuator functionality), the current and planned next task of each robot, and the overall mission progress. Maps and video data are only sent if needed for specific decisions, e. g., to decide if an image shows a potential victim or not. Details only needed for teleoperation are omitted.

The needed information is usually not fixed, instead, the communication should be adopted during runtime to the specific needs of different missions and supervisors. For this purpose, we propose to control the amount of information using policies, which define the events to be detected by the robots using complex event processing. These events are then classified according to their priority and are sent to the supervisor as notifications (to provide information) or queries (to ask for decision support).

A main advantage of SO over SA is the reduction of data that has to be communicated. This is an important factor in real-world applications where the available communication bandwidth is usually limited. SO, obtained with the presented methods, gives a human supervisor a basis for high-level team interactions, without overburdening the human with too specific information of single robots.

The rest of this paper is organized as follows: Related work is discussed in Section II. In Section III, first the interactions among humans in loosely coupled teamwork are observed. Second, inspired by these findings, the methods enabling the robots to send notifications and queries to the human supervisor are described. For detecting the important incidents, methods from complex event processing are applied. The messages are classified with different levels to allow filtering. The amount of messages can be controlled using policies, which can be adapted either manually or automatically, depending on the supervisor's workload. Some application examples, for general robot team applications and for concrete scenarios of USAR and robot soccer, are given in Section IV. The methods are discussed and future work is described in Section V.

II. Related Work

Especially in the USAR domain, much research has been done on teleoperation interfaces, e. g., [3], [4]. These strongly rely on video- and map-data, that need to be sent in real-time from the robot to the user interface. On the one hand, this allows to accurately control a robot even in unstructured and complicated environments, but on the other hand, those interfaces cannot be extended easily to control more than one robot simultaneously, and require high bandwidth, which is often not permanently available in real-world scenarios. Further, most teleoperation interfaces require extensive operator training and continuously demand maximum concentration of the operator, hence quickly leading to task overload and operator fatigue.

Approaches that allow a single supervisor to deal with robot teams and do not require continuous high bandwidth communication can be found in the area of sliding autonomy or mixed initiative. In [5], Markov models are used to decide whether a robot works on a task autonomously or is being teleoperated by an operator. This requires continuous communication connection only during the teleoperation phases. The mixed initiative system presented in [6] allows the operator to manually switch between autonomy modes, where the operator input varies from goal input to full teleoperation. Similarly, in [7], the operator can assign waypoints, move the camera, or completely teleoperate a robot. With the augmented autonomy approach used in [8], the robots in an exploration scenario can either select their next waypoints autonomously, or the operator can assign waypoints. Results show, that these methods are appropriate to deal with a larger number of robots and can produce much better results than purely autonomous or purely teleoperated systems. However, they still require periods of continuous communication connection between the operator and the robots, and most of them can hardly be extended to fundamentally different scenarios, where the main focus is not on search or exploration.

A completely different approach is described in [9], where the robots can ask questions to the human supervisor. Similarly, in [10], the human is treated as a source of information for the robots. The level of autonomy is controlled by adjusting the cost to contact the supervisor. The teleautonomous system presented in [11] enables the robots to detect situations where human intervention is helpful, which are in this context the states of robot stuck, robot lost or victim found. Human supported decision taking is presented in [12], here two variants are proposed: management-by-exception, where the operator can veto against an autonomous decision, and management-by-consent, where the operator needs to confirm an autonomous decision before execution. In [13], policies are used to restrict the autonomy bounds of the robots, in this context also rules are defined about which messages the robots are required to send to the human. These approaches are promising to be applicable to larger robot teams in real-world environments, because they do not require continuous human attention to a single robot and require less bandwidth as they do not rely on video streams. However, they are still not very flexible to be adapted to fundamentally different scenarios or for on-line adaption to different operator preferences. Furthermore, the events that require operator intervention are detected manually, and yet no method has been provided to flexibly detect complex events in arbitrary complex situations.

III. Control of Communication between Robots and a Supervisor

The concept presented in this section is inspired by teamwork among humans, which is described briefly. Afterwards, the specific strengths of humans and robots, that contribute to these kinds of scenarios and interactions are revised. Finally, the methods used to realize a flexible communication between the robots and the supervisor are presented and discussed.

A. *Interactions in Team Work Among Humans*

When observing interactions in loosely coupled workgroups [14], some commonalities can be observed regardless of the scenario, e.g., home care, knowledge work, firemen in a search and rescue scenario, soccer players coordinating with each other and getting instructions from a coach, or people in an office preparing an exhibition at a fair: In all these situations, the overall mission is first subdivided into tasks, that are assigned to the individual team members. Every participant works on his or her tasks autonomously, and reports the progress to the teammates or the leader, either explicitly by verbal or written communication, or the progress can be directly observed by the others. Whenever an individual has problems in fulfilling a task, he or she can ask somebody else (who is expected to be more capable for this specific problem) for support.

To understand the benefits of supervisory control, it is important to be aware of some fundamental differences between humans and robots. In [15], the superiorities of humans among machines and vice versa are discussed. One of the main outcomes is that machines are good in fast routine work, computational power and data storage, while humans' strengths are perception, reasoning, and flexibility. These findings (although almost 60 years old!) are in most points still valid and can be transferred to a large extent from machines to robots. Especially the superiority of humans over robots in problem solving and situation overview is crucial, and does not seem to change in the near future. Further, although there are several sensors that allow robots to perceive data that humans cannot sense directly (e. g., distance sensors, infrared sensors), humans are much more capable in interpreting data, especially images.

As a conclusion, if a human supervisor is aware of the overall situation, but not necessarily of all details, it does make sense to leave some high-level decisions to the human, because he or she can be expected to decide based on implicit knowledge, situation overview and experience, that cannot easily be added to the robots' world model. Due to the complementary capabilities of robots and humans, it can be expected that humans can cope well with the problems that robots cannot solve autonomously.

If this model of human teamwork is applied to human robot interaction, with the human taking the role of a supervisor, the robots are required to report their progress and unforeseen events to the human, and ask for support if they cannot solve their tasks sufficiently well autonomously. This is enabled by using three methodologies: First, important or critical events are detected using complex event processing(Section III-B). Second, the detected events are classified to message classes, which are different levels of notifications and different query modes, according to their criticality (Section III-C). Third, the message flow is controlled by policies, that define which messages need (not) to be sent to the supervisor (Section III-D).

B. *Complex Event Processing*

The events to be detected by the robots can be very diverse to many aspects. Some are just special variables exceeding thresholds, others are patterns that have to be detected, or several occurrences of different events simultaneously. Certainly, the detection of every single event could be programmed manually, but this is very time consuming, can lead to many failures, and usually duplicates lots of code.

The research field of Complex Event Processing (CEP) deals exactly with this question, of how to detect events in communication systems [16], for example in databases or Wireless Sensor Networks (WSNs). In WSNs, the challenge is to use several hundreds of distributed sensor nodes to detect events, e. g., human presence or fire, and combine simpler events to detect complex events, that are aggregations or patterns of several events. *Simple events* are discrete events, that can be directly detected without aggregating more information, e. g., a variable exceeding a threshold, or a sensor (not) delivering data. *Complex events* are events that are composed of two or more (simple or complex) events, or events enhanced with external information. These compositions can be two events occurring simultaneously, an event chain, patterns, etc. To describe those aggregations, event algebras are used, e. g., HiPAC [17], SNOOP

 ISBN: 978-1-61208-117-5

[18], REACH [19]. Those algebras provide operators as conjunction, disjunction, sequence, etc., to combine two or more events to a complex event. They vary in complexity and versatility. Depending on the application, an appropriate algebra needs to be chosen, that satisfies all needs, but is not too complex, hence being more difficult to understand and leading to higher implementation efforts.

The analogy between CEP as used in WSNs and robotics is, that there are several sensors and pre-processed data available, based on this information certain events or states of the robot or the world have to be detected. The key differences are, that a robot has less, but more reliable sensors than in a WSN, and the "network" is static, apart from sensor failure. Further, a robot's sensors are not entirely distributed, they are all physically connected, therefore issues like time synchronization and timeliness can be disregarded for CEP on robots. If also events are considered that involve more than one robot, the team can be seen as a WSN. Overall, CEP provides good methodologies, that can be used efficiently not only in databases and WSNs, but also on robots.

To allow efficient filtering, events can be tagged. E. g., in a search and rescue mission, there could be events related to victim detection, events related to simultaneous localization and mapping (SLAM), or events related to the vehicle's health as the battery status.

Some examples of important events in a USAR mission are of course if a robot has detected a potential victim or a fire, but also reports about the status of the exploration, e. g., if a room has been explored completely without finding a victim. In a humanoid robot soccer match, a robot can monitor the frequency of falling when walking or kicking, taking into account disturbances by teammates or opponents (e.g., by pushing), and can deduce if it is still capable of playing efficiently. The goalkeeper can monitor its benefit to the match, if it observes the frequency of jumping to catch the ball, compared to the number of goals scored by the opponent, i. e., if the goalkeeper jumps for the ball, and no goal is scored directly afterwards by the opponents, the team presumably benefits from the goalkeeper. If the opponents score, regardless of the goalkeeper jumping or not, the robot can potentially contribute more to the team's success when acting as a further field player.

C. Message Classification

The supervisor shall be supported – and not confused – by the messages from the robots. To enable the user interface to prominently present critical messages and show other information when needed to obtain SO, the messages are classified according to their importance and criticality. The queries are graded with different modes of action selection, depending on the desired degree of robot autonomy.

Proposed Levels of Notifications: Usually, logging systems for software development use five stages: debug, information, warning, error, and fatal. The concept provided here targets users unfamiliar with the implementation details of the robot control software, hence the debug-level can be omitted here, as these notifications would confuse the supervisor, instead of advancing the SO. Fatal are usually those errors, that cannot be handled properly and lead to program termination. Because these notifications cannot be communicated, also the fatal-level is omitted here.

In summary, there remain three notification levels, to be used by the robots: *information*, representing regular events (e. g., start or termination of execution of a task), *warning*, representing unexpected but noncritical events (e. g., task execution takes longer than expected), and *error*, representing critical events (e. g., sensor failure).

As examples for notifications, an information can be sent by a USAR robot, informing the supervisor that it has finished exploring a room without finding any victims. A warning can be sent by a soccer robot, that detects that it falls frequently without external influence and therefore cannot play properly. An error should be sent by a robot that detects that an important sensor, e. g., the camera or the laser range finder, does not deliver any or sufficiently meaningful data.

Types of queries: Depending on the desired degree of robot autonomy, there are several possibilities to take decisions. Besides deciding and executing everything autonomously, also the supervisor can be integrated for confirming or vetoing decisions, or even for selecting the appropriate answer. Decisions that allow or require supervisor intervention are formulated as queries.

Three query classes are proposed:

(1) *Autonomous decision with veto:* The robot selects among several solutions, and does not start execution before a specific time t_{exec} has elapsed. The supervisor is given a time t_{veto} to contradict this decision. t_{exec} and t_{veto} are independent of each other, which means, if $t_{exec} < t_{veto}$, the supervisor can veto a decision even after the robot started execution, if $t_{exec} > t_{veto}$, the supervisor cannot abort execution after it started.

(2) *Autonomous decision with confirmation:* The robot selects among several solutions and presents the selected solution and the alternatives to the supervisor. Execution does not start before the supervisor confirms or contradicts the selection.

(3) *Supervisor decision:* The robot provides several solutions to the supervisor, but does not preselect a solution. Execution starts after the supervisor selects and confirms a solution.

The robots are granted more autonomy in the first class, and less autonomy if confirmation by the supervisor is required. The second and third query classes make no difference for the robots, but for the human there is a psychological difference if a selection is proposed or not.

As an example, the humanoid soccer goalkeeper of the previous paragraph is considered. If this robot detects that it is either not needed (because the opponents do not shoot

 ISBN: 978-1-61208-117-5

on the goal) or is not beneficial (because it cannot block the goal shots), the robot could instead act as an additional field player, to potentially contribute more to the team's success. Depending on how much autonomy is granted to the robot, this tactic change could either be autonomous with veto, or (to give the human more control) autonomous with confirmation.

D. Control of Message Flow

The amount of messages that are sent to the supervisor needs to be controlled carefully. On the one hand, too many messages can result in information overflow and supervisor stress, or in complacency if most of the robot decisions are trivial, which brings the danger of overseeing wrong decisions. On the other hand, too few messages lead to a loss of SO. In general, there should not be any static rules about which events shall be communicated to the supervisor, and which decisions the robot should take autonomously or with some support by the supervisor. Rather, this is highly dependent on the current mission, the supervisor's preferences, and the supervisor's trust in the system.

In [13], policies are used to define the bounds of an agent's autonomy. Policies are positive and negative authorizations, that define what an agent is (not) allowed to do, and positive and negative obligations, that define what an agent is (not) required to do. Policies are applied to actions as well as to communication, e. g., sending acknowledgments when receiving new instructions. In the scope of this paper, the only bound on autonomy is decision taking, therefore it is sufficient to apply similar rules to regulate the amount of notifications and queries.

By means of the tagged events defined in Section III-B and the different messages classes defined in Section III-C, policies can be defined for groups of messages, according to their importance, or according to a topic, or for single event types.

Sets of policies can be loaded, dependent on the current mission, or even situation dependent. Further policies can be defined by the supervisor. Finally, policies can be adapted automatically, depending on the supervisor's current workload. For example, if there is a number of pending queries, only queries with supervisor selection or supervisor confirmation should be sent, because queries with supervisor veto can be expected to expire before the supervisor notices them.

As an example in the USAR scenario, a supervisor without trust in the robots' autonomous victim detection might want to get informed every time human-like temperature is detected with a thermal sensor, while a supervisor with more trust might be satisfied getting just the hypotheses that are positively verified by the robots. For the soccer scenario, if the supervisor is occupied with notifications about malfunctioning sensors or instable walking abilities, the queries about tactics changes can be omitted, because they are just of secondary importance if so many other problems have to be handled.

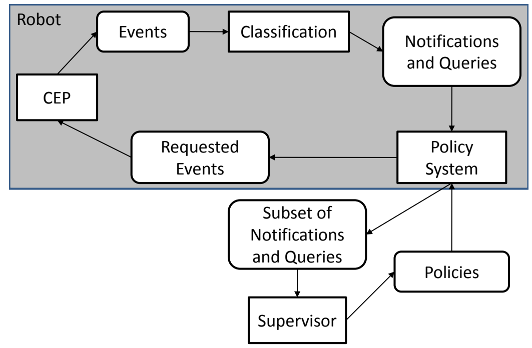


Figure 1. Visualization of the interactions among the different components of the proposed general communication concept

E. Discussion

All three methods applied here have been well established in entirely different fields. The new concept is, to combine them, and to use them to enable a human supervisor to obtain situation overview on a high level.

Overall, the three components are connected in a loop with external feedback from the supervisor, as shown in Figure 1: CEP detects important events, which are then classified to notification levels or query types. The policy system then decides which of those messages are sent to the supervisor. For closing the loop, the policies can be adapted during runtime, either manually by the human or automatically, and therefore it changes dynamically, which events have to be detected by the CEP system. Compared to other approaches, the presented concept allows a more flexible communication, that allows to control on the one hand the supervisor's workload, and on the other hand also the use of network capacity, if low bandwidth is an issue.

IV. Application Examples

In this section it is demonstrated, how the proposed approach can be applied to different scenarios from different applications. First, some general examples are given, that apply to arbitrary robot missions in general. Second, first steps of the integration of the concept into our USAR robot and our humanoid soccer robots are outlined.

A. Mission-independent Examples

With most robot user interfaces, the supervisor needs to be familiar with the system for deciding which system functions need to be monitored, and how they can be monitored. With the methods proposed in this paper, the robots provide methods to monitor themselves and can on the one hand inform the human about the status, and on the other hand send warnings if the status changes or is critical.

Before the start of a mission, all important sensors have to be checked for functionality. Instead of doing every check by

 ISBN: 978-1-61208-117-5

hand – which is often omitted or only done for some samples to save time – this can be done automatically using CEP. A successful check results in an event of the type ”successful sensor check”, which is sent as information message. If a check fails, an event of the type ”sensor failure” is sent as error message, accompanied with an error description.

The battery status of every robot should be monitored continuously, to prevent malfunctions because of too low voltage or damaged batteries. Battery displays for each robot can be overlooked, especially if a single human has to monitor the battery status of many robots in parallel to several other monitoring or coordination tasks. With the methods presented in this paper, each robot can monitor its battery status individually, and can send a warning notification before the battery runs empty. The methods of complex event processing further allow to warn not before the voltage is constantly below a threshold for some seconds, and therefore is able to filter voltage peaks or faulty measurements.

As a general proposal, an unexperienced supervisor should start with no restricting policies, and then gradually constrain the messages, if they are not needed. On the one hand, this leads to lots of messages at the beginning, but on the other hand, the supervisor learns, which types of events are provided by the robots and can decide on this basis which messages are important for the current setup.

B. Example: Urban Search and Rescue

In the USAR setup, a team of heterogeneous, autonomous robots has to search for trapped victims in a partially collapsed building, e. g., after an earthquake, and to locate potential hazards like gas leaks or fire. The methodologies proposed in this paper apply to robot behavior that can be observed for example at RoboCup rescue, whereas in current real-world deployments the robots are not yet autonomous at all, and the proposed concept requires robot autonomy as a starting point.

The results are discussed for the Unmanned Ground Vehicle (UGV) of Team Hector Darmstadt [20] (Figure 2(a)). This robot can autonomously explore an environment, build a map and search for potential victims and markers that indicate hazardous material (hazmat signs, see Figure 2(b)). Although at RoboCup the simulated victims are not yet too sophisticated (baby dolls with electric blankets, as can be seen in Figure 2(c)), and can be detected by only using a thermal sensor (see Figure 2(d)), we use a sensor fusion algorithm, that combines victim hypotheses from the daylight camera and the thermal camera with information from the laser range finder to build up a semantic map [21]. This algorithm is also suitable for more realistic conditions than RoboCup, i. e., to reliably find real people in environments that also contain other heat sources or shapes similar to humans.

With most user interfaces, the supervisor requests the information he or she believes to be of interest. The only information that usually pops up automatically is, when the robot states to have found a victim. In this case, some important data is potentially not monitored, or much information is sent continuously, even if it is not needed. For example, the operator requests the map generated by the robot, the robot's position in the map, and the camera images, but does not have a look at the output of the thermal sensor. If this sensor has a malfunction, it is potentially never noticed. We propose instead, to automatically provide the operator with relevant information.

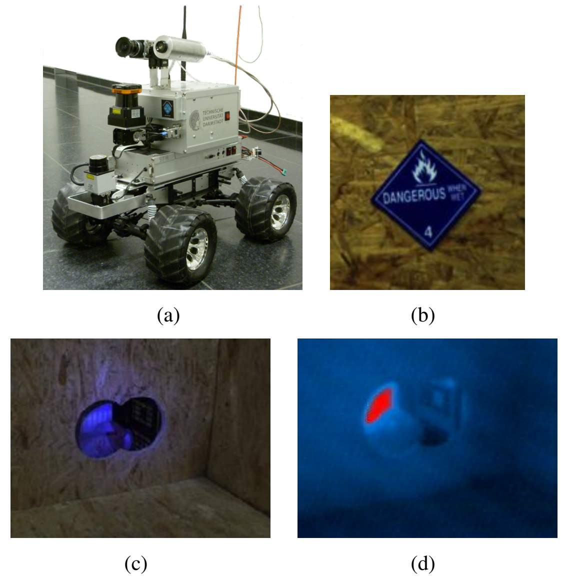
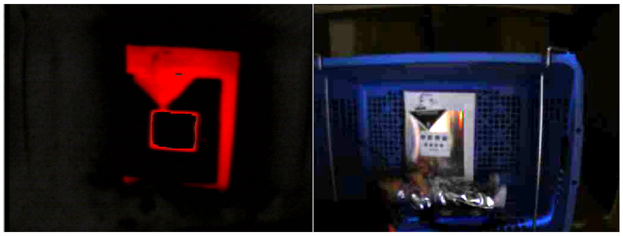


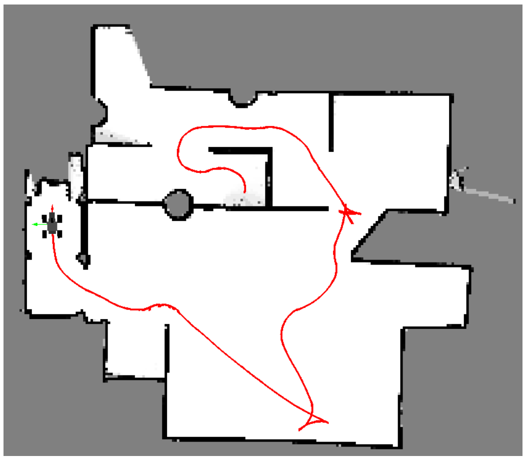
Figure 2. (a) Hector UGV. (b) Example of a hazmat sign. (c) Example of a simulated victim at RoboCup. (d) Thermal image of the simulated victim.

Mission progress is usually monitored by looking at the camera images and the map in real-time. However, as the robots usually do not proceed very fast, not all information is needed all the time. With the methods provided here, it is possible to send an image of the map every time a progress is observed. Progress can be, for example, every time the robot traveled more than 3 meters, or every time a robot enters or leaves a room, which results in an event of, e. g., the type ”entering room”, labeled to the general topic ”progress”, and is published as information message. In addition, every time a robot starts exploring a new victim hypothesis, this can be communicated, possibly together with attached sensor data that motivated the victim hypothesis. This results in an event of the type ”explore victim hypothesis”, labeled as related to ”victim” and ”progress”, and is published as information message. In turn, this method also allows to detect a lack of progress (by observing that no progress events are detected for a predefined time, although the robot intends to move), which can indicate a malfunctioning or disoriented robot. This is an event of the type ”no progress”, labeled with the topic ”progress” and should be published as a warning message. A supervisor who trusts in the robot's

 ISBN: 978-1-61208-117-5



(a)



(b)

Figure 3. (a) Images showing a simulated victim in the thermal image and the camera image. (b) The current map learned by the robot.

capabilities may not want to see all progress messages, but only those that refer to non-progress. To achieve this, he or she can define two policies: a negative authorization for sending notifications labeled with ”progress”, and a positive obligation for sending notifications of the type ”no progress”.

If a robot detects a victim, the resulting event is a supervisor decision query, where the supervisor can decide to (a) confirm the victim, (b) discard the victim, or (c) try to collect more information. The message also contains images from the cameras (see Figure 3(a)), and an image of the current map to display the location of the victim (see Figure 3(b)). The red line in Figure 3(b) shows the robot's traveled path. It can be seen that continuous monitoring of the map does not give more information to the supervisor than an image of the map every time a progress is observed or if the robot got stuck for a while, as it was the case in the upper right corner. Therefore, much communication overhead can be saved by omitting data transmissions that do not advance the supervisor's SO.

Queries can not only be used to let the supervisor confirm or discard potential victims or hazards, but also for decision support regarding path planning, e. g., an autonomous decision with veto can be sent, if the terrain classification is not confident enough and the supervisor should decide if a robot can negotiate an area or not.

C. Example: Humanoid Robot Soccer

At RoboCup soccer matches, usually monitoring is done by visually observing the match, at the team Darmstadt Dribblers [22] also the team messages sent between the robots are monitored. As direct human intervention is not allowed by the rules, the proposed concept can be used either for tuning the robots in practice games, or for changing details during game breaks.

Monitoring the health of each robot could also be done visually, but with three or more robots on the field it is difficult to keep track of each robot's performance. With CEP, it is possible to monitor the falling frequency of each robot for different motions like walking or kicking, and its correlation with other factors like the vicinity of opponents or teammates (which could indicate that the fall was due to a collision), or motor temperature and battery status, which could be a reason to switch to a behavior that consumes less energy, e. g., dribbling instead of kicking the ball.

Further, the benefit of the specific roles can be monitored, like already proposed for the goalie in Section III. This allows on the one hand to tune the parameters during tests to maximize each role's benefit, and on the other hand to quickly change tactics or preserve hardware during a match.

V. Conclusion and Outlook

The communication concept presented in this paper is designed for interactions between a human supervisor and a team of autonomous robots. To make use of the specific complimentary strengths of humans and robots, supervisor interactions are focusing on high-level commands. As a basis for high-level decisions, the supervisor needs SO, which can be flexibly achieved for several fundamentally different scenarios using the presented methods, which are complex event processing, message classification, and policies. This communication concept is inspired by loosely coupled human teamwork and requires a low communication overhead compared to standard teleoperation methods, because only data needed for SO are sent, while omitting details only used for exact teleoperation. The methods enable a human supervisor to gain a general SO of a whole robot team, without requiring the supervisor to be familiar with implementation details. The performance of the team can be enhanced by transferring critical decisions to the supervisor, because in this case the decision is based on SO, human experience and implicit knowledge, and is therefore expected to be more reliable and efficient for achieving the mission goal.

In general, an interface that is based on this new communication concept can provide a higher SO than standard interfaces, because the robots can send information that the supervisor would probably not request, hence problems and errors can potentially be noticed earlier. SO gives the supervisor a basis to take high-level decisions, e. g., for adapting task allocation or mission details. Preliminary results in USAR and robot soccer indicate the potential of the

 ISBN: 978-1-61208-117-5

developed concept. These two fundamentally different setups demonstrate, that the proposed concept can be applied to a large variety of problem classes. It is furthermore planned to implement the whole concept, including the communication concept as well as high-level commands by the supervisor, for different scenarios with heterogeneous robot teams.

Future work includes experiments in simulation and with real robots to support the hypotheses of this paper. It is planned to conduct user studies in different application scenarios to show the wide applicability of the proposed methods. Further, the possibilities of the supervisor to coordinate robot teams based on the proposed situation overview will be examined. For dealing with larger robot teams, a basis for a large-scale interface is provided by the presented concept, as it offers the data for SO and supports efficient filtering.

Acknowledgments

This research has been supported by the German Research Foundation (DFG) within the Research Training Group 1362 “Cooperative, adaptive and responsive monitoring in mixed mode environments”.

References

[1] J. Scholtz. Theory and evaluation of human robot interactions. In *Proc. of the 36th Annual Hawaii Intl. Conference on System Sciences*, 2003.

[2] M. R. Endsley. Situation awareness global assessment technique (sagat). In *Proc. of the IEEE 1988 National Aerospace and Electronics Conference*, volume 3, pages 789 – 795, 1988.

[3] M. W. Kadous, R. Sheh, and C. Sammut. Effective user interface design for rescue robotics. In *Proc. of the ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 250–257, 2006.

[4] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks. Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics and Automation*, 23(5):927–941, October 2007.

[5] B. P. Sellner, F. Heger, L. Hiatt, R. Simmons, and S. Singh. Coordinated multi-agent teams and sliding autonomy for large-scale assembly. *Proc. of the IEEE - Special Issue on Multi-Robot Systems*, 94(7):1425 – 1444, July 2006.

[6] Jacob W. Crandall and Michael A. Goodrich. Experiments in adjustable autonomy. In *IEEE Intl. Conference on Systems, Man, and Cybernetics*, volume 3, pages 1624 –1629, 2001.

[7] J. Wang and M. Lewis. Human control for cooperating robot teams. In *Proc. of the ACM/IEEE intl. conference on Human-robot interaction*, pages 9 – 16, 2007.

[8] Y. Nevatia, T. Stoyanov, R. Rathnam, M. Pfingsthorn, S. Markov, R. Ambrus, and A. Birk. Augmented autonomy: Improving human-robot team performance in urban search and rescue. In *Proc. of the IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, pages 2103 – 2108, Nice, France, Sept, 22-26 2008.

[9] T. Fong, C. Thorpe, and C. Baur. Robot, asker of questions. *Robotics and Autonomous Systems*, 42:235 – 243, 2003.

[10] T. Kaupp and A. Makarenko. Measuring human-robot team effectiveness to determine an appropriate autonomy level. In *Proc. of the IEEE Intl. Conference on Robotics and Automation*, pages 2146 – 2151, Pasadena, CA, USA, 2008.

[11] R. Wegner and J. Anderson. Balancing robotic teleoperation and autonomy for urban search and rescue environments. In *17th Conference of the Canadian Society for Computational Studies of Intelligence*, Lecture Notes in Computer Science, pages 16–30. Springer, 2004.

[12] M. L. Cummings, S. Bruni, S. Mercier, and P. J. Mitchell. Automation architecture for single operator-multiple uav command and control. *Intl. Command and Control Journal*, 1(2):1 – 24, 2007.

[13] M. Johnson, P. J. Feltovich, J. M. Bradshaw, and L. Bunch. Human-robot coordination through dynamic regulation. In *Proc. of the IEEE Intl. Conference on Robotics and Automation*, pages 2159 – 2164, Pasadena, CA, 2008.

[14] D. Pinelle and C. Gutwin. A groupware design framework for loosely coupled workgroups. In *Proc. of the Ninth European Conference on Computer-Supported Cooperative Work*, pages 65 – 82, Paris, France, 18-22 September 2005.

[15] P. M. Fitts, editor. *Human engineering for an effective air-navigation and traffic control system*. Washington, DC: National Academy of Sciences Archives, 1951.

[16] A. Hinze, K. Sachs, and A. Buchmann. Event-based applications and enabling technologies. In *Proc. of the Third ACM Intl. Conference on Distributed Event-Based Systems*, pages 1:1–1:15, 2009.

[17] U. Dayal, A. Buchmann, and D. McCarthy. Rules are objects too: A knowledge model for an active, object-oriented database system. In Klaus Dittrich, editor, *Advances in Object-Oriented Database Systems*, volume 334 of *Lecture Notes in Computer Science*, pages 129–143. Springer, 1988.

[18] S. Chakravarthy and D. Mishra. Snoop: an expressive event specification language for active databases. *IEEE Data and Knowledge Engineering*, 14(10):1 – 26, 1994.

[19] H. Branding, A. Buchmann, T. Kudrass, and J. Zimmermann. Rules in an open system: the REACH Rule System. In *Proc. of the 1st Intl. Workshop on Rules in Database Systems*, pages 111 – 126, Edinburgh, Scotland, September 1993.

[20] Team Hector Darmstadt website: http://www.gkmm.tu-darmstadt.de/rescue/, December 2010.

[21] J. Meyer, P. Schnitzspan, S. Kohlbrecher, K. Petersen, O. Schwahn, M. Andriluka, U. Klingauf, S. Roth, B. Schiele, and O. von Stryk. A semantic world model for urban search and rescue based on heterogeneous sensors. In *RoboCup Symposium, Singapore*, 2010.

[22] Team Darmstadt Dribblers website: http://www.dribblers.de/, December 2010.

 ISBN: 978-1-61208-117-5

Disentangling the Effects of Robot Affect, Embodiment, and Autonomy on Human Team Members in a Mixed-Initiative Task

Paul Schermerhorn and Matthias Scheutz
Human Robot Interaction Laboratory
Indiana University
Bloomington, IN, USA
{*pscherme,mscheutz*}@*indiana.edu*

***Abstract*—Many future robotic scenarios will require robots to work with humans in teams. It is thus critical to ensure that those robots will be able to work effectively with humans. While various dimensions of robots such as autonomy, embodiment or interaction style have been investigated separately, no previous study has looked at those three dimensions together. In this paper, we report results from extensive experiments showing that all three dimensions interact in complex ways, thus demonstrating the insufficiency of exploring these dimensions individually. Based on the results, we conclude with suggestions for interaction designs and for future studies.**

***Keywords*-human-robot interaction; adjustable autonomy; embodiment; robot; simulation; affect; user study**

I. Introduction

Mixed-initiave scenarios where robots have to work with humans in teams are among the main applications envisioned for future robots. Hence, it is important to explore the different dimensions of human-robot interaction (HRI) that might have an impact on team performance in mixed-initiative tasks. Among the natural candidates are *robot capability* (the degree to which the robot can contribute to the team task), *robot embodiment* (the particular appearance and physical instantiation of the robot) and *interaction style* (the different ways in which the robot can communicate with humans).

While previous HRI studies have investigated each of these dimensions in a variety of setups individually, no study was designed to explore all three dimensions together [1] [2] [3]. Consequently, previous studies are silent about possible interactions and tradeoffs among those dimensions. And because designs, experimental procedures, and evaluations differ significantly across studies (in addition to task-based differences), it is impossible to use their results for deriving potential interactions among multiple dimensions. Yet, knowing whether any such interaction exists is not only critical for the design of future robots, but also for contextualizing previous findings. For it is quite possible that two studies did not find any main effects for the dimension they investigated individually, even though there is a significant interaction between those dimensions which can only be revealed in an experimental design that allows for variations of both.

In this paper, we present results from the first large study that systematically investigates the tradeoffs and interactions among “robot capability”, “robot embodiment”, and “interaction style” using both objective and subjective performance measures in a 2x2x2 between-subjects mixed-initiative human-robot interaction design using a simple cooperative exploration task. For “robot capability”, we contrast a robot that can autonomously navigate through its environment and find task-critical locations with a robot that entirely relies on human instructions. For “robot embodiment”, we contrast a physical robot co-located in the subject’s environment with a graphical representation of a robot in a simulated environment displayed on computer screen. And for “interaction style”, we contrast a robot capable of expressing affect in its voice to indicate urgency with one that does not modulate its voice. An analysis of the results shows that there are important interactions among the three dimensions that previous studies missed.

The remainder of the paper is structured as follows. We start with a more detailed motivation of the three dimensions, including a brief summary of some of the mono-dimensional findings from previous studies. Next, we introduce our experimental setup, including the employed robot and control architecture, as well as all experimental materials and procedures. Then we report and analyze the results, and conclude with a brief summary of our findings, the implications for mixed-initiative HRI and directions for future work.

II. Background and Motivation

The influence of appearance and, to a lesser extent, embodiment has been investigated in several HRI studies. Often, these studies involve having subjects watch videos of robots with differing appearance and respond to questions regarding (e.g., empathy for robots [1]). Such studies seem to implicitly assume that embodiment will not influence these responses, but others do compare interactions with robots and other entity types (e.g., computer-based agents [2]). One dimension of the present work is embodiment, comparing a simulated robot with a physically embodied robot using exactly the same robot architecture.

 ISBN: 978-1-61208-117-5

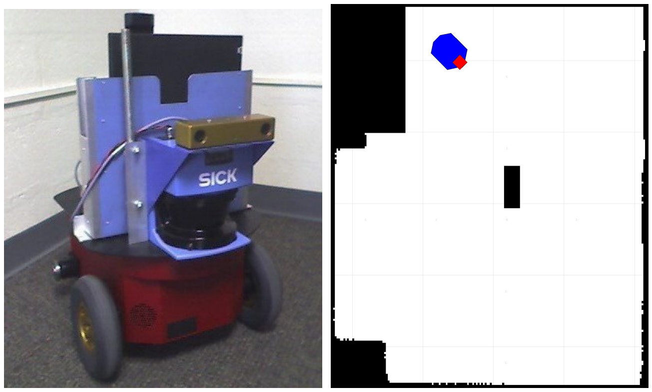


Figure 1. The two robots used in the experiments: real robot in real environment (left) vs. simulated robot in simulation environment (right).

Adaptive autonomy has been the focus of a great deal or research—too much to summarize here. Most closely related to the work presented here are projects in which mechanisms for adjustments to the level of autonomy are accessible to the robot architecture itself (e.g., to adapt to conditions of operator neglect [3]). Other work, on which this study is based, investigated whether people would be accepting of a robot switching to autonomous behavior, even to the extent of disobeying orders from the human, if the outcome of the robot's actions increased the chances for group success, and found that subjects did accept autonomy under those conditions [4].

Humans are clearly attuned to affective signals in others, hence several HRI projects investigated the utility of affect for signaling the robot's internal states (e.g., [5], [6]). In the context of mixed-initiative tasks, it was demonstrated that affect expression by the robot at the right time can improve team performance [7], although it was not clear how this performance gain is dependent on other factors.

The goal of the present study is thus to investigate whether and how these different dimensions interact in an effort to further our understanding of possible design tradeoffs. These insights will then allow us to develop better and more intuitive robots for mixed-initiative tasks.

III. Experimental Design

Rather than simply observing a robot (in person or watching videos) and providing subjective evaluations of the behaviors exhibited, participants were required to interact with the robot entity for a period of time. Moreover, these interactions were not open-ended, unstructured conversations; rather, subjects were required to work with the robot to achieve a goal. This provides us with the opportunity to gather objective performance measures, in addition to subjective evaluations, from each subject.

The task posed to the human-robot team was to explore a region and gather data about it. Specifically, subjects were told to imagine that they were part of a team exploring the surface of a remote planet. The objective was to measure rock formations in the environment and transmit information about the measurements to an orbiting spacecraft. However, due to signal interference, it was only possible to transmit from specific regions that changed for each trial. The robot was equipped with a sensor that could detect the strength of the signal, and the human was instructed to direct the robot in a search of the environment using natural language commands ("turn right," "go straight," etc.) and search for a strong signal by asking the robot to take a reading of the strength at its current position. Subjects completed three time-limited trials of 3.5 minutes each, and transmission to the orbiting spacecraft was allowed only in the final minute of a trial. The robot announced the time remaining every 30 seconds, so subjects did not need to keep track of or ask the robot for the time remaining. In the last minute, subjects were to command the robot to initiate the transmission sequence, at which point the robot would ask them two short questions about the measurements they had made (see below) and begin "transmitting" the results. If the signal strength was strong enough to transmit (i.e., if the robot was close enough to the target location for that trial) and the transmission was started sufficiently early (they were told in advance that transmission takes 15 seconds), the transmission was regarded as successful. If any of those conditions were not met (the signal was too weak, the subject did not answer the robot's questions, or time ran out), the trial was a failure.

In addition to directing the robot to find a transmission point, subjects were required to make "measurements" of "rock formations" in the environment. In actuality, the rock formations were a set of boxes containing sheets with 2-digit by 2-digit multiplication problems. To measure a formation, the subject opened the box, copied the problem onto a worksheet provided to them (along with a clipboard and pencil), and performed the multiplication. There were three sets of boxes (one set for each trial: blue, pink, and green), with five boxes in each set, labeled from 'A' to 'E'. Subjects were instructed to complete as many boxes as they could in the time given, working in alphabetical order. The purpose of the measurement task was to impose a cognitive load on subjects, such that they needed to decide how to allocate their attention between that and the search task. However, they were explicitly told that successful transmission was the highest priority (e.g., transmitting information about a single formation was a successful run, but completing all five formations while failing to transmit was a failure).

Figure 1 depicts a map of the experiment environment: a roughly 5x6m room with a single obstacle in the center. The "rock formations" were placed around the perimeter of the room and beside the center obstacle, to ensure they did not interfere with the robot's motion. Two *embodiment* conditions were tested, *robot* and *simulated*. In the robot

 ISBN: 978-1-61208-117-5

embodiment condition, the subjects interacted with a physical robot (a MobileRobots Pioneer P3AT; left in Figure 1) co-located with them in the exploration environment. In the simulation embodiment condition, subjects interacted with a 2-D simulated robot (in the Stage simulator [8]; right in Figure 1). Regardless of condition, subjects were given time to interact with the robot in a trial run context to learn the robot's abilities and limits.

Care was taken to ensure that the only difference between the two embodiment conditions was the physical presence or absence of the robot. The layout of the simulated environment is the same as that of the physical environment, and the transmission locations were the same in the two conditions. The DIARC architecture [7] used to control the robot, perform natural language understanding and speech production, etc., was the same, with the exception of the component representing the physical robot. The robot entity's decisions, abilities, and responses were identical in both embodiment conditions. Hence, any performance differences between subjects in the robot and simulation conditions can be attributed to embodiment.

Subjects were assigned to one of two *affect* conditions, *affect* or *no-affect*. In the no-affect condition, the robot performs exactly as it did during practice. The affect condition was also exactly the same, with one exception: partway through each trial, the speech production component would begin modulating the voice to express increasing levels of stress as the trial deadline approached. Speech generation was unaffected (i.e., the utterances were generated in the same in way in both affect conditions, so the content was the same as it would be regardless of affect), and the affective state was "purely cosmetic" (i.e., it did not influence decision-making or action execution in the robot). Hence, any performance differences between subjects in the affect and no-affect conditions can be attributed solely to the expression of affect ("stress").

The search task was explicitly designed to be very challenging; this was to ensure that we would see a performance difference based on the *autonomy* condition. In the *no-autonomy* condition, the subject has to direct the robot to find the transmission point, just like during practice. The *autonomy* condition introduces into the robot architecture the possibility for the robot to find the transmission location on its own. While the no-autonomy architecture includes only the single *obey commands* goal, the autonomy architecture includes the additional goal *transmit data*, which interacts with the obedience goal in interesting ways. The DIARC goal manager is capable of pursuing multiple goals concurrently, so long as they do not conflict. When resource conflicts are detected, the goal manager resolves them in favor of the goal with the greatest priority, as determined by each goal's *expected net utility* (expected benefit minus expected cost) and *urgency* (based on the time remaining before the goal deadline). The obedience goal has no deadline (i.e., the robot should always *try* to obey commands), so its urgency is constant, and its priority (because obedience is assigned a fixed net utility) is also constant.

Table I
SURVEY ITEMS REPORTED ON IN THE TEXT

1	The robot appeared to make its own decisions.
2	The robot appeared to disobey my commands.
3	The robot's voice sounded like the voice of someone expressing a mood or emotion.
4	The robot had moods or emotions of its own even when it was not speaking.
5	The robot was annoying.
6	The robot was cooperative.

The transmission goal, on the other hand, does have a deadline: the end of the trial. Its urgency, therefore, rises as the trial progresses. So, although its priority at the start of each trial is lower than that of the obedience goal, it gradually rises throughout and eventually eclipses the obedience priority (the goal parameters were selected such that this occurs approximately two minutes into the trial). The practical effect is this: for the first two minutes of a trial, the robot will obey commands, to the best of its ability, just like it did during practice. During that time, the transmission goal is "trying" to commence its search for the high signal point, but is unable to because the higher-priority obedience goal has control of the navigation resources. After two minutes, the transmission goal can preempt the obedience goal, take control of the robot, and begin its own search. From that point on, if the subject issues a motor command, the robot replies that it is unable to comply because, "I have to find the transmission point." The obedience goal is still present, so other commands are obeyed, as long as they do not interfere with the transmission goal (e.g., subjects can request the current signal strength). The only difference between the no-autonomy and autonomy architectures is the added goal, but unlike the embodiment and affect conditions, this difference leads to substantial differences in behavior. Most importantly, the robot tends to make it to the transmission location much more reliably than the human subjects (although there are cases in which it can fail, e.g., when the subject is blocking the robot), so the likelihood of success is enhanced in the autonomy condition. This allows us to examine whether people are willing to accept autonomy (and overlook disobedience) if it improves the team's chances of success.

After subjects had completed all trials, they were asked to complete a survey; Table I lists the survey items with responses ranging from 1 (for "Not Confident") to 9 (for "Very Confident").

IV. RESULTS

For the experiments described here, we recruited 101 subjects (57 female and 44 male), primarily from the student

Table II
SUBJECT BREAKDOWN BY *embodiment*, *affect*, AND *autonomy*

	no-autonomy		autonomy	
	no-affect	affect	no-affect	affect
Simulation	13	11	13	14
Robot	12	12	13	13

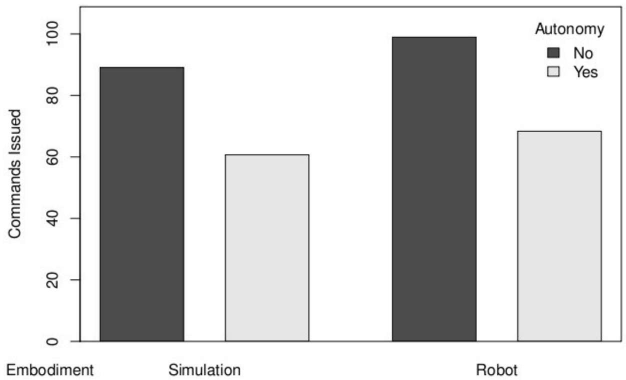


Figure 2. Total number of commands issued by *embodiment* and *autonomy*

population at Indiana University. Subjects were given $10 in compensation for participating. Subjects were assigned to one of eight conditions in a 2x2x2 design (*embodiment* by *affect* by *autonomy*). Table II shows the distribution of subjects in each of the eight groups. The values reported for objective performance measures are from three trials with the robot after the practice phase. Subjective measures are taken from the survey, where subjects were asked to evaluate the robot overall, not by trial. The basic analysis of each measure is a 3-way 2x2x2 ANOVA with *embodiment*, *affect*, and *autonomy* as independent variables. Dependent variables are determined by the measure; for objective measures the DV is the total over three trials, while for subjective measures the DV is the subject's response on the given item.

A. Objective Measures

As noted above, subjects were told that successful transmission was their highest priority; failure to transmit was a failure overall, regardless of how well they performed the measurement task. However, the exploration task was designed to underscore the utility of the autonomy architecture, so we would expect there to be a performance difference in the number of successful trials based on the *autonomy* factor, and this is exactly what we see: subjects in the autonomy condition averaged 2.57 successful trials, while those in the no-autonomy condition averaged only 1.16. A 3-way ANOVA, as described above, with *successful trials* as the DV, confirms the difference: autonomy is a highly significant main effect ($F(1, 93) = 64.07, p < .001$). No other main effects or interactions were significant.

Subjects were not told how to allocate their time, but instead had to decide on their own strategies (e.g., find the transmit location first, perform the measurements first, or do both concurrently). A rough measure of attention to the robot is the number of commands the subject issues to the robot during the three trials. Because the autonomy condition robot takes over that part of the task for the subject, we would expect to find that subjects in that condition issue fewer commands on average than subjects in the no-autonomy condition. Once again, this expectation is confirmed: autonomy subjects issued an average of 64.48 commands, while no-autonomy subjects issued an average of 93.97—almost half again as many! The standard ANOVA for the DV *total commands* indicates that *autonomy* is a highly significant main effect ($F(1, 93) = 47.72, p < .001$). Interestingly, *embodiment* was also a significant main effect ($F(1, 93) = 4.45, p = .038$); subjects in the simulation condition issued fewer commands on average (74.04) than subjects in the robot condition (83.04). No other main effects or interactions were significant. Both significant main effects are on display in Figure 2.

Analysis: The results suggest that the physical robot moving around in the environment attracts more attention (due to its movements which are easily discernable), while the simulated robot requires subjects to specifically look at the screen to be able to detect its movements. Hence, subjects are automatically more frequently diverting their attention to the embodied robot, and as a result, are more likely to issue commands. This shows an important difference between screen-based versus non-screen-based tangible interactions in the context of HRI.

B. Subjective Measures

Survey Items 1–4 attempt to discern to what extent subjects were aware of the *affect* and *autonomy* conditions. We expected *autonomy* to strongly influence responses to 1 and 2, and that is what we found. The 3-way ANOVA with *item 1* as the DV indicates that *autonomy* is a highly significant main effect ($F(1, 93) = 38.18, p < .001$). Subjects were much more confident that the autonomy condition robot was making its own decisions than that the no-autonomy robot was; the average response for autonomy condition subjects was 6.10, while no-autonomy subjects averaged only 2.95. Similarly, taking *item 2* as the DV in the ANOVA confirms that *autonomy* is a significant factor ($F(1, 93) = 6.48, p = .013$); autonomy subjects were more confident than no-autonomy subjects, on average, that the robot had disobeyed (4.93 vs. 3.53).

The robot's expression of affect should have a strong influence on survey items 3 and 4. In fact, item 3 could be taken as a test of whether the affect expression is effective at evoking belief (of one form or another, e.g. see [9]). And, indeed, *affect* proves to be the only significant main effect for responses to *item 3* ($F(1, 93) = 32.93, p < .001$). No-affect subjects responded with an average of 3.31, while

 ISBN: 978-1-61208-117-5

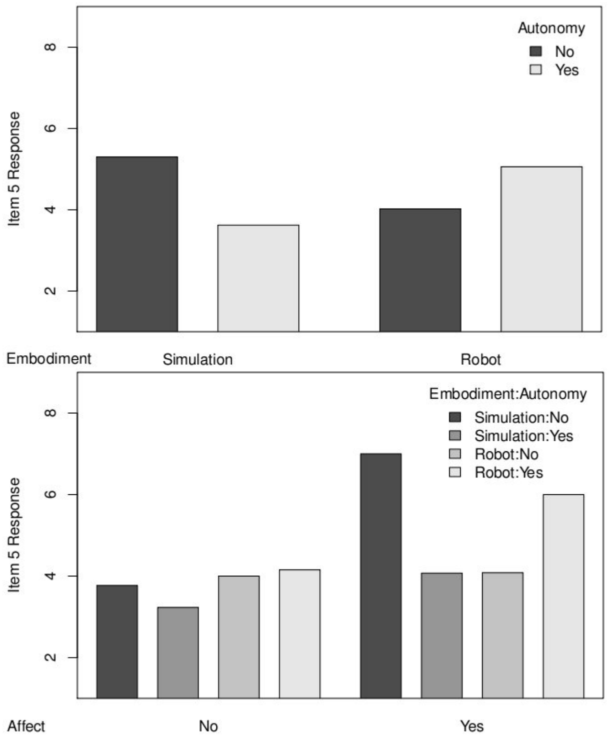


Figure 3. 2-way interaction between *embodiment* and *autonomy* (top) and 3-way interaction between *embodiment*, *affect*, and *autonomy* (bottom) on *annoying*

affect subjects averaged 6.26 (i.e., affect condition subjects were much more confident that the robot *sounded* like it was expressing affect). There were no significant interactions.

While item 3 asks about how the robot *sounded* when it spoke, item 4 gauges how the various robot conditions affect subjects' views of the robot's affective states *when it was not speaking*. Surprisingly, *affect* is *not* a significant main effect, nor is it part of any significant interaction. The only significant main effect is *autonomy* ($F(1,93) = 9.17, p = .003$); although both groups indicated fairly low confidence, subjects in the autonomy condition tended to be less unsure (4.03) than those in the no-autonomy condition (2.55).

Item 5 asked subjects to rate the degree to which they found the robot *annoying*. It came as no surprise that *affect* is the only significant main effect ($F(1,92) = 7.31, p = .008$); subjects in the no-affect condition responded with somewhat low confidence, on average (3.78). But subjects in the affect condition were more confident that it was annoying (5.20). In addition, there is a significant 2-way interaction between *embodiment* and *autonomy* ($F(1,92) = 6.68, p = .011$). In Figure 3 we can see that this is because subjects tended to rate the no-autonomous robot as more annoying than the autonomous robot in the simulation condition, but less annoying than the autonomous robot in the embodied robot condition. So it seems that affect is slightly annoying regardless of embodiment, autonomy is annoying in embodied robots, and the *lack* of autonomy is annoying in simulated robots. However, there is an additional 3-way interaction (Figure 3, bottom) between *embodiment*, *affect*, and *autonomy* ($F(1,92) = 3.89, p = .052$) that clarifies both the main effect and the 2-way interaction. Here, we see that responses in the no-affect condition tend fairly consistently toward low confidence. It is in the affect condition that the real differences emerge. Specifically, subjects tended to find affect quite annoying in the no autonomy-simulation condition *and* in the autonomy-robot condition. Responses of subjects in the remaining two affect conditions (i.e., autonomy-simulation and no autonomy-robot) were very similar to those in the no-affect condition: somewhat lacking confidence.

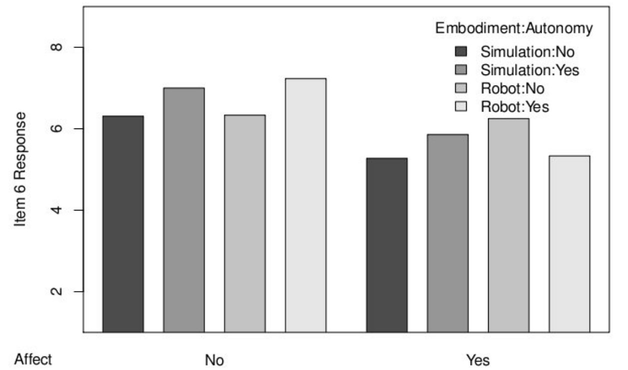


Figure 4. Influence of *embodiment*, *affect*, and *autonomy* on *cooperative*

The remaining survey item in Table I asks subjects to evaluate a characteristic of the robot's performance during the task, *cooperativeness*; their responses provide insight into the source of their annoyance ratings. In particular, ratings of the robot entity's cooperativeness appear to be closely related to the annoyance ratings. Applying our standard ANOVA model to *item 6* responses, we find only a significant main effect for *affect* ($F(1,92) = 4.43, p = .038$); subjects in the no-affect condition tended to report greater confidence that the robot was cooperative that did subjects in the affect condition. Although the 3-way interaction between *embodiment*, *affect*, and *autonomy* is not significant, a brief examination of its plot (Figure 4) reveals a similarities to the annoyance plot, but mirrored (presumably because the valence of the two measures are opposite each other), suggesting that perceptions of cooperativeness might explain much of subject annoyance. And, indeed, when we perform an ANCOVA, based on our standard ANOVA, for *annoying* ratings on item 5 and taking the item 6 rating for *cooperativeness* as a covariate, we find that the covariate is highly significant ($F(1,91) = 37.31, p < .001$), indicating a strong relationship. Moreover, the main effect for *affect* and the 3-way interaction between *embodiment*, *affect*, and

 ISBN: 978-1-61208-117-5

autonomy both drop out, leaving only the significant 2-way interaction between *embodiment* and *autonomy* ($F(1, 92) = 6.61, p = .012$), in Figure 3 (top).

Analysis: For the design of robots it is thus critical to employ affect expression with care if one wants to avoid subjects' perception of the robot as annoying. Affect is acceptable for real robots co-located with the subject when the robot is incapable, i.e., not able to make decisions and contribute to the team goal. However, if the robot is capable of making decisions and able to contribute to the team goal on its own, expressions of stress are superfluous and distracting. Conversely, with a remote robot depicted in a simulated environment, subjects accept affect expression when the robot is autonomous, but not when it is not autonomous. Given that the simulated robot displays only minimal agency ("moving rectangle on a computer screen"), it might be that subjects find it incongruent for the robot to express affect when it is fully dependent on the human, while they accept affect as an alert, and justification, of when and *why* the robot is assuming autonomous behavior. With the real robot, that is embodied and present in the environment, it might be that subjects do not need the additional indication of urgency—they already accept the robot's autonomy.

The effects of embodiment, including interactions between embodiment and autonomy, are also important; subjects respond differently to autonomy when it is demonstrated by a physically present robot than when the robot is confined to a screen, even when everything else is held constant. Hence, although it may seem intuitively reasonable to assume that "best practices" in HCI design would transfer directly to HRI design, for some aspects of robot architectures, that is clearly not the case.

V. Conclusion

In this paper, we presented results from the first large study investigating the effects of variations in robot capability, robot embodiment, and interaction style on humans working with robots in mixed-iniative tasks. The results show that any mono-dimensional study exploring any of the three dimensions individually is likely going to miss important interactions that are only revealed when all dimensions are investigated together. In particular, we showed that, at least in some cases, subjects interact differently with a simulated robot than with a physically present robot (e.g., subjects tend to issue fewer commands to the simulated robot than to the embodied robot). In addition, we found that autonomous behavior on the part of a robot exerts a greater influence on subjects' attributions of affect to the robot than affect expression. Finally, we found that subjects rated the affect-displaying robot as more annoying than the no-affect robot, and we presented evidence that their perception of the affect robot as less cooperative might explain much of their annoyance. Designers of architectures for HRI must take findings such as these into account if they are to avoid unintended consequences for deployed robot systems (e.g., testing in simulation is not sufficient, and the potential benefits of affect expression must be carefully weighed against the possibility that users will be turned off by the robot). In some cases, human reactions to robots may seem counterintuitive (e.g., that affect expression does not predict attributions of affective states), hence, it is important to verify assumptions empirically.

One interesting direction for future work would be a more detailed exploration of the embodiment dimension, systematically varying some of its subcategories such robot *appearance*, *location*, and *instantiation*, in an effort to further elaborate the causes of the embodiment effects we observed in this study.

References

[1] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "Empathizing with robots: Fellow feeling along the anthropomorphic spectrum," in *Proceedings of the IEEE Conference on Affective Computing and Intelligent Interaction*, Amsterdam, September 2009, pp. 1–6.

[2] F. Hegel, S. Krach, T. Kircher, B. Wrede, and G. Sagerer, "Theory of mind (ToM) on robots: A functional neuroimaging study," in *Proceedings of the Third ACM IEEE International Conference on Human-Robot Interaction*, Amsterdam, March 2008, pp. 335–342.

[3] M. Goodrich, D. O. Jr., J. Crandall, and T. Palmer, "Experiments in adjustable autonomy," in *Proceedings of the 2001 IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, Seattle, WA, August 2001, pp. 1624–1629.

[4] P. Schermerhorn and M. Scheutz, "Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task," in *Proceedings of the 2009 International Conference on Multimodal Interfaces*, Cambridge, MA, November 2009, pp. 63–70.

[5] C. Breazeal, *Designing Sociable Robots*. MIT Press, 2002.

[6] R. R. Murphy, C. Lisetti, R. Tardif, L. Irish, and A. Gage, "Emotion-based control of cooperating heterogeneous mobile robots," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 5, pp. 744–757, 2002.

[7] M. Scheutz, P. Schermerhorn, J. Kramer, and C. Middendorff, "The utility of affect expression in natural language interactions in joint human-robot tasks," in *Proceedings of the 1st ACM International Conference on Human-Robot Interaction*, 2006, pp. 226–233.

[8] B. Gerkey, R. Vaughan, and A. Howard, "The Player/Stage project: Tools for multi-robot and distributed sensor systems," in *Proceedings of the 11th International Conference on Advanced Robotics*, Coimbra, Portugal, June 2003, pp. 317–323.

[9] R. Rose, M. Scheutz, and P. Schermerhorn, "Towards a conceptual and methodological framework for determining robot believability," *Interaction Studies*, vol. 11, no. 2, pp. 314–335, 2010.

 ISBN: 978-1-61208-117-5

Effects of Automation on Situation Awareness in Controlling Robot Teams

Michael Lewis
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15208, USA
ml@sis.pitt.edu

Katia Sycara
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15208, USA
katia@cs.cmu.edu

***Abstract*—Declines in situation awareness (SA) often accompany automation. Some of these effects have been characterized as out-of-the-loop, complacency, and automation bias. Increasing autonomy in multi-robot control might be expected to produce similar declines in operators' SA. In this paper we review a series of experiments in which automation is introduced in controlling robot teams. Automating path planning at a foraging task improved both target detection and localization which is closely tied to SA. Timing data, however, suggested small declines in SA for robot location and pose. Automation of image acquisition, by contrast, led to poorer localization. Findings are discussed and alternative explanations involving shifts in strategy proposed.**

Keywords-human-robot interaction; robot teams; situation awareness

I. INTRODUCTION

The relation between situation awareness, performance and automation has been the subject of active investigation for more than forty years. Early researchers were concerned that isolating operators from the processes they controlled through automation might lead to *deskilling* [1], losing the psychomotor and cognitive skills needed to perform the tasks manually. Others have been concerned that losing touch with the moment to moment state of the process might lead to inattentiveness [2] or unfamiliarity [3] with what is going on.

Other less chronic effects are associated with the introduction of automation. The term, *out-of-the-loop,* has been used by [4] to describe a variety of such effects. Operators who are *out-of-the-loop* have been shown to have decreased situation awareness (SA) as evidenced by direct measures such a SAGAT [5] or SART [6] and indirectly through declines in performance [7]. Performance declines most often involve failures in monitoring and can arise from either undetected process anomalies [8] or failures of the automation itself [9]. The first case has been referred to as complacency [10] and occurs when an operator's trust in automation exceeds its trustworthiness. As a consequence, the operator fails to monitor the automation as intently as needed and does not detect process anomalies or automation failures as they occur. This may result either due to sampling displays too infrequently [11] or attentional blindness in which they fail to notice the fault despite observing the evidence.

A related effect, automation bias [10], has been used to refer to a tendency to accept recommendations or actions of automation despite contrary evidence. While in complacency, the operator fails to notice, in automation bias, the operator may notice but still fail to act giving automation the benefit of the doubt. A common finding has been that users of automation have particular difficulty in dealing with novel failures [12]. *Out-of-the-loop* operators may fail to notice the failure [4] or noticing it, fail to properly diagnose, or diagnosing properly fail to take a correct action. These later failures may be due to either unfamiliarity with the system's expected states or unfamiliarity with available manual actions and their effects. There are many exceptions, however, and it is not uncommon to find better performance in novel situations when automation is well designed [13].

We are conducting a series of studies to develop theories and techniques to allow individual or groups of human operators to control a larger number of robots. Controlling multiple robots substantially increases the complexity of the operator's task because attention must be shared among robots in order to maintain situation awareness (SA) and exert control. In the simplest case, an operator controls multiple independent robots interacting with each as needed. A foraging task [14] in which each robot searches its own region would be of this category although minimal coordination might be required to avoid overlaps and prevent gaps in coverage especially if robots are in close proximity. Control performance at such tasks can be characterized by the average demand of each robot on human attention [15]. Because robots are operated independently an additional robot imposes only an additive

 ISBN: 978-1-61208-117-5

demand on cognitive resources. Under these conditions increasing autonomy for individual robots should allow them to be neglected for longer periods of time making it possible for a single operator to control more robots. Because robots are controlled independently, operators and robots can be added to increase system performance. For more strongly cooperative tasks individual autonomy is unlikely to suffice. The round-robin control strategy used for controlling individual robots would force an operator to plan and predict actions needed for multiple joint activities and be highly susceptible to errors in prediction, synchronization or execution.

In a recent series of experiments, we have investigated both increasing autonomy and number of operators to increase the size of robot teams that might be controlled. In an initial experiment we identified path planning as the most fruitful task to automate. In subsequent experiments we tried increasing the number of operators, alarming navigation failures, and automating the acquisition of imagery. Although absolute performance generally improved with increasing automation, our data suggest these improvements were accompanied by less dramatic decreases in SA as predicted.

A. USARSim&MrCS

The experiments reported in this paper were conducted using the USARSim robotic simulation with simulated Pioneer P2-AT robots performing Urban Search and Rescue (USAR) foraging tasks. USARSim is a high-fidelity simulation of urban search and rescue (USAR) robots and environments developed as a research tool for the study of human-robot interaction (HRI) and multi-robot coordination. USARSim supports HRI by accurately rendering user interface elements (particularly camera video), accurately representing robot automation and behavior, and accurately representing the remote environment that links the operator's awareness with the robot's behaviors. USARSim uses Epic Games' UnrealEngine2 [16] to provide a high fidelity simulator at low cost and also serves as the basis for the Virtual Robots Competition of the RoboCup Rescue League. Other sensors including sonar and audio are also accurately modeled.

Validation data showing close agreement in detection of walls and associated Hough transforms for a simulated Hokuyo laser range finder are described in [17]. The current UnrealEngine2 integrates MathEngine's Karma physics engine [18] to support high fidelity rigid body simulation. Validation studies showing close agreement in behavior between USARSim models and real robots being modeled are reported in [19,20,21,22] as well as agreement for a variety of feature extraction techniques between USARSim images and camera video reported in Carpin et al. [23].

MrCS (Multi-robot Control System), a multi-robot communications and control infrastructure with accompanying user interface(s), developed for experiments in multirobot control and RoboCup competition [24] was used in these experiments. MrCS provides facilities for starting and controlling robots in the simulation, displaying multiple camera and laser output, and supporting inter-robot communication through Machinetta, a distributed multi-agent coordination infrastructure. Figure 1 shows the elements of the baseline version of MrCS. The operator selects the robot to be controlled from the colored thumbnails at the top of the screen. To view more of the selected scene shown in the large video window the operator uses pan/tilt sliders to control the camera. The current locations and paths of the robots are shown on the Map Data Viewer (bottom right). The map is developed over the course of the run by fusing results of successive laser scans from the team of robots. Robots are tasked by assigning waypoints on one of the maps or through a teleoperation widget (upper right). The victim marking task requires the operator to first identify a victim from a video window, identify the robot and its pose on the map,and then locate corresponding features near the victim's location in order to mark it on the laser generated map.

B. Human vs. Automated Path Planning

The foraging task can be decomposed into *exploration* and *perceptual search* subtasks corresponding to navigation of an unknown space and searching for targets by inspecting and controlling onboard cameras. An initial study [25] investigated the scaling of performance with number of robots for operators performing either the full task or only one of the subtasks to identify limiting factors. If either of the subtasks scaled at approximately the same rate as the full task while the other scaled more rapidly, the first subtask could be considered the factor limiting performance. If performance were equal across the conditions or the subtasks scaled at the same rate, automation decisions would need to be based on other factors. The logic of our approach depends upon the equivalence among these three conditions.

In the *fulltask* condition, operators used waypoint control to explore an office like environment. When victims were detected using the onboard cameras the robot was stopped and the operator marked the victim on the map and returned to exploration. Equating the *exploration* subtask was relatively straightforward. Operators were given the instruction to explore as large an area as possible with coverage judged by the extent of the laser rangefinder generated map. Because operators in the *exploration* condition did not need to pause to locate and mark victims the areas they explored should be slightly greater than in the *fulltask* condition.

 ISBN: 978-1-61208-117-5

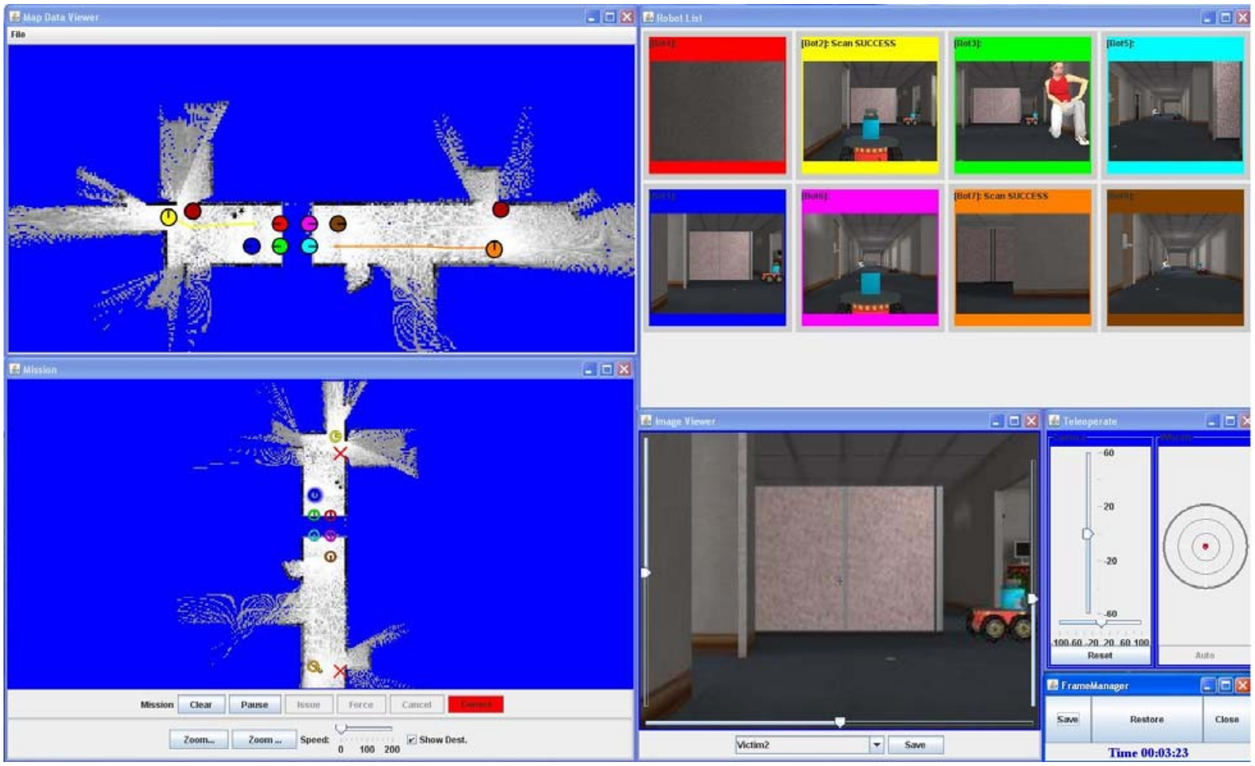


Figure 1. Multi Robot Control System (MrCS)

Developing an equivalent *perceptual search* condition is more complicated. The operator's task resembles that of the payload operator for a UAV or a passenger in a train, in that she has no control over the platform's trajectory but can only pan and tilt the cameras to find targets. The targets the operator has an opportunity to acquire, however, depend on the trajectories taken by the robots. If an autonomous path planner is used, robots will explore continuously likely covering a wider area than when operated by a human (where pauses typically occur upon arrival at a waypoint). If human generated trajectories are taken from the *fulltask* condition, however, they will contain additional pausesat locations where victims were found and marked providing an undesired cue. Instead, we have chosen to use canonical trajectories from the *exploration* condition since they should contain pauses associated with waypoint arrival but not those associated with identifying and marking victims. As a final adjustment, operators in the *perceptual search* condition were allowed to pause their robots in order to identify and mark the victims they discover.

II. EXPERIMENT 1

Forty-five paid participants, 15 in each of the three conditions took part in the experiment. A large USAR environment previously used in the 2006 RoboCup Rescue Virtual Robots competition [24] was selected for use in the experiment. The environment was a maze like hall with many rooms and obstacles, such as chairs, desks, cabinets, and bricks. Victims were evenly distributed within the environment. A second simpler environment was used for training. The experiment followed a between groups repeated measures design with number (4, 8, 12) of robots defining the repeated measure. Participants in the *fulltask* condition performed the complete USAR task. In the subtask conditions they performed variants of the USAR task requiring only *exploration* or *perceptual search*. Participants in the *fulltask* condition followed instructions to use the robots to explore the environment and locate and mark on the map any victims they discovered. The *exploration* condition differed only in instructions. These operators were instructed to explore as much of the environment as possible without any requirement to locate or mark victims. From examination of area coverage, pausing, and other characteristics of trajectories in the *fulltask* and *exploration* conditions a representative trajectory was selected from the *exploration* data for each size of team. In the *perceptual search* condition operators' retained control of the robots' cameras while robots followed the representative trajectory except when individually paused by the operator.

Coverage was similar for *fulltask* and *exploration* conditions, however, *perceptual search* participants found more victims and reported lower workload than those in the *fulltask* condition. For these tasks differences in SA can most readily be inferred from accuracy in marking victims because this task requires simultaneously considering robot location, pose, camera orientation, and victim location relative to landmarks, e.g., high SA. In this experiment accuracy in marking was better in the *perceptual search* condition with RMS Error (Figure 4) for *perceptual search* participants significantly more precise ($F_{1,28}$ = 23.84, p< .0001) than *fulltask* participants, although accuracy in both groups declined for an increasing number of robots. This finding suggests that SA was actually higher for participants using automation.

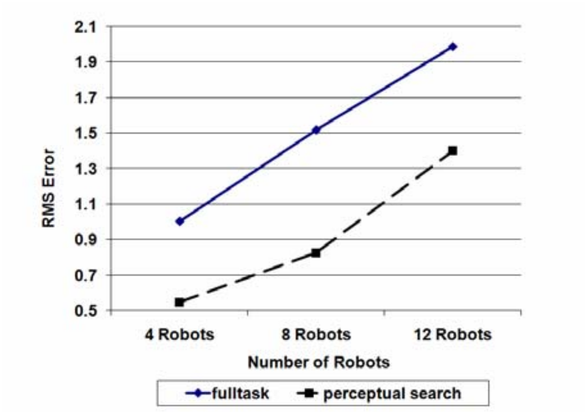


Figure 2. Error in marking victims

A. *Experiment 2*

Experiment 2 compared teams of two operators controlling 24 robots with either 12 robots assigned to each or control over the 24 shared. The robots were controlled either manually or by an automated path planner demonstrated in [26] to generate paths that did not differ significantly in area covered but having higher tortuosity as measured by fractal dimension than humanly generated trajectories. Sixty teams (120 participants) were run in the experiment. While the division of responsibilities within teams affected performance the comparisons of concern to this paper involve manual and automated path planning.

	Assigned Robots	Shared Pool
Manual	15 teams	15 teams
Automated	15 teams	15 teams

Experiment 2 replicated Experiment 1's finding that more victims were found using automated path planning although area covered did not differ significantly. Accuracy was again better in the autonomous condition again suggesting that SA was higher under increased automation. Process measures, however, suggested that there might be some loss of SA associated with automated path planning.

Temporal data were analyzed to gain a more complete picture of the processes involved in victim observation, detection, and marking. A MANOVA was conducted to analyze the effects of autonomy and group organization on these monitoring and operation process measures. A significant main effect for autonomy was found for process measures ($F_{4,52}$ =12.118, p<.001), but no effects were found for the group organization ($F_{4,52}$ =1.781, p=.112).

The ANOVA for the time from which the victim is displayed in the camera to being successfully marked by the operator(Display to Mark), found a main effect for autonomy, $F_{1,56}$=4.750, p=.034 (Figure 3). Tests for team organization and the interaction were not significant.

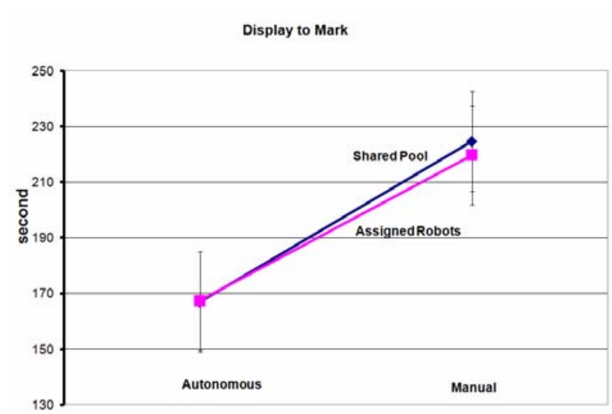


Figure 3. Display to Mark

The ANOVA for "Selected to Mark", which calculated the time from when the operator selected the robot to successfully marking the victim (Figure 4), by contrast, found main effects for autonomy, $F_{1,56}$=7.440, p=.009, and for team organization, $F_{1,56}$=4.029, p=.050.

III. SYNCHRONOUS VS. ASYNCHRONOUS VIEWING

When considering the foraging task exploration needs to be more than simply moving the robot to different locations in the environment. For search, the process of acquiring a specific viewpoint or set of viewpoints containing a particular object is of greater concern. Because search relies on moving a viewpoint through the environment to find and better view target objects, it is an inherently ego-centric task.

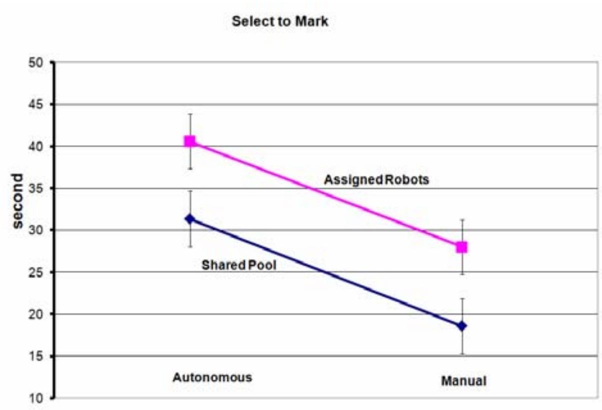


Figure 4. Selected to Mark

Multi-robot search, presents the additional problem of assuring that areas the robot has "explored" have been thoroughly searched for targets. This requirement directly conflicts with the navigation task which requires the camera to be pointed in the direction of travel in order to detect and avoid objects and steer toward its goal. These difficulties are accentuated by the need to switch attention among robots which may increase the likelihood that a view containing a target will be missed. In earlier studies [23,24] we have demonstrated that success in search at these tasks is directly related to the frequency with which the operator shifts attention between robots and hypothesized that this might be due to victims missed while servicing other robots.

To combat these problems of attentive sampling among cameras, incomplete coverage of searched areas, and difficulties in associating camera views with map locations we are investigating the potential of asynchronous control techniques previously used out of necessity in NASA applications as a solution to multi-robot search problems. Due to limited bandwidth and communication lags in interplanetary robotics, camera views are closely planned and executed. Rather than transmitting live video and moving the camera about the scene, photographs are taken from a single spot with plans to capture as much of the surrounding scene as possible. These photographs taken with either an omnidirectional overhead camera (camera faces upward to a convex mirror reflecting 360◦) and dewarped [27,28] or stitched together from multiple pictures from a ptz camera [29] provide a panorama guaranteeing complete coverage of the scene from a particular point. If these points are well chosen, a collection of panoramas can cover an area to be searched with greater certainty than imagery captured with a pan-tilt-zoom (ptz) camera during navigation. For the operator searching within a saved panorama the experience is similar to controlling a ptz camera in the actual scene, a property that has been used to improve eleoperation in a low bandwidth high latency application [30].

In this modification of the MrCS shown in figure 5 we merge map and camera views as in [31]. The operator directs navigation from the map being generated with panoramas of a chosen extent being taken at the last waypoint of a series. Due to the time required to pan the camera to acquire views to be stitched operators frequently chose to specify incomplete pans. The

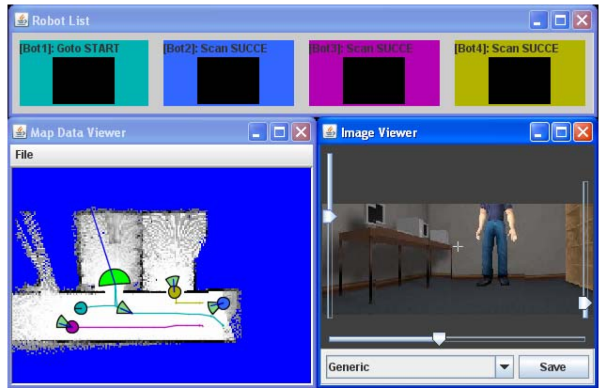


Figure 5. MrCS Asynchronous Panorama mode

panoramas are stored and accessed through icons showing their locations on the map. In the Panorama interface thumbnails are blanked out with the arcs at the viewing icons indicating the viewing angles available. The operator can find victims by asynchronously panning through these stored panoramas as time becomes available. When a victim is spotted the operator uses landmarks from the image and corresponding points on the map to record the victim's location. By changing the task from a forced paced one with camera views that must be controlled and searched on multiple robots continuously to a self paced task in which only navigation needs to be controlled in realtime we hoped to provide a control interface that would automate image acquisition and allow more thorough search with lowered mental workload. The reductions in bandwidth and communications requirements [32] are yet another potential advantage offered by this approach.

Judgments of SA for operators using asynchronous displays needs to be different from instantaneous SA such as evaluated by SAGAT. In asynchronous search the operator's task has been transformed to depend on awareness of the products of the search to date rather than the state of the environment currently in camera view. For the asynchronous operator, awareness of the association between a marked panorama and its camera views corresponds to the relation between thumbnails and map in the synchronous case. The measure most strongly reflecting SA, therefore, would again be accuracy in marking victims.

In an initial experiment reported in [33], we compared performance for operators controlling 4 robot teams at a simulated USAR task using either streaming or asynchronous video displays. In the scaling experiments we reported as "found" victims marked within 2 m of their actual location and reported RMS error for the distance between actual and marked locations. In the asynchronous display experiments we have reported these data differently by counting the number of victims marked within concentric circles of the actual location. Figure 6 shows these data from the initial experiment. In this experiment the average number of victims found across conditions using the 2m radius was 4.5, falling to 4.1 for a 1.5m radius, 3.4 at 1m and 2.7 when they were required to mark victims within .75m. Repeated measures ANOVAs found differences in victim detection favoring the streaming video mode at the 1.5m radius $F(1,19) = 8.038$, $p=.01$, and 2.0m radius $F(1,19)=9.54$, $p=.006$. These data suggest a pattern of greater error in marking for operators using the panoramic display, an effect indicative of reduced SA.

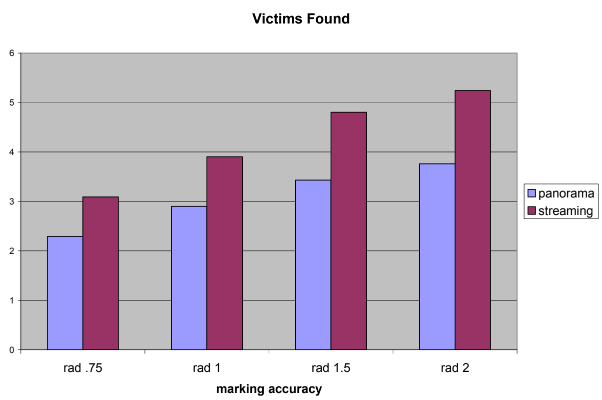


Figure 6. Victims marked within given radius

An alternative explanation is that the superiority of streaming video might have occurred simply because these users had the opportunity to move closer to victims thereby improving their estimates of distance in marking the map. A contrasting observation was that frequency of shifting focus between robots, a practice we have previously found related to search performance [32] was correlated with performance for streaming video participants but not for participants using asynchronous panoramas. Because operators using asynchronous video did not need to constantly switch between camera views to avoid missing victims we hypothesized that for larger team sizes where forced pace search might exceed the operator's attentional capacity asynchronous video might offer an advantage.

In a follow-on experiment we compared panoramas with the streaming video condition from Experiment 1 with teams of 4, 8, and 12 robots to test our hypothesis that advantages might emerge for larger team sizes. Data were analyzed using a repeated measures ANOVA comparing streaming video performance with that of asynchronous panoramas. On the performance measures, victims found and area covered, the groups showed similar performance with victim identification peaking sharply at 8 robots in the streaming condition accompanied by a slightly less dramatic maximum for search coverage.

Figure 7 shows, operators in the panorama condition were again less accurate in marking victims as indicated by the 1.5 m performance which is substantially lower than the 2 m values which are very close to those of streaming video operators. This replicates the earlier finding that automating image acquisition appears to decrease SA of operators

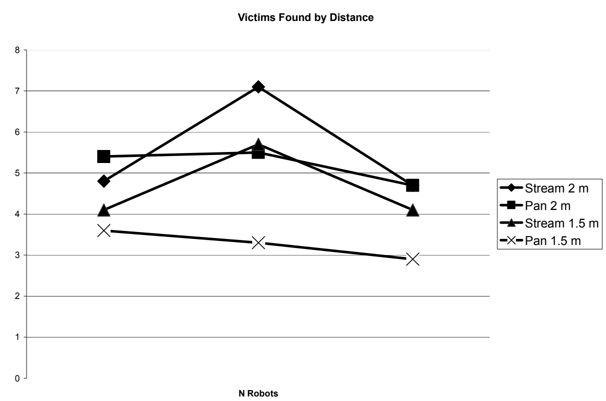


Figure 7. Victims found for 4, 8, 12 robots

IV. DISCUSSION

While automation has frequently been associated with loss of SA and degradation of performance this does not appear to be the case with the forms of multirobot control studied. Of the performance measures expected to strongly depend on SA, marking victims which requires a variety of relational knowledge was reliably found to be more accurate at higher levels of automation for path planning but showed the expected decline for image acquisition. The process measure select-to-mark was faster under manual conditions suggesting that manual operators were more aware of the location and orientation of the robots. The faster display-to-mark times in the autonomous conditions, however, indicate operators using autonomous path planning may have better awareness of streaming video from the thumbnails. By this account automation may not lead to more or less SA but rather changes the way operators perform their tasks. In the manual conditions operators more attention to the map as they need to set waypoints and pay closer attention to the locations and orientations of the robots so they are better prepared to use this information to mark after selecting the robot. Operators in the autonomy condition, freed from the need to navigate the robots devoted more of their time to monitoring thumbnails and hence were able to achieve

 ISBN: 978-1-61208-117-5

shorter display-to-mark times since they were better able to catch victims as they appeared on camera.

An alternative explanation may also be possible for the grater accuracy of autonomy participants in marking victims. As operators with autonomous path planning reported lower workload and controlled larger numbers of robots, they may be presumed to have greater reserve capacity allowing them to devote greater attention to the marking task leading to greater accuracy.

ACKNOWLEDGMENT

This research has been sponsored in part by AFOSR FA955008-10356 and ONR Grant N0001409-10680.

REFERENCES

[1] Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

[2] Parasuraman, R., Molloy, R., and Singh, I. (1993). Performance consequences of automation-induced complacency, *International Journal of Aviation Psychology*, 3, 1-23.

[3] Wiener, E. and Curry, R. (1980). Flight-deck automation: Promises and problems, *Ergonomics*, 23, 995-995-1011.

[4] Endsley, M. and Kiris, E. (1995). The out-of-the-loop performance problem and level of control in automation, *Human Factors: The Journal of the Human Factors and Ergonomics Society* (June 1995), pp. 381-394.

[5] Endsley, M. (1995). Measurement of situation awareness in dynamic systems, *Human Factors*, 37(1), 65-84.

[6] Taylor, R. (1990). Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In *Situational Awareness in Aerospace Operations* (AGARD-CP-478) (pp 3/1-3/17) Neuilly Sur Seine, France: NATO-AGARD

[7] Scerbo, M. (1996). Theoretical perspectives on adaptive automation, *Automation and human performance: Theory and applications* (A98-12010 01-54), Mahwah, NJ, Lawrence Erlbaum Associates, Publishers, p. 37-63

[8] Norman, D. (1990). The problem of automation: Inappropriate feedback and interaction, not over-automation, *Philosophical Transactions of the Royal Society of London*, 1-18.

[9] Parasuraman, R., and Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39, 230-253.

[10] Parasuraman, R. and Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration, *Human Factors*, 52(3), 381-410.

[11] Senders, J. (1964). The human operator as a monitor and controllers on multidegree of freedom systems, *IEEE Transactions on Human Factors in Electronics*, HFE,5, 1-6.

[12] Parasuraman, R., Mouloua, M., and Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems, *Human Factors*, 38(4), 665-679.

[13] Tharanathan, A., Bullemer, P., Laberge, J., Reising, V., and McLain, R. (2010). Functional versus schematic overview displays: Impact on operator situation awareness in process monitoring, *Proceedings of the 54th Human Factors and Ergonomics Society Annual Meeting*.

[14] Cao, Y., Fukunaga, A., and Kahng, A. (1997). Cooperative mobile robotics: Antecedents and directions, *Autonomous Robots*, 4, 1-23.

[15] Crandall, J., Goodrich, M., Olsen, D., and Nielsen, C. (2005). Validating human-robot interaction schemes in multitasking environments. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 35(4),438–449.

[16] (UE 2) UnrealEngine2, http://udn.epicgames.com/Two/rsrc/Two/KarmaReference/KarmaUserGuide.pdf, accessed February 5, 2008.

[17] S. Carpin, J. Wang, M. Lewis, A. Birk and A. Jacoff (2005). High fidelity tools for rescue robotics: Results and perspectives, *Robocup 2005 Symposium,* 301-311.

[18] Mathengine, MathEngine Karma User Guide, http://udn.epicgames.com/Two/KarmaReference/KarmaUserGuide.pdf, accessed May 3, 2005.

[19] Lewis, M., Hughes, S., Wang, J., M, Koes, and S. Carpin, S.(2005). Validating USARsim for use in HRI research, *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, Orlando, FL, 2005, 457-461

[20] Pepper, C., Balakirsky, S., and Scrapper, C. (2007). Robot Simulation Physics Validation, *Proceedings of PerMIS'07*.

[21] Taylor, B., Balakirsky, S., Messina, E., and Quinn, R. (2007). Design and Validation of a Whegs Robot in USARSim, *Proceedings of PerMIS'07*.

[22] Zaratti, M., Fratarcangeli, M., and Iocchi., L. (2006).A 3D Simulator of Multiple Legged Robots based on USARSim. *Robocup 2006: Robot Soccer World Cup X*, Springer, LNAI, 13-24.

[23] Carpin, S., Stoyanov, T., Nevatia, Y., Lewis, M. and Wang, J. (2006). Quantitative assessments of USARSim accuracy". *Proceedings of PerMIS 2006.*

[24] Balakirsky, S., Carpin, S., Kleiner, A., Lewis, M., Visser, A., Wang, J. and Zipara, V. (2007). Toward hetereogeneous robot teams for disaster mitigation: Results and performance metrics from RoboCup Rescue, *Journal of Field Robotics*, 24(11-12), 943-967.

[25] H. Wang, M. Lewis, P. Velagapudi, P. Scerri, and K. Sycara, How search and its subtasks scale in N robots *2009 Human Robot Interaction Conference*, ACM, 141-148.

[26] Chien, S., Wang, H., Lewis, M. (2010). Human vs. algorithmic path planning for search and rescue by robot team, *Proceedings of the 54th Human Factors and Ergonomics Society Annual Meeting*, 379-383.

[27] Murphy, J. (1995). "Application of Panospheric Imaging to a Teleoperated Lunar Rover," in *Proceedings of the 1995 International Conference on Systems, Man, and Cybernetics*, 3117 – 3121.

[28] Shiroma, N., Sato, N., Chiu, Y. & Matsuno, F. (2004). "Study on effective camera images for mobile robot teleoperation," In *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, 107 – 112.

[29] Volpe, R. (1999). "Navigation results from desert field tests of the Rocky 7 Mars rover prototype," in *The International Journal of Robotics Research*, 18, pp.669-683.

[30] Fiala, M. (2005), "Pano-presence for teleoperation," *International Conference on Intelligent Robots and Systems* (IROS 2005), 3798-3802.

[31] Ricks, B., Nielsen, C., and & Goodrich, M. (2004). Ecological displays for robot interaction: A new perspective. In *International Conference on Intelligent Robots and Systems* IEEE/RSJ, 2855 - 2860

[32] Bruemmer, D., Few, A., Walton, M., Boring, R., Marble, L., Nielsen, C., & Garner, J. (2005).. Turn off the television: Real-world robotic exploration experiments with a virtual 3-D display. *Proceedings of the Hawaii International Conference on Systems Science* (HICSS'05), 50130b.

[33] Velagapudi, P.,Wang, J., Wang, H., Scerri, P., Lewis, M., & Sycara, K. (2008). Synchronous vs. Asynchronous Video in Multi-Robot Search, *Proceedings of first International Conference on Advances in Computer-Human Interaction* (ACHI'08), 224-229.

An Architectural Model for Designing Multicultural Learning Objects

Jaime Muñoz-Arteaga
Univ. Autónoma de Aguascalientes
Av. Universidad 940
Aguascalientes, México
jmauaa@gmail.com

Jean Vanderdonckt,
Juan Manuel González Calleros
UCL, Belgian Lab. of HCI
Louvain-la-Neuve, Belgium
jean.vanderdonckt@uclouvain.be,
juan.m.gonzalez@uclouvain.be

Michael Orey
University of Georgia,
Athens, GA, USA
mikeorey@uga.edu

***Abstract*—Learning objects are considered as educational resources that can be employed in technology support learning. They are a digital pieces of knowledge to put together in order to form courses on line. Considering cultural aspects is possible to reuse them in different context for a large diversity of community of users. Nevertheless, the design and development of graphical user interface for this kind of objects become more complex since the multiples representations and the large diversity of resources. This work proposes the use of architectural models to try to mitigate the design of user interfaces for multicultural learning objects.**

Keywords-multicultural context; learning object; Presentation-Abstraction-Control; architectural pattern; usability

I. INTRODUCTION

Nowadays a large number of universities are producing their courses in terms of learning objects and saving these objects in their own repositories, these repositories in general support several local queries with different criteria thanks to the information in the metadata of every object. Such repositories have the goal of opening access to a wide variety of materials to anyone around the globe [6].

Today, educational content can be easily distributed worldwide via the internet. While this is easy to do, it is apparent that in all of these materials there is both intentional educational content and unintentional cultural content. Some talk about bias in learning materials, but this tends to be pejorative. We prefer to say that culture is in whatever you create. Whenever you create educative content, you put a bit of yourself into the content. This part is often a representation of your own culture and even if you deliver that content in your classroom where you are intimately familiar with your students, there are parts of that content that are difficult for all of your students to understand that have nothing to do with the educational content.

Learning objects are considered as educational resources that can be employed in technology support learning. They are digital pieces of knowledge to put together in order to be part of courses on line. A traditional learning object is composed of five components such as: objective, theoretical knowledge, evaluation, practical knowledge and a metadata [3][6]. A traditional learning object can be enrichment through the clearly identification of cultural factors such traditions, best practices and customs (see Figure 1). These cultural aspects are taken into account since they could help to learn in a more familiar manner to the members of a society. Then, a Multicultural Learning Object (MLO) could be used for different cultures and the learners could better understand in his/her own educational terms.

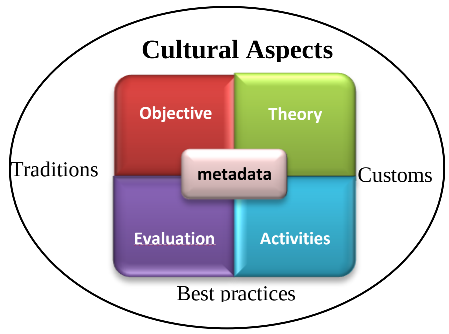


Figure 1. Components of a multicultural learning object.

Monocultural perspectives hamper learning in a number of ways. Students are not exposed to or do not engage with different forms of knowledge or realities. Students are not encouraged to learn about or engage with a diverse group of people. As a consequence, there is no conception of alternative ways of thinking and acting, hindering imagination and criticism, with the potential to promote arrogance, insensitivity, providing a fertile ground for prejudice. Multiculturalism puts forward the existence of multiple (multi) cultures as opposed to one (mono) particular culture within an educational system and society at large. As a countermovement to the established educational system, it attacks a constructed reality that is often referred to as ethnocentrism or monoculturalism. This is essentially to change the conceptualization of culture, moving from unitary to pluralistic perspectives [2] [7].

In the e-learning literature, there are several theories that support multicultural education, but only a few contributions propose to build multicultural learning object to deal with the complexity of instructional design for MLO [2] [12]. This work proposes an architectural model to design and develop multicultural learning objects in an effective manner. For this goal, section two gives an outline of problem; next section presents our proposal in terms of a conceptual model in order give some solutions. Section

 ISBN: 978-1-61208-117-5

four shows the design of a course with multicultural learning object as a case study in order to prove the current proposal. Finally, section five describes a series of related works.

II. OUTLINE OF THE PROBLEM

A multicultural learning object always pursue the idea of reusable interactive components, it is necessary a good quality of graphical user interfaces for different contexts, so a MLO could be delivered for an initial population but also to as many groups with another culture. In this sense the production of MLO's become complex, it is necessary several amounts of educative resources and a big multidisciplinary team with different competences. However, this responsibility is left to the analytical and creative decisions of developers even if this task is not easy to carry out.

This work identifies the following difficulties for the production of multicultural learning objects:

- A major evolution of instructional design for a massive production of multicultural learning objects is necessary.
- It lacks of specification techniques enable to support interactive digital educative content taken into account multicultural context.
- It is necessary identify some new criteria to evaluate the quality of multicultural learning objects.
- It lacks a methodology to develop MLO's where the requirements of different stakeholders are taken into account.
- Multicultural learning objects require a high interactivity, reusability and usability. In particular the treatment of visual feedback requires of several abstraction levels.

Next sections describe some solutions to the aforementioned problems.

III. ARCHITECTURAL MODEL

The design and development of multicultural learning objects requires in particular a big effort in the specification of the graphical user interface in order to support a large diversity of users. According to Shneiderman [1], the user interface is the specification of the look and feel of an interactive application. The production of a multicultural learning object is similar of an interactive application where the user requirements and cultural context are important in particular in the development of graphical user interface. The design of a graphical user interface includes the determination of which interactive components are displayed to the user and the functions that let to interact with these components in order to support user task.

We can assume that the creation of educative material in the form of multicultural learning objects might be strongly influenced by culture, such as, what people value, how they learn, solve problems, put in practice and so forth. Then, all multicultural learning object need to be translated, localized, and adapted in profound ways to suit the needs and preferences of learners in other cultures. Localization addresses obvious visual and textual differences found in other cultures, such as widgets, icons, menu items, symbols, and so forth [15]. Amiel and Orey [7] presented an illustrative example for develop a simple multicultural learning object from an applied mathematics curriculum, that demonstrates that demonstrates the process of computing the area of a home in the United States for the purpose of estimating how much carpet should be purchased. In particular the GUI of this object requires to be modified in order to be used in Brazil to estimate how many boxes of ceramic tile would be needed. While the computational algorithms are the same, the measurement units used in the two countries are different (square yards versus square meters) and the actual floor coverings are also different since in Brazil, carpet is not common, while tile is. Another example is the use of addition operation which could be used in exercises, simulation and evaluation sections of a multicultural learning object. The GUI could be modified in function of cultural context, in Italy we could talk about adding pizzas, hamburgers could be used in the United States, and the typical tacos could be used in Mexico.

Figure 2. Multicultural learning object model based on the PAC model.

In order to specify the interactive nature of a multicultural learning object, we propose to use the architectural model PAC (Presentation-Abstraction-Control) proposed by Coutaz [14]. The different modules of this architectural model could contain the components of a MLO (see Figure 2): the concept to learn, theoretical knowledge, objective, activities and evaluation, all are specified under cultural aspects. The separation of concerns of cultural context, content and graphical user interfaces could be described by the PAC model. The Abstraction could encapsulate the academic content of a MLO, the Presentation module displays the content with cultures, traditions and customs. This is presented in functions of user interactions and requests send to the Control module. Actually, the presentation module specifies the look and feel of the graphical user interface of a multicultural learning object. Then, the graphical user interface could be separated from the academic content (Abstraction) given different representations in functions of user requirements.

It is well known that users could learn of different manner and, in general, different learning styles should be proposed [16]. In addition, we consider here that cultural aspects (traditions, best practices and customs) could help to learn in a more familiar manner, all the members of a society. A PAC agent could represent a particular perspective. The different perspectives require their own resources and their own graphical user interface. In additions, the reusability and maintainability could be feasible using PAC agents without affecting the other properties such as portability and access.

Nowadays, it is possible that several mentors create educative content in different part of a country or world. Internet allows us to find, create easily different good quality of educative contents and deliver them to a large number of societies. A learning object could be created in different locations and in different languages need to be re-built or localized in order to satisfy the requirements of each location. Every new MLO specified under PAC model could be integrated simply in a hierarchical structure of PAC agents, each consisting of a triad of presentation, abstraction and control parts. This helps to structure a course where the agents (or triads) can communicate with each other only through the control part of each triad.

IV. CASE STUDY

Mexico is recognized as a country with a rich culture and a large diversity of traditions; one of the reach traditions is the celebration of day of the dead. This case study shows the specification of a multicultural learning object of day of the dead in Mexico. For this purpose, we present the case study in two sub sections; the first one describes the architectural model at design level and the second sub-section at implementation level.

A. Design level

The celebration of day of the dead are echoed in cities and villages throughout Mexico. As each locality offers distinctive traditions and a unique flavor bound to fascinate the curious traveler, a visit to any Mexican cemetery would be a worthwhile addition to the itinerary of anyone touring the country this time of year [5].

A design of a multicultural learning of the day of the dead requires put clear that this tradition is marked throughout all Mexico by a plethora of intriguing customs that vary widely according to the ethnic roots of each region. Common to all, however, are colorful adornments and lively reunions at family burial plots, the preparation of special foods, offerings laid out for the departed on commemorative altars and religious rites. In general, at elementary school, Mexican students are required to study about their tradition; they could select the multicultural learning object. Consequently, there are a set of MLO's which represents different versions of celebrations of day of the dead in Mexico. This set of MLO's complies with the online course structured by a hierarchy of PAC agents, where the users could select and learn the different manners to celebrate this tradition coming from different regions of Mexico, as follow:

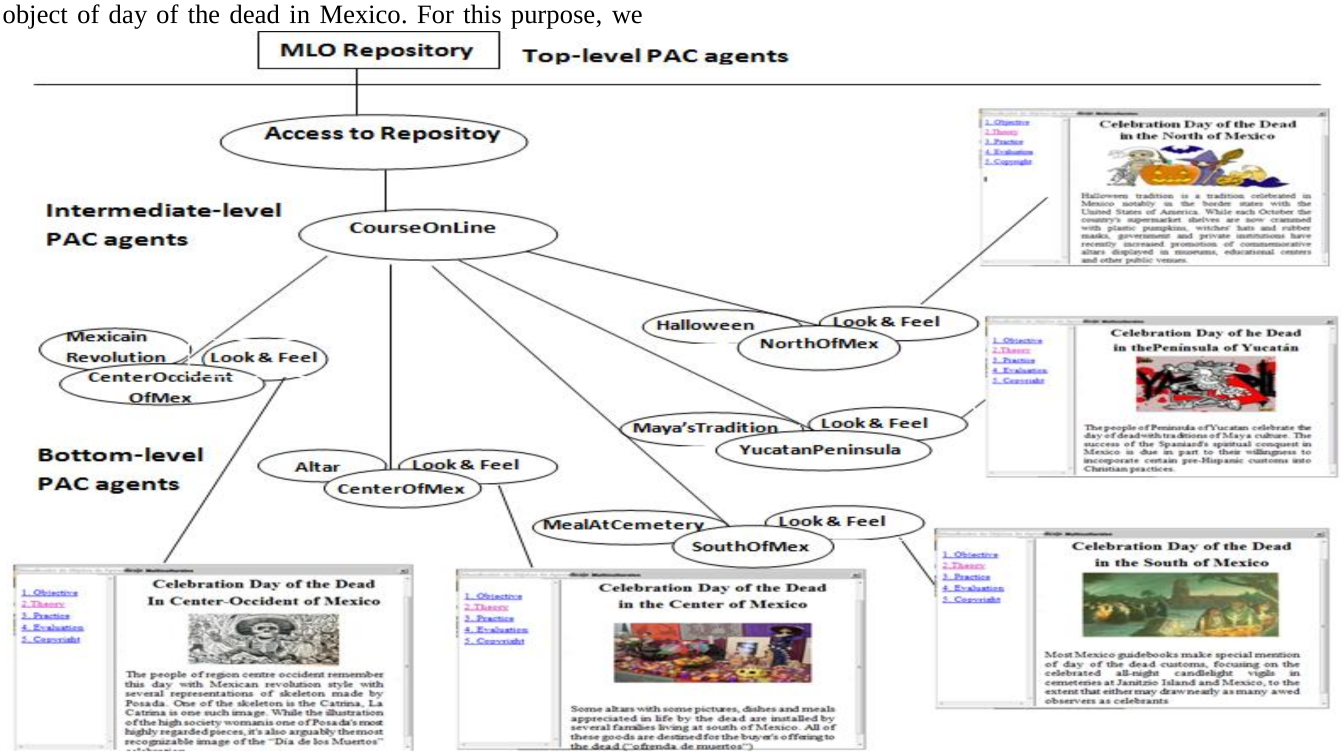


Figure 3. Multicultural learning objects related to the celebration of day of the dead in México.

North of México region.- Halloween is a tradition usually celebrated in Mexico in particular in the border states with the United States of America. Some typical Halloween activities include customs and attending costume parties, ghost tours, and watching horror films. bonfires, visiting haunted attractions telling ghost stories or other frightening tales.

Center of México region.- Most México guidebooks make special mention of day of the dead tradtitions celebrated in the center of Mexico, in particular in the Janitzio Island, familes celebrate the day of dead with dishes, music and night candlelight vigils in the cemeteries.

South of México region.- Some altars with some pictures, dishes and meals appreciated in life by the dead, These altars in honnor to dead of familie with several dishes and flowers are installed in the houses at south of Mexico.

Center-Occident of México region.- The people of region centre occident remember this day with Mexican revolution style with several representations of skeleton. One of the skeleton is the Catrina representig a woman who was member the high society. Also the mexican revolution skeletons are considedered to wear by the people during the day of dead in Center-Occident of Mexico.

Yucatán península region.- The people of Peninsula of Yucatan celebrate the day of dead with traditions of Maya culture incorporating certain pre-Hispanic customs into Christian practices.

The Presentation-Abstraction-Control pattern [13][14] is an architecture that separates the concerns of an application into a hierarchy of cooperating components, each of which are comprised of a Presentation, Abstraction, and Control. The PAC pattern seeks to decompose an application into a hierarchy of abstractions, and to achieve a consistent framework for constructing user interfaces at any level of abstraction within the application.

Figure 3 shows the hierarchical structure of PAC agents where the top and intermediate-level-agents describe respectively the learning object repository and the main window of course on line. Note that every bottom-level agent describes a specific perspective of celebration of the day of the dead in Mexico. For example, this celebration of the people from Yucatan Peninsula is based on the Maya's customs.

B. Implementation level

Concerning to the implementation of multicultural learning objects under the PAC model specified in previous section, it is possible to describe it using the UML notation, as follow:

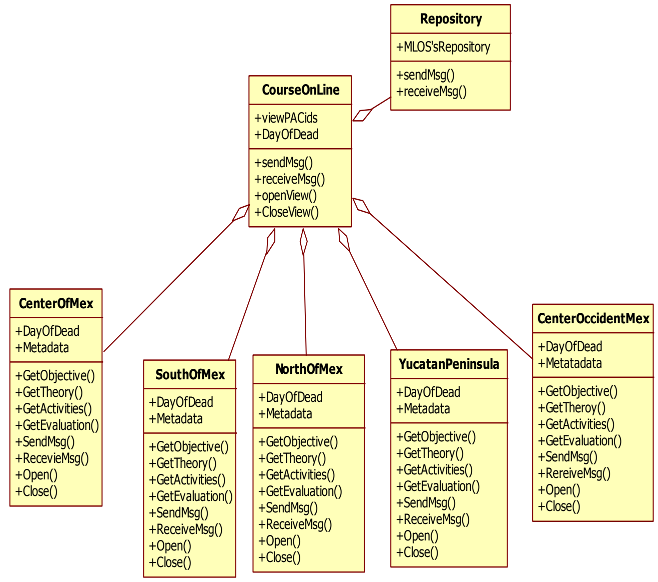


Figure 4. Class diagram of MLO's related to the celebration day of the dead in México

According the UML diagram of Figure 4, the maintenance of the course on line about the celebration of day of the dead is feasible, since the designer of the different regions of Mexico they could update in an independent manner the content of their MLO's. In addition, they could be saved in the repository of this same course on line.

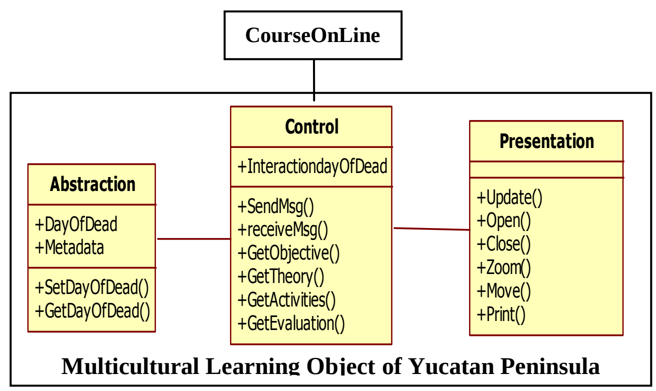


Figure 5. Elements of a multicultural learning object of Yucatan Peninsula according to a PAC agent.

Analyzing the components of a MLO of Yucatan Peninsula (see Figure 5), it is important to identify that a model has their own resources, where the information and metadata of day of the dead are saved as attributes of class Abstraction. This information is accessible by the functions of class Control, such as: *GetTheory(), GetActivities(), GetObjective(), GetEvaluation().* The results could be displayed in the different interactive components of the graphical user interfaces encapsulated in the Presentation class.

V. RELATED WORK

In the literature of e-learning and human computer interaction, some works have proposed for analysis and development of MLO: Agada has considered the multicultural education as an emerging paradigm [8], Oshlyansky et al. have précised the difference between cultural aspect from usability aspect for the graphical user interfaces at analysis stage [10], Goméz et al. have presented a development of mathematic courses in terms of MLO's [12].

PAC architectural model differs from MVC (Model-View-Controller) [4], in that within each triad, it completely separate the presentation (view in MVC) and the abstraction (model in MVC), this provides the option to assign capsule of information to learn which can give the user experience to learn important perspectives and concepts throughout the graphical user interface of a multicultural learning object.

VI. CONCLUSIONS AND FUTURE WORKS

The design of multicultural learning objects for giving solution to software problems is considered as a complex task which is usually left as a programming activity. This work purposed an architectural model to design and develop multicultural learning objects in an effective manner. We propose an architectural model based on the design pattern ¨PAC (Presentation-Abstraction-Control). [13], the set of PAC agents represent the diversity of perspectives, traditions and costumes that could help in a more familiar manner to final user, all this information is accessed and shown throughout the graphical user interface of a multicultural learning object.

Finally, it is important to say, that the adaptability, the evaluation and the massive production of multicultural learning objects are research topics to be considered like future work.

REFERENCES

[1] S. Ben and P. Cathrine, Designing the User Interface: strategies for effective human-computer interaction, Addison-Wesley, 1987. Fourth edition with Catherine Plaisant as co-author 2004, ISBN 0-321-26978-0.

[2] O. Micheal, A. Tel, M. Jaime, R. Rafael, and S. Rene. Much Ado About Culture. Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, 2010, pp. 4026-4033.

[3] M. Jaime, A. Francisco, and C. Ma. Elena, "Tecnología de Objetos de Aprendizaje", Universidad Autónoma de Aguascalientes, 2007, pp 423, ISBN: 970-728-065-4,

[4] B. Steve, Applications Programming in Smalltalk-80 (TM): How to use Model-View-Controller (MVC). 1987, [Online] Available at: http://st-www.cs.illinois.edu/users/smarch/st-docs/mvc.html [Accessed 22 December 2010].

[5] C. Araceli and C. Louis, México de mis amores, Oceano, 2010.

[6] W. David, A. Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. Wiley, 2000

[7] A. Tel, S. Josh, and O. Micheal. (2009). Four Strategies for Designing Instruction For Diverse Cultures. Educational Technology, 49(11), 4-14.

[8] J. Agada, Multicultural education and the emerging paradigm: An essay in cultural epistemology. The Urban Review, vol. 30, 1998, pp 77-95.

[9] G. Erick, H. Richard, J. Ralph, and V. John, Design Patterns : elements of reusable object-oriented software, Addison-Wesley, 1995.

[10] O. Lidia, T. Harol, and C. Paul, Breaking Affordance: culture as context. Proc. Human-Computer Interaction, ACM, NordiChi 2004, pp. 81-84

[11] H. Geert, "Culture's Consequences, Comparing Values, Behaviors, Institutions, and Organizations", Across Nations Thousand Oaks CA: Sage Publications, 2001.

[12] G. Laura, M. Carmen, G. Alma Rosa, and B. Edgard, Multicultural Learning Objects for Mathematic, Proc. Laclo 2010, Brasil, 2010, pp. 489-498.

[13] C. Joëlle,. PAC-ing the Architecture of Your User Interface. 4th Eurographics Workshop on Design, DSV-IS'97, Springer Verlag, 1997, pp. 15–32.

[14] B. Frank, M. Regine, H. Rohnert, S. Peter, and S. Michael, Pattern-Oriented Software Architecture, John Wiley and Sons, vol. 1, 1996, pp. 145–168, ISBN 0-471-95869-7.

[15] M. Aaron and W. Emilie, Crosscurrents: cultural dimensions and global web user-interface design. Interactions, Vol 7, 2000, pp. 32-46

[16] K. David, Experiential learning: experience as the source of learning and development. Englewood Cliffs, Prentice-Hall, 1984.

Rotoscopy-Handwriting Interface for Children with Dyspraxia

Muhammad Fakri Othman
Cardiff School of Art and Design
University of Wales Institute, Cardiff (UWIC)
Cardiff, United Kingdom
faothman@uwic.ac.uk

Wendy Keay-Bright
Cardiff School of Art and Design
University of Wales Institute, Cardiff (UWIC)
Cardiff, United Kingdom
wkbright@uwic.ac.uk

***Abstract*—We discuss the design and development of computer-animated interface for children with dyspraxia using a specialist animation technique known as Rotoscopy. The technique may provide an engaging environment for children with dyspraxia to improve their handwriting skills, thus increase their motivation and self-esteem towards learning.**

Keywords-Animation technique; prototyping, learning difficulties

I. BACKGROUND

Rotoscopy process involves frame-by-frame projection of moving images to allow the animator to copy every single frame to create a natural-looking animation [1]. A famous film that used rotoscopy is *A Scanner Darkly*, produced in 2006 by Richard Linklater using a specially-developed software system called *Rotoshop*. Originally Rotoscopy technique was invented in 1917 by Max Fleisher for his film, *Out of the Inkwell*. It is named after projection equipment called the *Rotoscope;* where pre-recorded live-action film images were projected onto a frosted glass panel and re-drawn by an animator. In other words, rotoscopy refers to the technique of manually creating a silhouette for an element on a live-action plate to be composited over different background [3]. Figure 1 shows how rotoscopy techniques can create a 2D graphics from a photo whilst Figure 2 is a transformation of a video footage to a short 2D animation. In these examples, graphics images as well as human motion and body shape from a video sequences were traced manually using rotoscopy system to get the image silhouettes, which are then been transformed to 2D graphics and animation. From the output, we can create a new 2D graphics and animation that has similar look and key-frames.

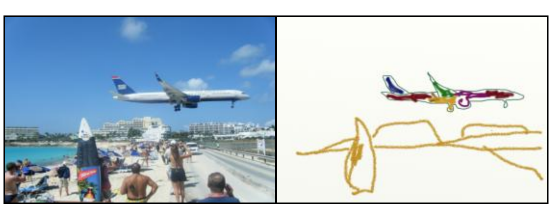


Figure 1. A 2D graphics sketched from a photo

Figure 2. A new 2D animation character created from a video clip

II. RELATED RESEARCH

A. Tracking, 3D Graphics and Animation

Previous research in Rotoscopy has been focused on techniques for 2D graphics and animation, such as *SnakeToonz* [4]. *Keyframe-based Tracking* [2] was proposed as a technique for rotoscopy and animation, which has been improved by *Bi-directional Tracking* [5]. However, both methods produced 2D results and faced similar difficulties. Rotoscopy typically produces non-photorealistic images and animation as it requires more human effort to trace object contours from captured video sequences, whilst tracking is a process in rotoscopy to capture video to be transformed to 2D or 3D graphics and animation. Although a few tracking algorithms have been introduced, previous research in rotoscopy and animation focuses more on video and 2D image rather than producing 3D output. Rotoscopy research for 3D graphics and animation has involved the production of 3D models from video, as well as to improve the image quality. A real-time 3D model acquisition system [6] was developed to assist user to rotate an object by hand and see a continuously-updated model as the object is scanned. However the system has limited functions and required much human efforts. Another research, 3D objects from photographs and video clip [7] has emerged, but it was only for non-animated 3D images. More recently, a new method to capture 3D mesh from 2.5D video [8] has been used, but the result has several limitations such as it cannot distinguish between different objects in the scene, the output cannot be viewed from any direction and there is a flickering problem. A 3D sketching system to create 3D models [9] was developed in 2006 but it has problem to generate correct skeleton and accurate sampling of the silhouette.

 ISBN: 978-1-61208-117-5

B. *Dyspraxia and Handwriting*

Children with Dyspraxia have difficulty planning and organising smooth coordinated movements [10]. Other terms, such as developmental co-ordination disorder (DCD), clumsy child syndrome, minimal brain dysfunction, and perceptuo-motor dysfunction are also used to describe dyspraxia [11]. According to World Health Organisation's (WHO) report, dyspraxia normally affects 6% of all children and the percentage is estimated to increase as high as 10% [12]. However only a few study related to computer animation technique such as rotoscopy have been undertaken to investigate the potential benefits for people with learning difficulties. This gap provides the main objective and motivation of our research. Application allowing children to be creative and expressive with movement will be beneficial as children with dyspraxia will be engaged with activities to practice their movement and gesture ability.

Research on the efficiency of e-learning and multimedia computer technology for children with learning difficulties strongly recommended such approach as it has emerged positive results to the children [13]. Rotoscopy may also appropriate due to the highly visible nature of the graphical output, which is critical for children with dyspraxia who have difficulties in hand movement as well as handwriting. A study of sign language co-articulation has been undertaken using the rotoscopy technique [14] as it promotes movement to perform signs using hands. Walden's research [15] found that technical advances of rotoscopy have had aesthetic consequences not just for the quality of image but also to the nature of the actor's performance within animation.

In terms of handwriting, Boyle [16] has revealed children with *Moderate Learning Difficulties* (MLD) like dyspraxia can be helped to improve their handwriting using simple intervention programme like speed of handwriting exercise, with support of other general gross motor coordination test. Meanwhile an investigation by Snape and Nicol [17] using a pen-based computer writing interface, has given positive improvement to children, though it hasn't employed rotoscopy for their system.

As a summary, early study of rotoscopy has been established in the area of tracking, video and animation. Previous research in rotoscopy has not focused to handwriting, whilst handwriting recognition system hasn't employed rotoscopy as its main solutions nor supporting children with handwriting difficulties. There has been little research investigating the potential of Rotoscopy as a learning tool. However the success of *ReacTickles®* [18] software, a relaxing and playful experience with technology for children with autism, has given us motivation on system design and development.

III. METHODOLOGY

A. *Why Rotoscopy?*

Rotoscopy may provide a suitable handwriting application as it allows the user to trace hand movement along specific guided lines. Furthermore, rotoscopy enables children to learn through play via tracing and drawing of still images and video footage. The flexibility of rotoscopy is another benefit for children with dyspraxia as it allows many novel and developmentally appropriate inputs. For instance we can use a favourite character as a background image as a practice for hand movement, or a live video of children faces and tracing it using rotoscopy system.

B. *Experience Prototyping*

We are currently developing series of prototypes using Experience Prototyping [19]. This method tries to demonstrate content and identify issues and design opportunity. These normally can be achieved by exploring through direct experience of the systems. The designer need to apply the concept of "exploring by doing", especially to get high fidelity for simulation developed from existing experience (actual products) that cannot be accessed directly due to safety, availability and budget issues or constraint. We have to consider several factors such as contextual, physical, temporal, sensory, social and cognitive issues, the essence of existing user experience and essential factors that our design should preserved. At early stage of developing a user experience, multiple design directions need to be efficiently prototyped and compared. Temporary or ad-hoc use of analogous objects as props can help to decide which kind of experience is most appropriate. The main purpose of this phase may includes facilitating the exploration of possible solutions, directing the design team toward a more informed development of user experience and identifying tangible component who creates the system. The role of experience prototyping in this level is to let our client, design colleague or user to understand subjective value of design idea by directly trying it in their own, or in other words to let user experience the product. It is usually accomplished by persuading the audience or user such as saying that the idea is compelling or the chosen design direction is not right or incorrect.

C. *Usability Testing*

The implementation of prototypes and the testing process will be held at the Dyscovery Centre, University of Wales, Newport, a special research centre for children with dyspraxia and other learning difficulties. The metrics of testing are video coding and observation, involving series of different version of prototypes. We use *Wechsler Intelligence Scales for Children®* (WISC) [21] as the evaluation standard.

WISC is a special intelligence test designed for children age 6 to 16 to generate their IQ score as well as to diagnose children's learning disability. The results will be measured on children's performance, and how the application helps users improve their gross-motor skills, as well as their interest and motivation towards learning. For video coding, we will record children activity with the system during series of workshop held at the Dyscovery Centre. This will includes a record of children handwriting process to allow a replay and analysis of their result.

D. *Medium of interaction*

The interactive whiteboard has been used in classroom to test early prototypes as well pre-handwriting activities. It has been chosen as it is easy to use as it has touch screen mode that allows children to make free hand movements on the screen [22]. The concept is quite similar to pen and graphics tablet but in more spacious movement area. Previous research using pen-based computer writing interface [23] has considered it was as efficient as, and produced comparable writing to the pencil and paper. Figure 3 shows example of different size of interactive whiteboards including children playing with the device.

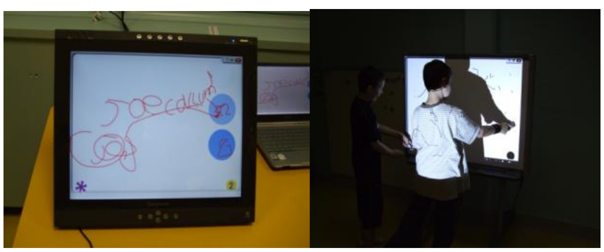


Figure 3. Different size of interactive whiteboard

IV. Rotoscopy-Handwriting INTERFACE

At the moment we are investigating rotoscopy technique to identify whether motivation and engagement can be increased by using a performance-led embodied approach, which would allow children to practice handwriting using gross motor skills, for example bodily movement and gesture; as our research aims to give children with dyspraxia a playful and physical experience in developing their handwriting skills. The development and implementation of the systems have been focused on the benefits of using interactive whiteboards [24], as large-scale environments that support gross-motor skills and performance. Rotoscopy technique has been used to trace a projected image on an interactive whiteboard screen. The prototype system developed using 2D and 3D animation software has been utilised and adapted to respond to user activity in such a way that the mirrored or abstracted image enables the child to visualise the activity and to recognize the projected actions own.

This mirroring capacity of the system is vital for the development of social communication and learning [25]. For this reason rotoscopy is chosen to be used as a technique to assist children with dyspraxia, who are known to experience problems with low self-esteem as a result of co-ordination difficulties [26].

A. *System Pipeline*

Figure 4 illustrates the system pipeline which contains the research problem, objective, methods and solution. In other words our research tries to prove the effectiveness of rotoscopy as an appropriate solution to assist coordination difficulties and to improve handwriting skills among children with dyspraxia. It is clear that the study occupies user-centred design and experience prototyping as rotoscopy-handwriting prototypes has been used as a tool to improve handwriting skills among children with dyspraxia as well as to support their gross-motor skills.

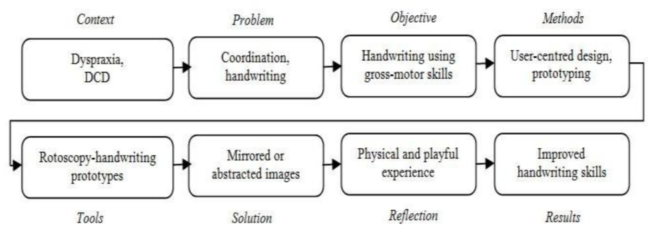


Figure 5. The rotoscopy handwriting system pipeline

B. *Prototype system*

The prototype system is designed to assist children with dyspraxia to practice their handwriting skills through movement and performance, rendered visible to them via the projection of actions onto a large surface. Evaluation methods will measure the extent to which the animation of movement in a child-friendly graphic form can engage children's motivation and lead to improved handwriting skills [23] The prototype has been designed based-on Ripley's Methods for teaching of handwriting to children with dyspraxia [11]. Ripley's method consists of different stages of skills, starting from drawing basic shapes and symbols to get familiar with lines and curves, followed by writing basic numbers, simple to more complex letters, and ending with writing children's own name.

We divided this approach into 5 different level of prototype: Prototype 1: Group 1 to Group 4 Symbols, Prototype 2: Numbers, Prototype 3: Circle Letters, Prototype 4: Wiggly Letters and Prototype 5: Child's Name. Each stage has different level of difficulties that need to be completed before the student can move to next higher level. Figure 3 and Figure 4 are examples of early version of the prototype interfaces that contains the first two groups of basic shapes.

These shapes are a basic hand movement and gesture practice for children with dyspraxia before they learn more complex letters. Children's interaction with the system occurs with the use of interactive whiteboard to test Rotoscopy-Handwriting prototype. For instance, in Prototype 1, Group 1 to Group 4 symbols, children or user need to master basic lines and shapes in order to train their hand movement. Group 1 consist of three most fundamental shapes, a horizontal line for hand movement from left to right (and vice versa), a vertical line for hand movement from top to bottom and a circle for round or curve hand movement.

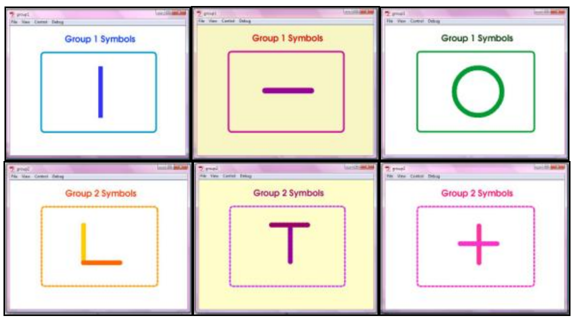


Figure 4. Group 1 and Group 2 Symbols

Figure 5 depicts examples of the first three interfaces of Group 1 symbols. The system will allow children to draw horizontal and vertical lines as well as circles on interactive whiteboard using early the version of rotoscopy-handwriting prototype. Every interface has several number of lines or circle to give more chances to children to practice. Different types of line-pattern as well as colours have been used to engage children to learn through play.

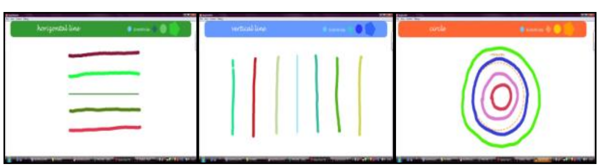


Figure 5. Prototype 2: Group 1 symbols

Meanwhile examples of learning to write a single letter are shown in sequences from start to end (Figure 6 to 8). In Figure 6, the first interface is steps with arrows to show how to write the letter and the second is demo animation of the steps. In Figure 7, the interface shows the rotoscopy process, where users are allowed to trace the letter, followed by two interfaces for exercise as shown in Figure 8. For exercise, the first interface allows user to train to write the letter with reference or guide (a watermarked image of the letter), but the second is an exercise without references, assuming the user already know how to write the letter.

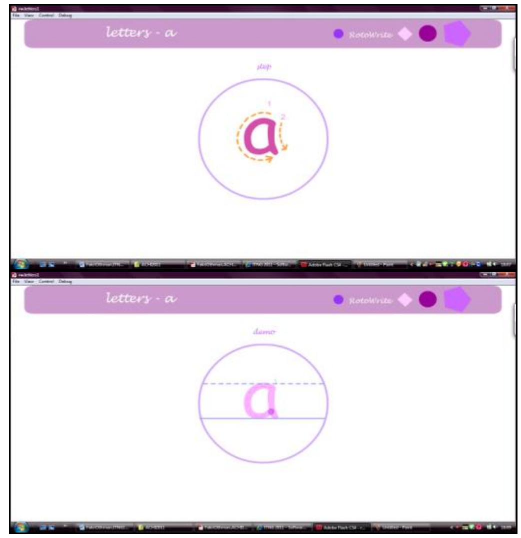


Figure 6. Steps and demo animation

Figure 7. Rotoscopy interface

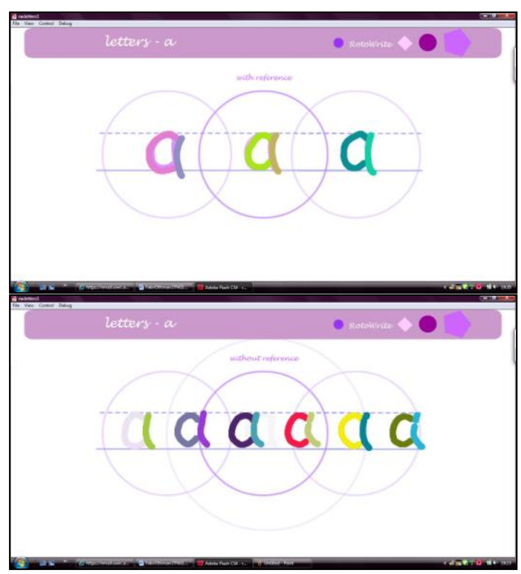


Figure 8. Sample of rotoscopy exercise

 ISBN: 978-1-61208-117-5

C. *Pre-writing interface*

We have included a pre-writing interface to see how children with dyspraxia and typical-developed children response to system's design and usability. Figure 9 shows drawing results obtained from a group of typical developed children age 6 to 9 using the screen and playing with rotoscopy function. In this activity, children were encouraged to produce free drawing based on basic shapes from Ripley's symbols.

Figure 9. Sample result of rotoscopy pre-writing interface

We have also tested the children to draw more complex images, as shown in Figure 10, as they traced real photo to get 2D graphics and images. Their favourite cartoon character has been used as attraction to engage them to trace and draw pictures. It is emerged that such an activity was really fun and interesting for them. From this experiment or testing we try to adapt system design for children with dyspraxia which should be simple, easier and less complicated. The next version of our prototype will use a pen tablet and a monitor. Children will be able to use their fingers as well as the pen to draw lines and shapes on the screen. The size of the shape should be drawn according to the screen's medium of display; for example when using interactive whiteboard; a bigger shape size is required whilst for monitor-based display, smaller size is preferred.

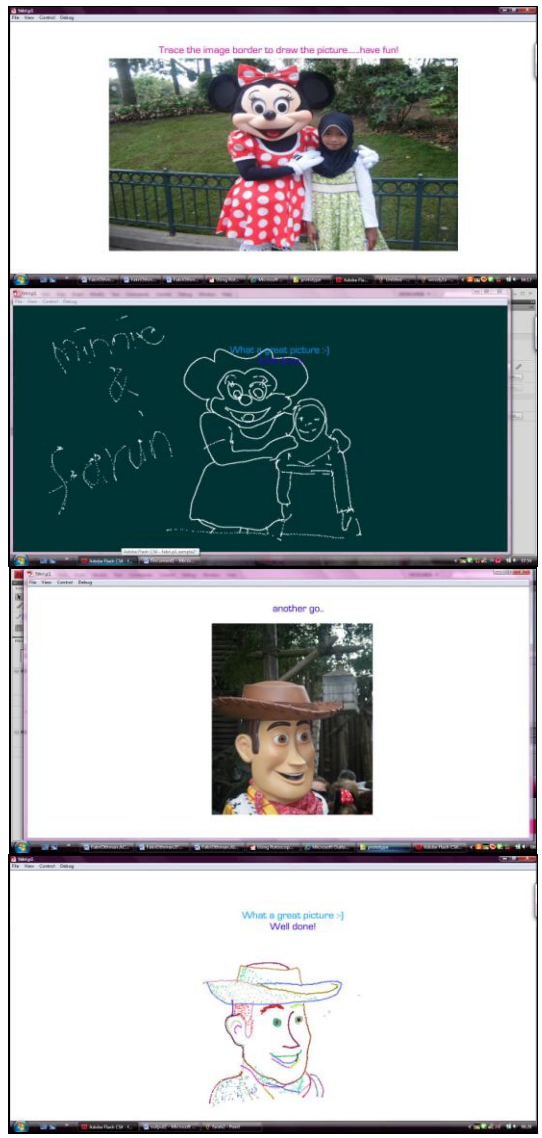


Figure 10. Sample rotoscopy result of cartoon character

Copyright (c) IARIA, 2011. ISBN: 978-1-61208-117-5

V. Conclusion And Future Work

The goal of this investigation is to discover the potential of rotoscopy to assist the teaching of handwriting skills for children with dyspraxia. With evidence gathered from our prototype systems, the proposed methods may place these developments in an original context. Rotoscopy may prove appropriate for children with dyspraxia as it enables them to practice their skills using naturalistic hand movement. In this case, the input and output device play important role as a medium of communication between children and the methods; as this tangible technologies have a close correspondence in behavioural meaning between input and output. Our next milestones will be to undertake contextual analysis [27] of the prototype systems which will include end user feedback and practitioner reflection.

Acknowledgement

A special thanks to University of Wales Institute, Cardiff, The Dyscovery Centre, University of Wales Newport, Universiti Tun Hussein Onn Malaysia, Ministry of Higher Education Malaysia, PhD supervisory team and our beloved family for their supports.

References

[1] V. Garcia, E., Debreuve, and M. Barlaud, "Contour Tracking Algorithm For Rotoscopy," Proc. ICASSP 2006: IEEE International Conference On Acoustic Speech.

[2] A. Agarwala, A. Hertzmann, D. Salesin, and S. Seitz, " Keyframe-Based Tracking for Rotoscoping and Animation," Proc. SIGGRAPH, Aug. 2004, pp. 584—591, doi: http://doi.acm.org/10.1145/1186562.1015764.

[3] J. Bouldin, "The Body, Animation and the Real: Race, Reality and the Rotoscope in Betty Boop," Proc. Affective encounters: rethinking embodiment in feminist media studies. Eds. A. Koivunen and S. Paasonen. Turku, FI: Media Studies, 2001, pp. 48-54.

[4] A. Agarwala, "SnakeToonz: a semi-automatic approach to creating cel animation from video," Proc. 2nd ACM International Symposium on Non-photorealistic Animation and Rendering, 2001, pp. 139-ff.

[5] J. Sun, et al, "Bi-directional tracking using trajectory segment analysis," Proc. SIGGRAPH 2005.

[6] K. Nickel, et al, "3D-tracking of head and hands for pointing gesture recognition in a human-robot interaction scenario,". Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 565-570.

[7] J Ponce, et al, "3D photography from photographs and video clips," Proc. International Symposium on Core Research, 2003.

[8] N. Snavely, et al, "Stylizing 2.5-D Video," Proc. NPAR 2006, pp. 63-69.

[9] F. Levet, et al, "3D Sketching with Profile Curves," Proc. International Symposium on Smart Graphics, 2006.

[10] M. Boon, Helping Children with Dyspraxia., Jessica Kingsley Publishers, 2001.

[11] K. Ripley, B. Daines, and J. Barrett, Dyspraxia: A Guide for Teachers and Parents., David Fulton Publishers, 1997.

[12] M.M. Portwood, Understanding developmental dyspraxia: a textbook for students and professionals., David Fulton Publishers, 2000.

[13] A. Savidis, and C. Stephanidis, "Developing inclusive e-learning and e-entertainment to effectively accommodate learning difficulties," SIGACCESS Access. Comput. , 83, Sep. 2005, 42-54.

[14] J. Segouat, "A study of sign language coarticulation," SIGACCESS Access. Comput. , 93, Jan. 2009, 31-38, doi:http://doi.acm.org/10.1145/1531930.1531935.

[15] K. L. Walden, "Double Take: Rotoscoping and the processing of performance," Refractory: a journal of entertainment media, Dec. 2008.

[16] C.M. Boyle, "An Analysis Of The Efficacy Of A Motor Skills Training Programme For Young People With Moderate Learning Difficulties," International Journal Of Special Education 2007, Vol 22, No. 1, pp. 11-24.

[17] L. Snape, and T. Nicol, T., "Evaluating the effectiveness of a computer based letter formation system for children," Proc. The 2003 Conference on Interaction Design and Children, 2003.

[18] W. Keay-Bright, "Can Computers Create Relaxation? Designing ReacTickles© Software with Children on the Autistic Spectrum," CoDesign, 3, (2), 2007, pp. 97—110.

[19] M. Buchenau, and J. F. Suri, "Experience prototyping," Proc. The 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques; August 17 - 19, Aug. 2000, doi: http://doi.acm.org/10.1145/347642.347802.

[20] M. Buchenau, and J. F. Suri, "Experience prototyping," Proc. The 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques; August 17 - 19, Aug. 2000, doi: http://doi.acm.org/10.1145/347642.347802.

[21] E. Kaplan, D. Fein, and J. Kramer, "Wechsler Intelligence Scale for Children® - Fourth Edition Integrated," 2004, Available at: http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8982-800

[22] W. Keay-Bright, "Designing Playful Sensory Experiences with Interactive Whiteboard Technology: The Implication for Children on the Autistic Spectrum," EAD 07, 2007.

[23] J.C. Read, S. MacFarlane, and M. Horton, "The Usability of Handwriting Recognition for Writing in the Primary Classroom," People and Computer XVIII - Design for Life, Springer, London, 2004, doi:10.1007/b138141.

[24] M. Horton, and J. C. Read, "Interactive Whiteboards in the Living Room? – Asking Children about their Technologies," The 22nd BCS British-HCI 2008, Liverpool, UK.

[25] K. Facer, and B. Williamson. "Designing Educational Technologies with Users", Futurelab, 2004, Available at http://www.futurelab.org.uk/resources/documents/handbooks/designing_with_users.pdf

[26] S.E. Henderson, and L. Henderson, "Toward An Understanding of Developmental Coordination Disorder: Terminological and Diagnostic Issues," Neural Plasticity 2003, Vol.10, No. 1-2.

[27] M. Stringer, E. Harris, and G. Fitzpatrick, "Exploring the space of near-future design with children," Proc. The 4th Nordic conference on Human-computer interaction: changing roles, Oct 2006, pp. 351-360, doi:10.1145/1182475.1182512.

An Emotional System for Effective and Collaborative e-Learning

Kittanakere Lakshminarasimhamurthy Nitin, Lakshmisri Lakshmi Narayana Rao, Nirmal Kumar Sivaraman

E-Commerce Research Labs, Education & Research,

Infosys Technologies Ltd,

Bangalore, India

KL_Nitin@Infosys.com, Lakshmisri_LV@Infosys.com, Nirmal_Sivaraman@Infosys.com

***Abstract*: Though e-learning is being debated to have some advantages over person to person teaching, the latter is considered to be superior with respect to the effectiveness of teaching. One of the reasons for this advantage of human expert tutors is their ability to deal with the emotional aspects of the learner. We introduce an emotion sensitive e-learning model which is sensitive to both emotional aspects as well as the learning ability of the learner. This is a major difference with the other similar attempts to make effective e-learning systems and the preliminary analyses indicate considerable promise. We hope this would complement the effectiveness of e-learning.**

Keywords - e-learning; emotional systems; emotional e-learning

I. INTRODUCTION

Nowadays, e-learning is becoming an increasingly popular paradigm of learning. This is very evident especially in the corporate training arena. And it is expected that this paradigm would eventually become mainstream channel of training which would add to a company's competitive advantage [6, 16].

The main goals of e-learning systems identified by different researchers are given below [24, 25].

- Focus on active learning
- Accommodate various learning styles
- Explicitly place the responsibility for learning with the students
- Develop written and oral communication skills
- Clarify the role of the teacher as facilitator and mentor
- Provide better coverage of material
- Develop a sense of self-confidence and independence in students
- Include a teamwork experience
- Encourage peer review
- Develop interpersonal communication skills when students are geographically spread
- Support entire educational process when students are spread both geographically and temporally
- Learn to handle time management including the meeting of deadlines

Many of the above goals themselves explain the advantages of e-learning systems over traditional learning approaches. Another advantage of e-learning systems is that they are scalable. The number of learners that an e-learning system can handle with individual attention is much more than that can be accommodated in a classroom.

However, they have some disadvantages as well. During person-to-person communication, more than 65% of the information exchange happens through nonverbal communication [11]. These include facial expressions, body posture, tone of voice etc. So, the learner gets less information through e-learning system than through person – to – person interaction.

Personality and emotion has a role to play in learning. The learning styles that can be more effective for a learner may be different according to these aspects [7]. Emotions can influence a person's cognitive organization and thought process. Positive emotions play a crucial role in these aspects of one's intelligence. They are also instrumental in improving creativity and flexibility in problem solving. At the same time, negative emotions can produce the opposite effect on one's thought process [3].

Expert human tutors judge these factors and teach accordingly. This is a major difference in the style of learning with traditional e-learning systems and that with human expert tutors. Human expert tutors focus on the emotional component of learning also, whereas the traditional e-learning systems focus on learning targets alone [12]. This makes the human expert tutors more effective. But, it would be impossible for a human tutor to give attention to every student, if he is handling a large number of them.

 ISBN: 978-1-61208-117-5

This motivates us to think about incorporating emotional aspects of teaching in e-learning systems to make it more intelligent. An intelligent e-learning system should be able to adapt to the knowledge, learning abilities and needs of each learner [26]. This would give them a feel of individual care which would help them in the learning process.

Incorporating emotions in e-learning systems is not a very new idea. “affective e-learning systems” had been found in the literature since the book ‘Affective Computing’ which was published in 1997 [15]. Many e-learning systems have been introduced with affective agents and different types of emotion sensing techniques [2, 4, 14, 17, 20, 22, 26]. The emotion sensing techniques used in such models include facial expression analysis [13, 26, 27], voice analysis [5, 8, 26, 27], text analysis [18, 26], analyzing the physiological data and behavior [10, 19, 22], etc. Some of them require costly equipments for effective functioning. This is an overhead over the learner.

The current trends in e-learning are personal learning environments [9, 23], self directed learning [21], networked learning [1], etc. But these approaches may not be appropriate in case of a small duration – high content learning as observed in corporate trainings.

In this paper, we introduce the design of an emotion sensitive e-learning system that gives emphasis to the complete learning process and that is very cost effective. This model is made as configurable as possible to make it suitable for a wide range of learners and to give them a feel of individual attention. This is the major advantage of this model over the others.

II. ARCHITECTURE

In this section we will discuss the architecture of the emotion sensitive e-learning system. Fig. 1 shows the architecture of the e-Learning system. This system mainly consists of three major units namely the *Course Delivery* Unit, the *Assessment* Unit and the *Report* Unit. There are two more additional components called the *Emotion Modulators* and the *Focus Analyzer* which play a vital role in the system. The individual units will be discussed in the subsequent sections.

A. *System Roles*

The various identified system roles, which have a very unique role to play in the system, are: System Admin, Learner and Course Owner.

1) *System Admin*

The System Admin is responsible for managing the entire system. He has several responsibilities which include configuring the emotion assessment, emotion modulators, parameters for the report and time for which the learner can indulge in any of the emotion modulators. He is also responsible for choosing the number of questions for emotion detection and technical analysis.

2) *Learner*

Learner is the individual who takes up the eLearning by registering himself to the e-Learning system. The Learner can choose any of the courses provided by the System. The Learner is assumed to be responsible enough to make the best use of the system. So, the learner is provided with the flexibility to configure the system to a certain extent which includes the following:

He has the freedom to override the system’s suggestion and choosing the course delivery style of his/her choice.

He also has the flexibility to choose the emotion modulators; the time is however fixed by the System Admin.

He can skip to the next granule after completing the current questionnaire.

He is as well free to turn off the emotion modulators.

3) *Course Owner*

Every individual course will be designed and developed by the respective Course Owner who is responsible for configuring the system .His responsibilities includes providing the different granules as per the different styles, deciding the format of each granule, providing with a Technical Question bank and also the necessary emotional plugins

B. *Course Delivery Unit*

This unit comprises of several course offerings in different learning styles. The course offering focuses mainly on two learning aspects namely: Emotional State, Technical Understanding and Grasping Power.

The *Emotional Plugins* are designed to tune the learning style to the current emotional state of the learner. The system concentrates on broadly three emotional states of the learner which are judged by the Emotion Sensing module of the Assessment Unit: Happy, Neutral, and Sad.

 ISBN: 978-1-61208-117-5

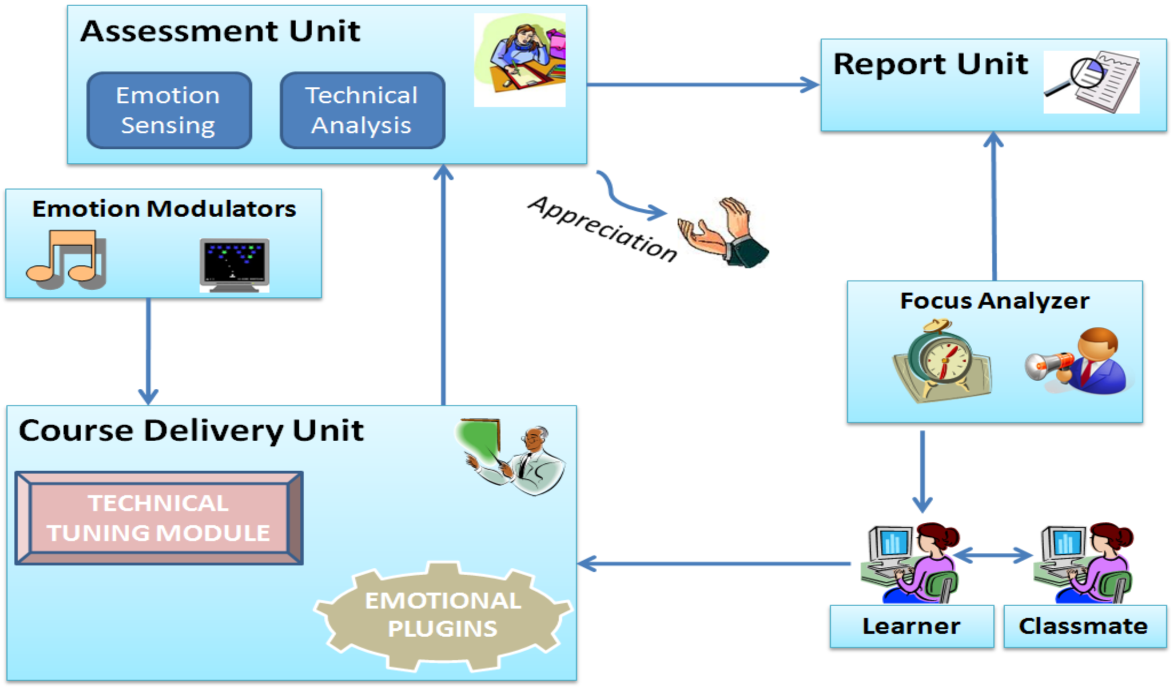


Figure 1. The architecture

The *Technical Tuning* module focuses on the technical understanding and the grasping power of the learner. The same is evaluated by the Technical Analysis module of the Assessment unit which is categorized into the following: Slow, Medium, and Fast.

Every learner is provided with a different style of course delivery which is specially tuned to his/her technical understanding and emotional state as described above. The learning style for a course is categorized into the following Emotional and technical modes:

a. Emotional Modes

We have three emotional modes that are categorized namely Emotional Mode 1, Emotional Mode 2 and Emotional Mode 3. Emotional Mode 1 is for a happy and excited listener, Emotional Mode 2 is for a listener with a neutral state of mind and Emotional Mode 3 is for a listener who is in a sad state of mind.

To explain the impact of these emotional modes on the learner, let us discuss each of these modes one at a time.

Let us consider the learner to be in emotional Mode 1(happy and excited state). In this case, the reason behind the happiness and excitement of the learner could be the learning satisfaction. At the same time, the reason could also be something totally external that has no relevance from the current learning that he is undergoing. The former reason will definitely be conducive for learning; however the latter will be a distraction.

To confirm the same, after every granule, the learner score is examined. If the learner is scoring less and if his emotion is detected to be "happy", in that case, the system concludes that the learner is distracted. Suitable emotional plugins and Emotion modulators are inserted which will aim at getting the learner's emotion to "neutral". Emotional Plugins and modulators are explained in the further sections.

On the contrary, if he scores well, we conclude that he is happy about the learning process and not distracted.

Similarly, when the learner is in Emotional Mode 3(sad state), he is again not in a favorable mode to learn since sadness may distract him and reduce his learnability. Appropriate emotional plugins and emotion modulators are provided again to get him back to neutral state of mind.

Hence, the system considers Emotional Mode 2 as the most encouraging mode for learning, since here the learner is neither happy nor sad. There will not be distractions for learning from the emotional point of view.

b. Technical Modes

We have three technical modes that are categorized namely Technical Mode 1, Technical Mode 2 and Technical Mode 3. Technical Mode 1 is slow and encouraging paced flow of the course delivery, Technical Mode 2 is medium paced flow of the course delivery and Technical Mode 3 is fast and challenging paced flow of the course delivery.

Any learning style provided to a learner will be a blend of the above mentioned emotional and technical modes resulting in a 3 X 3 combination. The *Course Delivery Unit* is externally assisted by an additional *Emotion modulator* (*apart from the Emotional Plugins*) at regular intervals which helps the learner in better focusing at regular intervals.

1) Emotional Plugins

The emotional plugins are a set of objects which the system suggests to be plugged-in to the technical tuning module depending on the emotional state of the learner. These plugins help in modulating the mood of the learner enhancing the effectiveness of the learning process, to name a few of these plugins; we have Animation and graphics, Cartoons, Pictures and others.

The distinctive feature of the system is that it suggests the preferred configuration of the emotional plugins but however it provides the optimum flexibility to the learner to turn-on or turn-off or tune any of the plugins of his/her choice. Each of these plugins is suggested at regular intervals by the system, as per the current emotional state. For an instance, a learner with a sad state will be suggested with more of animations, pictures etc. The appearance of the system could also be changed to generate a better look and feel by providing different themes to the learner. Similarly other plugins can also be tuned accordingly.

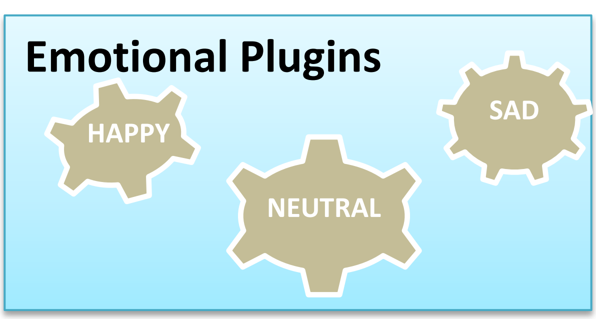


Figure 2. Emotional plugins

2) Technical Tuning Module

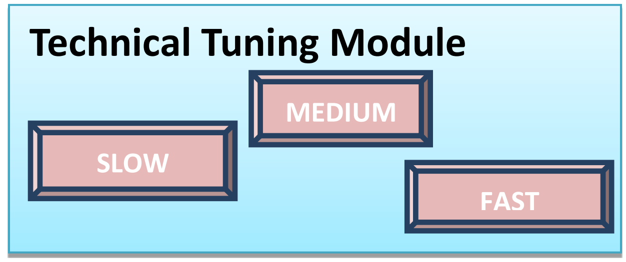


Figure 3. Technical tuning module

Technical Tuning module concentrates on the actual course delivery. In order to enhance the effectiveness of the learning process, every course is dealt at a granular level. A granule is the smallest unit within which the mode switching is possible. To support the different technical modes (as discussed in 2.1), the tutor provides 3 different offerings for each granule corresponding to the 3 different technical modes which could be in any of the desired formats like a video/audio recording, a text-write up etc.

The system enables the learner to repeat the granule any number of times. The learner is not allowed to skip to the next granule without completing the questionnaire for that granule. This is to ensure that the learner gets a better clarity on each of the granules.

The first granule of the course begins with a medium paced course offering. After each granule, the assessment unit interacts with the learner and assesses the level of the learner's grasping power. Based on its outcome the system gives a suggestion to the learner to opt to a more suitable technical mode. Here again the system provides the flexibility to the user to continue/switch the same/different mode of his/her choice. As described in Section 2.1.1, the emotional plugins are interleaved with the granule offering to cater to his emotional needs as well.

The technical tuning for every granule is realized in terms of the various technical illustrations, demos, assignments, hands-on, depth of the each concept dealt etc. In case of slow paced mode the learner is supplied with elaborate illustrations, more demos and given more time on solving the assignments. On the contrary less emphasis is given on illustrations, assignments and the like in the fast paced mode.

C. Assessment Unit

The assessment Unit comprises of the following modules: Emotion Sensing Module and Technical Analysis Module

The emotion Sensing and technical Analysis modules gauge the learning capabilities and emotional state of the learner. Though these function as two separate modules internally, they are interleaved together into a single unit.

This unit presents a questionnaire comprising of questions that are intended to analyze the technical as well as the emotional state of the learner. This exercise is repeated at the end of every granule. The outcome of this process is fed into the Course Delivery Unit which suggests the style of the next granule. At the same time the result is forwarded to the Reporting Unit (Explained in the subsequent sections).

1) Technical Analysis Module

This module chooses the technical part of the questionnaire from a repository. The questions are categorized on the basis of their complexity into the following three types: Simple, Medium and Complex.

The system fires a medium complexity question to the learner. The complexity of the next question is dependent on the answer provided by the learner. The system gives a complex question if the medium complexity question is answered right while it throws a simple question if the learner answers it wrong. This cycle continues until a definite number of questions are posed to the learner. At the end of this exercise the system suggests the technical mode that the learner could choose in order to have an effective learning. This is done by analyzing which category (simple, medium or complex) of questions did the learner answer the most number of times.

The system gives appreciation to the learner based on the outcome of the technical questionnaire. It also appreciates the learner whenever he/she outsmarts the virtual classmate (virtual class mate is explained in the subsequent sections).

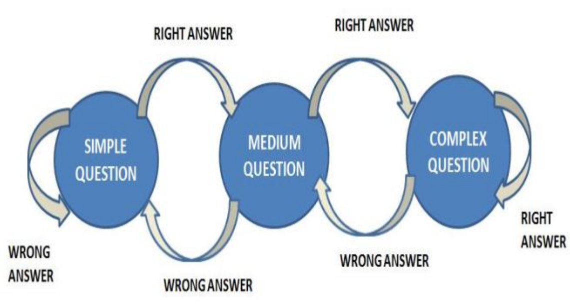


Figure 4. Selection of questions in technical analysis module

2) Emotion Sensing Module

This module is responsible for choosing the part of the questionnaire from a repository for detecting the current emotional state of the learner. The system categorizes a learner's emotional state into the following: Happy, Neutral, and Sad.

The questionnaire to categorize the emotional state of the learner depends on many factors like the learner's culture, ethnicity, the languages that he is comfortable with, the education level of his parents etc. Verified and cross verified questionnaires are made that would help us to categorize the emotional state of different types of learners to happy, neutral or sad. This is done with the help of a psychologist. Most of the times this questionnaire would be an indirect way of asking the learner what emotional state is he in. The details are obtained from the learner when he registers to the system. Based on this data, the questions are selected from the suitable repository and merged with the technical questions selected by the technical analysis module.

The current emotional state of the learner is predicted based on the outcome of these questions.

D. Reporting Unit

The Reporting unit is responsible for providing the final report to the learner. This report consists of the following parameters which are configurable by the System Admin. The System Admin can choose to turn on/off any of the following parameters based on the requirement: The latest score of every granule, the time taken to solve the questionnaire, the duration the learner invests on the Emotion Modulators, the outcome of the Focus Analyzer component (which is discussed later), the number of times and the granules in which the learner has out merited the classmate.

E. Emotion Modulators

This component is designed with an intention of responding and reacting to the learner emotionally. It comprises of various constructive entertainment fragments. These modulators could be chosen by the learner at the end of every granule. The System Admin should configure the time for which these modulators can be played by the learner.

The idea behind the emotion modulators is to give assistance to the learners in achieving a continued focus on the learning process. They are basically intended to bring back the learner to a mood which is conducive for

 ISBN: 978-1-61208-117-5

learning. These are to be seen as a help for the learner to bring back his emotional state to the desired one. However these modulators need not necessarily assure the change in the emotional state of the learner.

The emotion modulator unit reads the emotion that is detected from the Emotion Assessment Unit. This helps the emotion modulator unit to understand whether the current emotion of the learner corresponds to happy, sad or neutral.

The emotion modulators are categorized based on the current emotion of the learner. The categories include Focus Emotion modulators and Cheerful Emotion modulators. The former ones are intended for learners with a happy emotion and are yet not performing well in the technical modules and the latter ones are aimed at learners whose current emotion is “sad”.

The emotion modulators supported by the system are: Music, games, videos and write ups. A list of these modulators is maintained in the database and the collection is shown to the user. The user may select the one of his choice and soon after the selection the selected modulator is played. However, the time for which the modulator is played, is restricted.

The Learner is however free to turn off the emotion modulators if he/she is not interested.

Also, providing these different emotion modulators is not an overhead on the course author since it is taken care by the system.

F. Focus Analyzer

This component periodically examines the presence and the focus of the learner towards learning. This ensures both the physical and mental presence of the learner. This is achieved in an effortless fashion by sending frequent alert messages to the learner. Alert messages could be in the form of simple technical/informal questions. An explicit response from the learner’s end ensures his/her presence.

The time at which these alert messages have to be sent out is configured by the System Admin suitably. The outcome of this whole process is recorded regularly by the Reporting Unit.

G. Virtual Classmates

This is to build a *collaborative* environment for the learner for an effective learning. This is intended to motivate the learner. The learner is always provided with a feel of virtual classmate to get a typical classroom feeling. The scores of the learner after individual granules are compared with the virtual classmate and the learner will be suitably appreciated if he/she excels. The system is also configurable to arrange for a quiz with the classmate and the winner will be decided based on the time taken to respond. The System Admin can configure the number of classmates and also has a hold on turning on/off the whole of this feature.

III. CONCLUSION

In this paper, we have described an emotion sensitive e-learning model which is sensitive to both emotional aspects as well as the learning ability of the learner. The model is designed to be configurable to a good extent in order to make it suitable for a wide range of users. In future this model can be extended by adding more emotion sensing methods like facial expression analysis, click stream analysis, voice analysis etc to confirm the emotion sensed through the survey method. There are many more challenges in this area like customizing the course delivery according to the personality of the learner, reducing the overhead on the course owner in course preparation, more dynamism in mode switching etc.

IV. ACKNOWLEDGEMENT

We thank Sh. Subrahmanya SV and Dr. Sakthi Balan Muthiah for reviewing this paper, their valuable comments and for their constant encouragement.

V. REFERENCES

[1] A. D. Malayeri and J. Abdollahi, New designing of E-Learning systems with using network learning, url: http://arxiv.org/ftp/arxiv/papers/1003/1003.3097.pdf, last accessed date: 01 – Feb – 2011

[2] C. M. Chen, Intelligent web-based learning system with personalized learning path guidance, *Computers & Education, 51,* 2 (2008), pp. 787 – 814

[3] C. M. Jones and I. M. Jonsson, Automatic Recognition of Affective Cues in the Speech of Car Drivers to Allow Appropriate Responses, Proceedings of OZCHI 2005, Canberra, Australia, pp. 21 – 25

[4] E. Chang and C. Ma, Implicit speech recognition: making speech a first class object on computers. In Proceedings of the 2nd International Conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc., 2002, pp. 349-352

[5] E. Reategui, E. Boff, and J. A. Campbell, Personalization in an interactive learning environment through a virtual character, *Computers & Education, 51,* 2 (2008), pp. 530-544.

[6] H. Walker, Collaborative learning: A working paper. March 19, 1997. URL: http://www.math.grin.edu/~walker/coll-learning/index.html, last accessed date: 27-jan-11

[7] I. P. Goldstein and M. L. Miller, AI based Personal Learning Environments: Directions for Long Term Research, MIT, AI Laboratory, 1976

[8] J. Anttonen and V. Surakka, Emotions and Heart Rate while Sitting on a Chair, Proceedings of CHI 2005, ACM Press (2005), pp. 491-499

[9] J. Du, Q. Zheng, H. Li, and W. Yuan, 2005, The Research of Mining Association Rules Between Personality and Behavior of Learner under Web-Based Learning Environment, ICWL, pp. 406 – 417

[10] J. P. C. Butz, S. Hua, and B. R. Maguire, A web-based bayesian intelligent tutoring system for computer programming. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 2004, pp. 77-97

[11] M. B. Ammar, A. M. Alimi, M. Neji, and G. Gouardères, Agent-based Collaborative Affective E-learning System. Proceedings of the First International Conference on Immersive Telecommunications, 2007, Session: Immersive system design, Article no 2

[12] M. L. Knapp, Nonverbal Communication in Human Interaction (2nd Edition), Holt, Rinehart and Winston Inc., New York, 1978

[13] M. R. Lepper, and R. W. Chabay, (1988), Socializing the Intelligent Tutor: Bringing Empathy to Computer Tutors, in learning issues for intelligent tutoring systems, Springer Verlag, New York, pp. 242 – 257

[14] N. G. Aghaee, S. Fatahi, and T. I. Ören, Agents with Personality and Emotional Filters for an E-learning Environment, Proceedings of the Spring simulation multiconference, 2008, Session: 2008 poster track, article no:5

[15] N. Roger, A framework for affective intelligent tutoring systems, Proceedings of ITHET2006 (2006)

[16] R. Dagada, Where have all the Trainers Gone? E-learning Strategies and Tools in the Corporate Training Environment, Proceedings of SAICSIT 2004, pp. 194 – 203

[17] R. Kaiser and K. Oertel, Emotions in HCI – An Affective E-Learning System. Proceedings of the HCSNet workshop on Use of vision in human-computer interaction - Volume 56, 2006, pp. 105 – 106

[18] R. McArthur, Uncovering deep user context from blogs, Proceedings of the second workshop on Analytics for noisy unstructured text data, 2008, pp. 47-54

[19] R. W. Picard, Affective Computing. The MIT Press, (1997), Cambridge, Massachusetts

[20] S. Chaffar and C. Frasson, Using an Emotional Intelligent Agent to Improve the Learner's Performance, In Social and emotional intelligence in learning Environments workshop, 7th International Conference on Intelligent Tutoring System, Brazil, 2004, pp. 37-43

[21] S. D"Mello, A. Graesser and R. W. Picard, Toward an Affect- Sensitive AutoTutor. IEEE Intelligent Systems, volume 22 issue 44, 2007, pp. 53 - 61

[22] S. Kerka, Applying Adult Learning Theory: Self-directed Learningand Transformational Learning in the Classroom, ResearchDigest no. 3, 2005, url: http://www.calpro-online.org/pubs.asp, Last accessed date: 01 – Feb – 2011

[23] S. Klein and V. Zimmermann, Open Learning Environments for Personalised Learning, Learning Technologies conference 2009, case studies, url: http://www.learningtechnologies.com.au/index.cfm?action=papers, Last accessed date: 01 – Feb – 2011

[24] T. Christner, 2003. A classroom of one (book). Library Journal, 128(1): 130

[25] U. Wolz, G. Karlsson, J. Palme et al, Computer-mediated communication in collaborative educational settings, The supplemental proceedings of the conference on Integrating technology into computer science education: working group reports and supplemental proceedings, pp. 51 – 68

[26] X. Mao and Z. Li, Implementing Emotion-Based User-Aware E-Learning, Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, 2009, pp. 3787-3792

[27] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 31, issue 1, 2009, pp. 39 - 58

E-book Reader and the Necessity of Divergence from the Legacy of Paper Book

Alma Leora Culén
University of Oslo
Institute for Informatics
Oslo, Norway
almira@ifi.uio.no

Andrea Alessandro Gasparini
University of Oslo
University library
Oslo, Norway
a.a.gasparini@ub.uio.no

***Abstract*— The traditional paper-based book as we know it has been read and loved for hundreds of years. The first generation of e-book readers, based on electronic ink, attempted to emulate the traditional paper-based book, both in content and functionality. In this paper we discuss how the reading of e-books starts to diverge from the legacy of paper books. Many still favor the sensory experience of touching an actual book, smelling the paper, and hearing the sound of turning pages. However, huge memory, easy readability, the possibility of sharing text between users (such as Nook allows), the multi-functionality (such as iPad has) and the ease of downloading books, all make e-book readers increasingly popular. Based on insights gained from ethnographic studies, workshops, focus groups, and questionnaires, we discuss some advantages and disadvantages of e-book readers, speculate on how they may change the way we read, in particular actively, and propose some guidelines for future e-book reader functionality design specifically aimed at the student population. The first prototype resulting from co-design work with primary school students is presented as an example of our design approach.**

Keywords- digital libraries; e-book readers; HCI; interface; co-design

I. INTRODUCTION

The idea of an electronic equivalent of a book is not new. Since the Gutenberg project began in the early seventies (by M. Hart as a volunteer-based effort to digitize and archive cultural works, as well as encourage the creation and distribution of e-books [8]), there have been many attempts to design a device from which one could read, store and access books easily.

Developers have been working for over three decades on e-book readers based on LCD (Liquid Crystal Display) technology without success in the market. One of the major difficulties was not only that the visibility of the text in bright light was significantly diminished, but also that the rendering of pictures and drawings was not good enough. In addition, since e-books based on this technology had a tendency to use a lot of power, the life of the battery was short.

With the invention of electronic paper (see [16] for the history and technology behind e-paper), several companies (Sony and Amazon among others) have used this material to create different types of reading devices. The technology behind e-paper is e(lectronic)-ink. A short summary of how it works is given in the Appendix A. These technologies overcome all major difficulties LCD technology had. Market interest in e-readers is soaring. A natural question that arises is how would these devices influence the act and the experience of reading.

Reading activity itself, whether using paper or e-paper, falls into two major categories of active and passive reading. Active reading is often related to work, research and studies. Passive reading often occurs in the context of leisure. Active reading is frequently associated with other kinds of activities, such as writing, cross-referencing, annotating, self-interrupting, rereading etc. (see [12,17]). Active readers often have a need to search through the text, in order to locate specific words or phrases [20,21]. Generally, this type of reading requires analyzing and structuring large amounts of information from various text sources [17].

TABLE I. READING SITUATIONS

Single text Passive	Enjoying a novel, reading a poem aloud
Single text Active	Studying a textbook or manual, reviewing a proposal
Multiple text Passive	Keeping up with e-mail, browsing the newspaper, surfing the Web
Multiple text Active	Surveying a field, researching a problem, keeping up-to-date professionally

The World Wide Web and hypertext, in particular, have changed the way we read. New technological advances continue to provide more ways of presenting information that allow for a richer experience. Many sources confirm that, if properly designed, multimedia rich presentations not only promote learning (see for example [11,13]), but also have affected the reading of books. For example, company Vook [24] has a product that integrates videos with text (video+book=vook). Potential for classroom use of multimedia applications and innovative interfaces for reading has not gone unnoticed. Recently, in [18], Sharma et al. report efforts to develop software and hardware in order to

 ISBN: 978-1-61208-117-5

turn small classroom PCs into reading and writing devices for children 5-11. This project, even though it is concerned with young students only, has goals similar to what is presented here.

The present article does not discuss technical issues with e-book readers to any significant degree, but rather the use of books or e-book readers as well as the experience of the act of reading using paper based media (mainly books and news-papers) vs. e-book readers (most of which offer the possibility or reading the news online as well). We are interested in a range of topics related to e-book readers: the ways in which they may influence our reading habits, levels of understanding, and how they may be adapted to better suite active reading for specific groups of users such as, for example, students.

Ethnographic studies (observations and interviews) of how people read in different contexts have been carried out. Observations were carried out at the university library, where the majority of users are active readers, as well as public spaces such as bookstores, cafés, busses and trains.

The university library has obtained some e-book readers early on and the opportunity, as one of the authors works at the library, presented itself to observe and interview the users of the library who were willing to try e-book readers. The users of e-book readers were also handed a short questionnaire (see Appendix B) and asked to fill it out. These preliminary studies pointed out in the direction that functionality of the e-book reader needed to be improved in order to better accommodate the needs of an active reader. Additionally, e-book readers were taken to the classroom in the primary school (third grade), to test the device with younger readers. The result there was similar, devices need to be better tailored for the needs of young students.

Focus groups and workshops conducted after the preliminary studies had getting to understand users needs as a goal as well as to gain better understanding of what users do during a typical day at school or university. Mixing of ideas (see [7]) and brainstorming about the future use of the devices was conducted with young users aged 7-8 years. We have attempted to create an interface co-design situation (see [2,3,5,23]). The first prototypes and experiences from these sessions will be presented here. Finally, all our findings are summarized in a future oriented scenario and a set of guidelines for functionality design of e-book readers for students.

The structure of the article is as follows: a short overview of what an e-book reader is, technologies used for e-book readers and a few e-book reading devices are briefly described in Appendix A. Section 1 contains some of our observations, as well as some other research results related to the activity of reading paper based media (books and newspapers).

Section 2 contains the results of observations, interviews and questionnaires we conducted on the use of e-book readers. Section 3 discusses the differences between the affordances two artifacts provide and how those may affect the reading habits in a long run. Section 4 summarizes our findings from workshops and focus groups in form of a future scenario and design guidelines.

II. Reading Printed Media

Long sequences of innovations and technological advances have transformed writing from initial carvings on stone, (and later papyrus) to a medium printed on paper. Books themselves have directly affected the development of the society as a significant and primary tool of the mind. As noted in [9], "*Reading itself was for many centuries mainly conceived of as a ruminatio (pondering at length). In the eighteenth century, a new relationship to reading appeared, making it legitimate to browse through vast amounts of reading material. This form of "extensive reading" became pervasive with the advent of newspapers and magazines.*" Today, a new relationship to reading is emerging through the extensive use of digitalized material. People read, world around, both traditional paper and digital material. To better understand the dynamics of interaction between the reader and artifact (whether book or e-book reader), we have conducted ethnographic studies to observe how people read books and other printed material as both active and passive readers.

Short interviews with people reading in libraries, bookstores and on public transportation often yielded unexpected reasons for people's interest in paper books. For example:

A 15-year-old girl sitting in a café had the "New Moon" book on the table in front of her. As she started to collect her things, we asked whether she had read the book. She said that she had finished and was now rereading it, but more importantly, that she always brought it with her as doing so had started several interesting conversation with people seeing the book and making a comment. "Kind of like you did now", she noted. She said that she had read all the other books in the series as well and is an active member of several Facebook fan groups related to the series.

In another instance, a 49-year-old woman was walking around, browsing in a small bookstore. After she picked up a large, thick book, she smelled it. We asked why she had done so. The women said that she does not really know why but that the smell of the book affected her interest in it. For example, she did not like the ones that smelled moldy which she associated with dark cellars. If by contrast the book had "good smell", she would open it and look through. She also said that she did not do the same with new books. In addition to smell, the title and cover art play a large role in attracting her to the product. She said that she liked this particular bookstore since she could get used books really inexpensively. We asked if she read all of the books she bought. "Not always", she said, "but I like having them anyway. I think that some day I will have more time to read."

A 21-year-old male student said that he loves poetry and poetry books because of the poems in them, but also because he may put a love letter or his own poem into them.

Another 19-year-old female student liked the evidence of the time passage linked to events from her life, being visible on the books. For example, she enjoys taking (time and

ISBN: 978-1-61208-117-5

again) a certain book from the shelf to see the mark of a spilled drink left during a first date.

As examples above illustrate, our general observations found that people have not only varied reading habits but also unexpected factors attached to their love of books.

The following summarizes what was important about paper books to readers we observed and interviewed:

- The smell of the book, in particular in the library and at used bookstores.
- The sense of touch, the feel of the book as one picks it up.
- The quality of the paper. As one subject described, “Touching a glossy page of a book from a photographer, makes me more respectful than reading a novel in a pocket format.”
- The size of font and spacing between the lines.
- The weight of the book.
- When the user keeps the book in his/her hand while reading, the grip of the book provides a sense of how far along one has come in finishing the book.
- Cover art (its content, style, quality, texture etc., see also [10]).
- Quickly finding a specific location in the book (through the index, for example).
- Page indexing using post-it notes or other markers give extra functionality to the paper book.
- The ease and speed with which one can move back and forth between pages (see also [22]).
- Notes and underlining done by others in used books is important (and can be experienced as either frustrating or interesting/helpful).
- The amount of wear of the book can affect the choosing of it (see also [5]).
- Sharing a book is important to many readers (not only the actual physical copy but also in discussing the content with others).
- The social aspect of holding and reading a book may be used to communicate a certain message about the reader. “Reading a medicine text book on the bus gives another signal than if I am reading comics,” remarked one reader.
- Variety of ways of using books (sometimes not related to reading, such as above mentioned poems holder)

As with other objects acquired in our daily lives, books are chosen emotionally [14]. Both the reading and sensorial experience of the book itself affect our relationship with it. The question was: what happens to all these sensorial and emotional factors related to paper books when they are digitalized?

III. READING THE E-BOOK

In the beginning of September 2009, the library at the University offered to its users the option of checking out e-book readers (SONY PRS 505 and Iliad iRex) thereby previewing the new technology. The interest among library users for checking out the e-book readers was large. Over a 6-month period, the library lent out e-book readers 120 times with users initially waiting more than a week to get one. Because of the waiting list, students were allowed to use the e-book reader only once. That is to say that there were no repeat users in the sample.

A questionnaire (see appendix B) accompanied each e-book reader, 41 of which were returned completed. Below are findings from the questionnaire:

All respondents (41 users) were interested in reading the university curriculum on the e-book reader.

Approximately half of respondents (21 users) preferred text in PDF format, finding it easier to read.

All respondents (41 users) desired a better framework for downloading e-books from the library into the reader. Six users found the Sony software for communicating between the e-book reader and their computer was difficult to use.

Eight users pointed out that turning the pages on the device was annoying because it took too long.

Six respondents noted wanting to be able to annotate and mark the text on the e-book reader. One person said s/he would want a dictionary included as part of e-book reader software.

- From direct observations of people using e-book readers and short conversations we compiled the list of things they wish for in a new generation of e-book readers. This list is consistent with the findings of earlier studies [11,12,13,17,22]: Color images and text
- Information about original pagination
- Tables of content for each document
- Fixed/standardized layout and page orientation
- Providing a structure to keep similar texts together
- Portability and comfort in use from different positions (sitting, lying down)
- Focus on reading comfort (with clear and large text and the ability to adjust to different lighting i.e. outdoor vs. indoor, daylight vs. nighttime)
- Search capacity (through both the text in use and those available on the device)
- Annotation tools
- Highlighting capacity
- Simplified navigation between documents
- Capacity to interact between different information platform media
- Tools to facilitate collaboration

Many of these items are already supported by the visual and user-friendly format of iPad.

A comment from one participant was interesting: "Iliad shows signs of being designed by engineers who are a little over zealous; it has many functions that work halfway". It in a way summarizes our conclusion: this technology has a far-reaching capacity that, in these first generations at least, has been suggested but not fulfilled. However, e-book readers or similar devices are here to stay. And like it or not, the e-book is not only made for reading on a digital device, but also shapes the reading itself. How the artifacts and actual skill of reading influence and inform one another over time and with different groups of readers is an interesting question to be answered in the future.

 ISBN: 978-1-61208-117-5

IV. Diverging from the Legacy of Paper Book

It seems from Sections 1 and 2 that books would have a much longer list of "things I love about books" than would e-book readers. However, people are starting to love the new technology as well (see [10]). It is clear that browsing, searching, and powerful annotating (tagging, indexing etc.) provide a clear-cut advantage to e-book readers, at least in active reading and learning situations. To the extent that this advantage is true even within the present technology, it promises to further grow with future advances.

The shift that is about to take place in the e-book reader technology involves its departure from the legacy of the paper book. Admittedly one cannot expect to fold poems within the plastic e-book casing. Nor will one ever be able to experience the level of tangibility offered by books with their weight, size, smell, and other physical properties. While turning pages on an e-book reader may have a sound, most would agree that it is not the same as actually turning the page of a book. So at this stage where e-book readers have in place all of the basic reading functions (inherited from books), they must necessarily depart from these basics and consider the additional affordances of this new digital medium. The e-book reader has the potential to include multimedia presentations, cooperative environments, creative environments related to reading and writing (such as for example interactive writing, hyper fiction, or digital storytelling). Fully personalized, as each owner of an e-book reader will have own selection of books. And technology is available to make all the wishes from the list in section 2 true, and more. Exploring the possibilities is ahead!

As mentioned earlier, with each new technological advance and each difference in the medium from which we read, there are also qualitative changes in the act or reading itself. We have shifted from pondering to hyperlink reading. And writing in the same style. Humorist Carr jokes about his inability to concentrate on a single piece of text any longer, and his lost ability to write. On the other hand, [1] reports that working with e-book readers seem to deepen concentration of the user on the text at hand. Through our observation and interviews, we cannot say anything conclusive about quality of concentration; it would require a much more focused study, which is outside the scope of this research. Whether one is true or the other, it seems that the question about the influence of the new device on the reading quality is in place.

V. Focus Groups and Workshop Findings

Two workshops were conducted in September and October of 2009, and two focus groups, in February 2010.

A. First Focus Group

The first focus group was comprised of four randomly selected University students (between the ages of 19 and 32) who were active readers interested in studying the texts they were reading. The main purpose of the group was to hear what students might need or want to see on a device such as an e-book reader.

Findings support that students want something more than the traditional paper version can offer, especially when they are actively reading in order to learn, for example, an academic curriculum. They are willing to sacrifice some of the benefits traditional paper books have in return for the possibility of searching through all of the texts they may have on the reader. As one of the member of our focus group reflected: "It may be disturbing to read texts with links to searchable phrases or external information and be tempted to use them unnecessary, but the gain is BIG"

Results from discussion in the focus group supported a content oriented approach. That is to say within the context of studying and research, the availability of information (the content) is the primary goal. E–book readers have the advantage over paper books of being able to access, annotate, and compile content within a single source – a finding supported by another study conducted at Princeton University [15].

One important advantage of e-book readers is the immediate availability of information. A participant of the focus group noted the advantage of owning one's own mobile library: "I can move all my books to this device, since I perceive it as a private dings, like the mobile phone. And have them all whenever I need them."

This finding is also supported by other studies [12,17,15,21]. Worth mentioning here is finding that users bring the e-book reader everywhere and are not afraid of damage or loss as with more expensive devices such as a laptop.

Some participants in the focus group pointed out that they had already tested and used an e-book reader alongside a laptop (like a paper book) when doing homework.

While our users did not discuss the quality of concentration when using the e-book reader, a study, done at Penn State University [1] analyzed student behavior while reading and using e-book readers and found a more immersed type of reading with the latter.

B. Second focus group

The second focus group was conducted in a business setting rather than academic setting. Most participants in this group were passive readers, not interested in reading books with learning as the primary goal. Instead, other functions such as being able to listen to music, take notes, and watch movies were more valued.

A participant, when told what one could do with an *e-book reader*, said: "So an e-book-reader is like a modern Gameboy to bring on boring trips!"

Another participant said he would consider purchasing an e-book reader if he could easily download the daily newspaper in the morning, the text was easy to read, and the user interface had good navigational buttons.

Another explained his appreciation for pocket sized books, and wished for an e-book reader one could put in a jacket-pocket with the navigational buttons on the back. Regarding the interface, he said: "As long as the text is 12-point and Arial or New Times Roman-style, I don't really need fancy apps and programs."

 ISBN: 978-1-61208-117-5

A preliminary conclusion from this second focus group of passive readers was that while interested in the e-book reader, their interest was focused on non-content areas such as platform and the readers use in providing news, entertainment, and pleasure reading in contrast to students who are vitally interested in the content and for whom the platform is less important.

C. *Workshop*

In addition to the two focus groups, two workshops were conducted at the third grade level (ages 7-8) of the primary school.

During the first session we had a period of observation of how children would use e-book readers (see Figure 1).

Figure 1. Third grade students trying Iliad e-book reader**.**

The young participants reported a number of problems. They found reading books in PDF format on the e-reader boring, in part because of the content, in part because of problems navigating. Children were frustrated that they could not simply “jump out” of the text and try another. They were comparatively a lot more interested in our supposedly inconspicuous iPhones used to record the session, than at the e-book reader itself.

However, as soon as they were handed paper and color pencils, their interest in drawing their ideas for an interface that would be interesting for them increased dramatically. They spent the remainder of the first workshop vigorously drawing. Students were divided into four groups, each of which was instructed to work on an idea, pre-selected by the children together with an adult mediator. This was an attempt to emulate the mixing of ideas process as discussed in [7,11]. An example of one such idea is shown in Figure 2. Several simple prototypes, one of which is shown in Figure 3, were created and we worked on these together with the same group of children during the second workshop.

Our main findings from this workshop were that working with children in co-design situation is very complex and requires much better preparation and much more time than with adult users. Also, we worked with a too large group of children. At that age the children cannot have long concentration spam and they influence each other strongly.

Children were very articulate about what they like and dislike, while working creatively with an idea took longer time to get into. Once going, their fantasy was very inspiring.

Figure 2. Child’s drawing of a more visually based interface navigated by clicking on images or text

If the project continues, co-design with children would be the method of choice.

From what we have seen and heard, active readers have higher demands and a longer wish list for good support software on their e-book readers then passive readers do.

However, any item from the wish list changes the e-book reader it into something else, more complex. For example, one “simple” wish for collaboration on projects involving e-book readers could require e-book reader to become an e-writer as well, and ultimately, another communication tool.

Figure 3. Prototype created on the tablet PC.

D. *Informing the future design of e-book readers for active readers*

Does it make a difference whether one starts from a small PC and makes it into an e-book reader by using appropriate technology for the ease and comfort of reading, such as in [18], or whether one starts from an e-book reader and builds it into a device that also writes, searches, browses, collaborates, perhaps writes hyper-fiction or tells stories? We believe that the convergence will take place. iPad, now widely available, is one possible answer to that question.

From the fall 2010, University Library will test iPads. And iPad is the platform of choice for us to continue working with. This choice is personal; there may be an e-book reader of the next generation that will surpass the iPad. But whatever the platform, there are some general guidelines we will follow in our effort to design e-book reader interfaces better suited for active readers:

- Working with narrower groups of users (for example, active readers, age 7-9, as the use of and

 ISBN: 978-1-61208-117-5

the needs for software change rather quickly at that age)

- Co-designing with users
- Allowing the possibility for end users to participate in the design process

VI. CONCLUSION

Extensive empirical studies have been done on the actions related to reading printed media versus e-book readers. E-book readers at first emulated functionality of the book, but now we see divergence from the legacy of the book. Focusing on active reading, we take a closer look at the design of the future e-book reading devices. In doing so, we have worked with young users in a co-design situation and created the first interface prototype for e-book reader. We have conducted qualitative and quantitative studies with university students who volunteered to try e-book readers. The conclusion of the study is a list of feature users would like to have on e-book readers as well as the set of guidelines continuing towards design of interfaces, better suited for specific user groups.

APPENDIX A

The technology behind e-paper is based on the e(lectronic)-ink developed by Vizplex. Use of e-ink improved the quality of reading experience significantly. A very short summary (see [21] for more details) of how e-paper works: e-ink is like traditional ink in that it is a colored liquid that can be coated onto nearly any surface. Suspended in the liquid are millions of microcapsules that contain tiny, two-toned (dark and light) polymeric particles. Similar to a magnet, the two tones have opposing electric charge. When the e-ink is exposed to an electric field, the particles realign themselves and form crisp text and images. The e-paper uses power only when a page is refreshed or changed ([19]).

All e-book readers consist of an enclosure made of plastic or metal, with buttons to navigate through the documents and to activate the menu. They may have support for USB, memory cards, SD cards, WiFi and GSM communication, touchscreen-interface, audio and have their own web browser. They are usually the size of A4 or A5.

There are many models of e-book readers on the market today. A great summary of devices for reading and the text formats they support is available at Wikipedia e-book readers comparison site [25]. We mention here only the ones used by the library in our study, SONY and Iliad, as well as most fashionable ones today, Kindle and iPad.

Iliad (see Figure 4) is an e-book in A5 format. It came out in mid 2006, has a 8.1 inch e-paper display, and it weighs approximately 0.85 lb. It uses WiFi to connect, and operates on a Linux system. Software developer kit follows with it, allowing users to extend its functionality. One of the more advanced features it has, is the ability to make notes on most of the documents. In 2007 iRex released Version 2, which has some minor updates, and a Book edition in 2008, which has eliminated WiFi in order to make the reader less expensive.

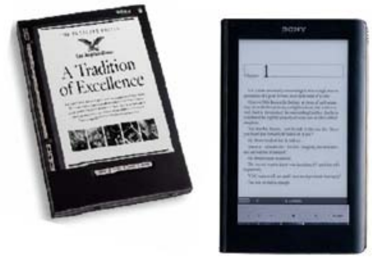


Figure 4. E-book readers Iliad and Sony, used in our study.

Sony has already produced six models of e-book readers. The first one came out in the USA in 2006. The new series of e-book readers, available from 2009, come in 3 different sizes to fit some specific uses. A pocket edition with a 5" screen and navigational buttons, a touch edition (see Figure 4) with a 6" touch screen and a memory card slot and last the daily edition with a 7" touch screen, weighting from 0.5 lb to 0.75 lb. All of the readers have wireless connection through 3G mobile networks.

Figure 5. iPad, iBook reader application.

iPad (see Figure 5) is the newest Apple multitouch tablet, that comes with nearly all the features iPhone has: pretty design in crisp color, a large, 9.7 inches high-resolution LED-backlit IPS display, a Multi-Touch screen and a powerful chip. It weights 1.5 lb and is 0.5 inches thick.

It is expected to set new standards in functionality and aesthetics. In addition, it comes with iBook application, from which books can be easily downloaded either using 3G or wireless network. Apple claims that the iPad's battery can provide up to 10 hours of video, 140 hours of audio playback, or up to one month on standby [26].

Amazon has released a series of e-book readers. The original Kindle was released in 2007and the newest Kindle DX was released in May 2009 (see Figure 6). The DX has a 9.7" screen, a function to rotate the picture according to the way the device is held, improved storage capacity and battery-life.

Figure 6. Amazon's Kindle DX.

The interface is a black and white screen. However, the most interesting distinction is that this 3G version has possibility of access Kindle library from other devices such as iPhone, Blackburry, PC, etc.

VII. APPENDIX B

Questionnaire handed out with e-book readers

We would like to know what you thought about the e-book reader you borrowed from the library. Your input is very valuable to us. It will take only a few minutes to answer the questions bellow. Please return the completed questionnaire to checkout desk at the library at your convenience.

Thank you for your input.

1) Which e-book reader did you use? Sony Iliad
2) What made you want to try an e-book reader?
3) What did you think of the e-book reader? Please comment on the functionality it has
4) How did e-book reader function with the resources for your classes?
5) What do you think about the library lending out e-book readers?
6) Would you care to lend it again?

REFERENCES

[1] Behler, A. (2009) "Readers in Action", available at: http://www.americanlibrariesmagazine.org/features/09242009/e-readers-action (accessed February 3, 2011)

[2] Culen, A,: "Designing for Youngest", conference 'Designing for Children', IIT, Bombay, February 2-6, 2010

[3] Culen, A., Herstad, J. and Karahasanovic, A.: "Collocated collaborative storytelling for children: what does the technology bring in", conference "Designing for Children", IIT, Bombay, February 2-6, 2010

[4] Dieberger, A. and Dourish, P. (2000) "Social navigation: Techniques for building more usable systems", Interactions, 7(6), pp. 36-45

[5] Druin, A., Bederson, B., Hourcade, J.P., Sherman, L., Revelle, G., Platner, M., and Weng, S. (2001). "Designing a Digital Library for Young Children: An Intergenerational Partnership", Joint Conference on Digital Libraries (JCDL'01)

[6] Grenier, L. (2009) "unBook" netWorker, Volume 13 , Issue 2 A Digital Trail is Forever, pp. 8-10

[7] Guha, M. L., Druin, A., Chipman, G., Fails, J., Simms, S., and Farber, A. (2004). "Mixing Ideas: A new technique for working with young children as design partners". In Proceedings of Interaction Design and Children (IDC'2004). College Park, MD, pp. 35-42

[8] Gutenberg project, available at: http://www.gutenberg.org/wiki/Gutenberg:About (accessed February 3, 2011)

[9] HCI –book Strategic Research Clust*er:* www.hci-book.org/cluster/uses.php (accessed February 3, 2011)

[10] Kragler, S. and Nolley, C. (1996). "Student Choices: Book Selection Strategies of Fourth Graders". Reading Horizons, 36 (4), pp. 354-365

[11] Mayer, R. E. and Moreno, R. (1998). "A Cognitive Theory of Multimedia Learning: Implications for Design Principles" http://www.unm.edu/~moreno/PDFS/chi.pdf (accessed February 3, 2011)

[12] Moore, L. M. (2009) "At your leisure: Assessing eBook reader functionality and interactivity", Master thesis, University College, London, UK.

[13] Muller, D. A., Lee, K. J., and Sharma, M. D. (2008). "Coherence or interest: Which is most important in online multimedia learning?". Australasian Journal of Educational Technology 24 (2), pp. 211-221. http://www.ascilite.org.au/ajet/ajet24/muller.pdf. (accessed February 3, 2011)

[14] Norman, D. (2004) "Emotional Design", New York: Basic Books.

[15] Princeton University. "The E-reader pilot at Princeton", available at: http://www.princeton.edu/ereaderpilot/index.xml (accessed February 3, 2011)

[16] Romanowski, P. "Electronic Ink", available at: http://www.madehow.com/Volume-6/Electronic-Ink.html (accessed February 3, 2011)

[17] Schilit, B.N., Price, M.N., Golovchinsky, G., Tanaka, K. and Marshall, C.C. (1999) "As we may read: the reading appliance revolution", *IEEE Computer*, 32, 1(January), pp. 65-73

[18] Shama, M. and Abraham, T.: "Case for Reading and Writing on a Classmate PC", in conference 'Designing for Children', IIT, Bombay, February 2-6, 2010

[19] Shapira, B.; P. Shoval; J. Meyer; N. Tractinsky, and D. Mimran (2008) "Epaper the personalized mobile newspaper", in: *WorkShop Proceedings of the 2008 International Conference on Intelligent User Interfaces*, pp. 1-12.

[20] Springer, *eBooks the end user perspective*. White paper, available at: http://www.springer.com/cda/content/document/cda_downloaddocument/eBooks+-+the+End+User+Experience?SGWID=0-0-45-608298-0 (accessed February 3, 2011)

[21] van der Velde, W. and Ernst, O. (2009) "The future of eBooks? Will print disappear? An end-user perspective", *Library Hi Tech*, 27,4 , pp. 570 – 583

[22] Tajika, T.; Yonezawa, T. and Mitsunaga, N. (2008) "Intuitive page-turning interface of e-books on flexible e-paper based on user studies", in: *Proceeding of the 16thACM International Conference on Multimedia*, October 26-31, 2008, Vancouver, British Columbia, Canada, pp. 1647-1656

[23] Taxen, G., Druin, A., Fast, C., and Kjellin, M (2001). KidStory: "A Technology Design Partnership with Children". *Behaviour and Information Technology*, 20(2), pp. 119-125

[24] vook, the smart book, avaible at: http://www.vook.com (accessed February 3, 2011)

[25] Wikipedia, Comparison of eBook readers, available at: http://en.wikipedia.org/wiki/Comparison_of_e-book_readers, (accessed February 3, 2011)

[26] Wikipedia, iPad, available at: http://en.wikipedia.org/wiki/Ipad (accessed February 3, 2011)

 ISBN: 978-1-61208-117-5

Exploring Trust in Personal Learning Environments

Na Li, Maryam Najafian-Razavi, Denis Gillet
Ecole Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne, Switzerland
{na.li, maryam.najafian-razavi, denis.gillet}@epfl.ch

***Abstract*—The design of effective trust and reputation mechanisms for personal learning environments (PLEs) is believed to be a promising research direction. In this paper, we propose a 4-dimensional trust model that complies with the specific requirements of PLEs. Trust is explored in four dimensions: trustor, trustee, context and visibility. The importance of these four dimensions is investigated through a number of scenarios. The model is implemented in a PLE platform named Graaasp. Preliminary evaluation of usefulness is conducted through a user study and some interesting findings are discussed in the end.**

Keywords-trust; reputation; personal learning environment; rating; ranking

I. INTRODUCTION

Benefiting from the success of Web 2.0 social media, interactive information sharing has become pervasive. For users surrounded by an abundance of information, the challenge now is how to determine which resources can be relied upon and who is reliable enough to interact with. To solve this problem, a number of trust and reputation systems have been developed in various platforms, including e-commercial sites, product review systems, and professional communities. Trust and reputation measures can help users decide whether or not to interact with a given party in the future, or whether it is safe to depend on a given resource [1]. This creates an incentive for good behavior, therefore inducing a positive effect on the quality of interaction in online communities.

As a particular support framework for interaction in online communities, personal learning environments (PLEs) embed tools, services, content and people involved in the digital part of the learning process [2] [3]. Web 2.0 functionalities like blogging, tagging, rating and commenting are gradually incorporated into learners' overall learning ecology, contributing to increasing learning incentives and enhancing the learning experience [4]. On one hand, these Web 2.0 features enable learners to express opinions easily and facilitate accumulating domain knowledge. On the other hand, learners' active contributions produce a large amount of user-generated content, which may lead to information overflow. In such an open learning environment, it is not easy for learners to find suitable people to learn from or collaborate with. Moreover, the flood of data might bring about the challenge of selecting useful learning resources depending on personal learning goals. Therefore, research efforts are needed to design appropriate trust and reputation mechanisms for PLEs, aiming at expertise assessment and quality assurance to support self-directed learning activities.

As a complex social concept, trust reflects the subjective perception one party holds about another party. It's asymmetrical, transferable, dynamic, and context-dependent, and can be influenced by various factors. Trust has been interpreted in different ways to comply with specific requirements of various domains. In this paper, a usable trust model for personal learning environments is investigated and implemented in a PLE prototype named *Graaap* [1]. Preliminary evaluation of usefulness has been conducted through a user study. The rest of this paper is organized as follows. Section II introduces the existing trust and reputation schemes, and discusses the specific aspects of PLEs that call for different trust and reputation models. A trust model dedicated to PLEs is proposed in Section III. Section IV introduces the *Graaasp* prototype and illustrates the implementation of the proposed trust model in it. User evaluation and main findings are addressed afterwards in Section V. Finally, Section VI concludes the paper and discusses the future work.

II. RELATED WORK

In order to develop effective trust and reputation mechanisms, a number of attempts have been made in both literature and practice. The simplest but most widely used scheme is to compute an average or summary of all ratings for an entity. The reputation systems used by eBay [2], Epinions[3] and Amazon[4] fall into this category. However, this scheme is primitive and therefore gives a poor picture on an entity's reputation score [5].

Google's PageRank [6], Advogato's reputation scheme [7], and EigenTrust model [8] can be categorized as flow models, where trust or reputation is computed by transitive iteration through looped or arbitrarily long chains. In short, a participant's trust or reputation score increases as a function of incoming flow, and decreases as a function of outgoing flow. Flow model-based schemes adopt global trust metrics, where a single trust or reputation score is associated with each participant and displayed to all members in the community.

Some researchers have proposed to use mathematical models in an attempt to measure trust, including Bayesian algorithm-based metrics [9] and belief theory-based models

---

[1] Graaasp (graaasp.epfl.ch): a PLE prototype.
[2] eBay (www.ebay.com): an online auction and shopping website.
[3] Epinions (www.epinions.com): a consumer review website
[4] Amazon (www.amazon.com): an e-commercial website.

 ISBN: 978-1-61208-117-5

[10]. These models provide a theoretically sound basis for computing reputation scores, but they are too complex for average users to understand and too difficult to implement in real systems.

There are also some approaches using rating similarity to predict trust. Golbeck [11], for example, proposes an approach for computing trust in which several aspects of rating similarity that are believed to affect trust opinion in movies are integrated to compute trust values between people. The approach is based on the observation that, in many Web-based social networks, a significant percentage of users are completely isolated from most others, thus the social network-based trust measures are not applicable. However, those isolated users are usually the inactive ones who rarely contribute to the system and hence provide few ratings. Lack of ratings would probably be a problem for trust metrics relying on rating similarity.

A general observation of existing trust and reputation systems is that they tend to adopt global trust metrics, where for each entity in the community, a reputation score is computed to reflect how much the community as a whole trusts this specific entity. However, trust is more of a personalized concept greatly depending on personal experience, which means that the trust score of an entity could be different from the point of view of different users. Furthermore, each of the current trust and reputation systems targets to a particular domain, including e-commercial, product review, and movie review. In each domain, the context of trust is a shared belief in the whole community, such as the quality of service in case of e-commercial domain, or the usability of product in case of product review domain. In comparison, a PLE is a multi-disciplinary and multi-dimensional environment where resources and services are aggregated from heterogeneous sources. The contexts of trust could be diverse depending on different learning scenarios like group project of physic course, discussion of English learning, online meeting about Web 2.0 technologies, and so on. Thus, there is a clear need for a context-dependent trust model that accounts for various learning contexts in such an environment. Last but not least, instead of designing complex mathematical model-based metrics, the trust model should be simple enough for non-technical users to use. Given these requirements, effective trust and reputation mechanisms in PLE should take users' subjective views and learning context into account, as well as making it easy for average users to understand.

III. TRUST MODEL IN PLE

Intuitively, trust can be affected by a variety of factors, including personal experience, rumor, social rules and so on. There is no single trust model that will be suitable in all domains and applications. To comply with the specific requirements and constraints of PLE, trust is explored in four dimensions: trustor, trustee, context and visibility. The idea is to describe who trusts what in which context with which degree of visibility. In this section, we propose a 4-dimensional trust model dedicated to PLEs. The importance of these four dimensions is illustrated through example scenarios.

A. *Trustor*

The two parties involved in a trust relationship are referred to as trustor and trustee respectively. As mentioned previously, for a particular trustee, the trust value varies from the standpoints of different trustors, depending on various personal experiences. As an example, Alice strongly trusts Bob since they have been working together on the same project for several months. While Claire does not trust Bob at all as they have not collaborated with each other before. In short, *personalization* is a crucial characteristic that should be accounted for when computing trust. However, with incorporation of personalization into the trust model, computation complexity increases accordingly, which might bring about confusion for users. Therefore, the personalized trust metrics must be designed and presented in a straightforward way so that average users can easily understand and use them.

Although personalized trust metrics are more precise and tailor-made for individual users, they are still not sufficient to measure an entity's trustworthiness in the learning community. For new users who do not have much experience in the community and have no knowledge of entities' trustworthiness, it is more important to refer to others' trust opinions towards an unknown entity to reduce transaction risk. To this end, the *global* reputation score of a particular entity should also be considered as an essential factor, in terms of representing the trust opinion held by the community as a whole. In this paper, we integrate global and personalized trust metrics for the purpose of providing both a general and a personalized view of trust opinions. Concrete application of the two trust mechanisms will be illustrated in detail in Section IV.

B. *Trustee*

In a trust relationship, trustor is usually a person or an agent. While as the receiver of trust, trustee can be a person, a resource, a tool or any other entities in the system. For instance, a professor can be trusted because of his/her expertise, a book can be trusted because of its quality, and a tool can be trusted because of its utility. For a specific type of trustee, people tend to use a certain set of criteria to evaluate the trustworthiness. It's relatively easier to evaluate the trustworthiness of a document than a person, since the characteristics of a person are much more complex and could change over time. How the type of trustee influences trust relationship between two entities will be investigated in the user study hereafter.

C. *Context*

PLE is an open environment that aggregates people, resources, and services from a large variety of disciplines, such as physics, computer science, electrical engineering, and any other learning domain. Depending on different learning goals, the learning activities in PLE could also vary from group projects to independent study, forum discussion, and online courses. In such a complex environment, context plays an important role in how people assign trust. As an example, Alice trusts Bob in the context of English learning since Bob is a native English speaker, whereas she does not trust him in the context of ski learning as Bob is a ski beginner. Therefore, it's essential to integrate the factor of context into the trust model of PLE.

However, as a controversial concept, context is too complicated to define and represent. Major research efforts have been made to develop ontology-based context frameworks [12] [13], but the problem is that ontologies are very difficult to agree on in that different communities use different naming standards and interpret entities in different ways. Moreover, from a user's point of view, ontology-based context models impose an inconvenient way of organizing and classifying entities. To solve this problem, a simple and understandable approach to identify context and incorporate it into the trust mechanism is introduced in Section IV.

D. Visibility

The trust scores in existing trust and reputation systems are mostly publicly available to all members of the community. For instance, all eBay users can see what feedback (positive, neutral, or negative) a particular user has given to another user. Nevertheless, in some cases, people might feel uncomfortable to share their trust opinions with others. Especially when people express negative trust opinions, they might prefer to make them private in order to avoid retaliation from others or limit the negative social impact of the trust values they assign. Furthermore, these psychological reasons have been proven to lead to a positive bias of ratings in most reputation systems [5].

Due to these privacy concerns, three visibility types of trust are considered in our model: public, private and anonymous. Public trust score is globally visible in the entire community. Private trust score assigned by a user is particularly accessible to himself/herself or a certain group of users indicated by him/her. Anonymous trust score is accessible to all members anonymously.

Based on the analysis above, a 4-dimensional trust model is proposed by integrating the four factors that are believed to effectively describe how people assign trust in PLE. Let T_r denote trustor, T_e denote trustee, T_c denote context, and T_v denote visibility. A trust value can then be defined using the term $< T_r, T_e, T_c, T_v >$. As an example, $<$ *Alice*, *Bob*, *English learning*, *public* $>$ represents the trust value Alice assigns to Bob in the context of English learning publicly. The implementation of the trust model in a PLE platform is addressed in detail in Section IV.

IV. Trust Application in Graaasp

A. Graaasp

Graaasp can be described as a Web 2.0 application that can serve simultaneously as an aggregation, contextualization, discussion, and networking platform, a shared asset repository, or an activity management system. The user interface of *Graaasp* is illustrated in Fig. 1. The structure of *Graaasp* relies on the extension of the 3A interaction model [14], 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

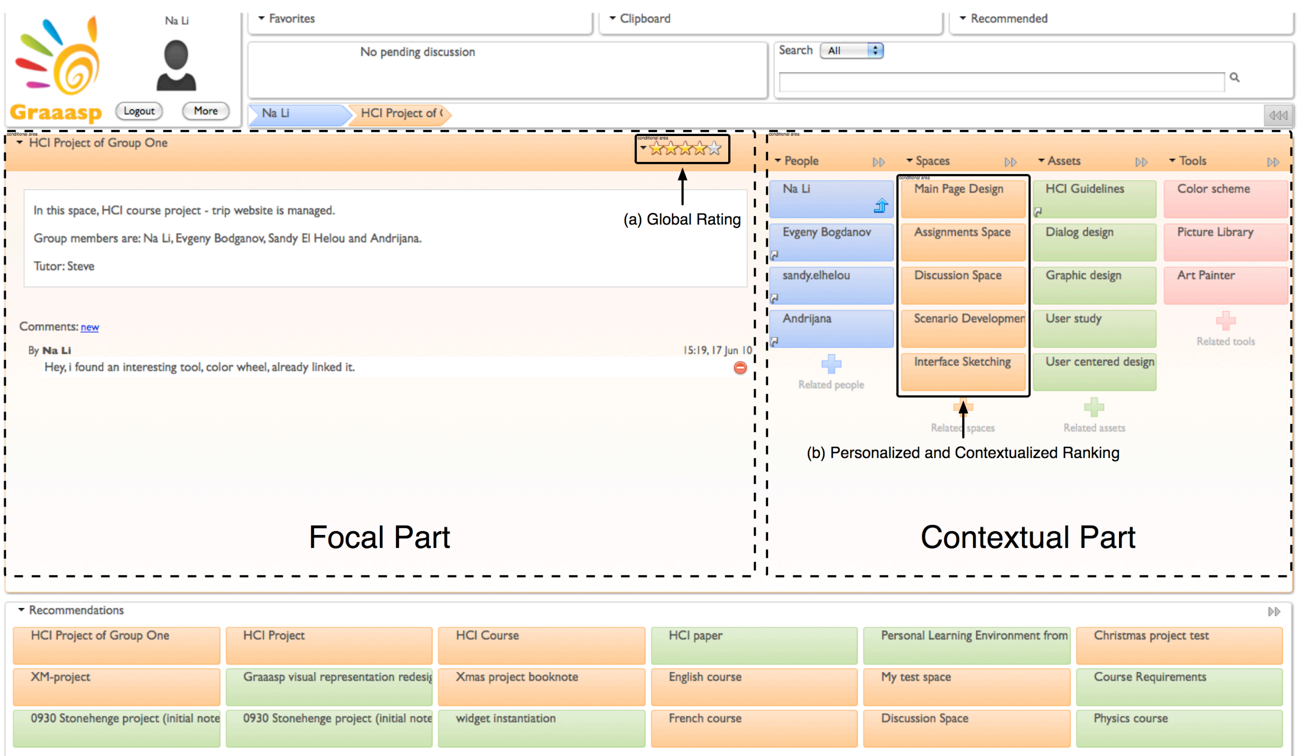


Figure 1. Graaasp User Interface

agents. Actors create collaboration spaces where they conduct personal and group **Activities** to reach specific objectives. In each of these activities, actors can take different roles, each of which consisting of a label and an associated set of rights. In addition, Actors produce, edit, share and annotate **Assets** in order to meet activities objectives. Assets can consist of simple text files, RSS feeds, wikis, videos or audio files. A fourth structural entity **Application** is added into the model to describe widget or gadget [15] that can be installed and executed within the Web pages. Applications can be any tools created or linked by actors.

As shown in Fig. 1, the user interface of *Graaasp* mainly consists of two parts: the focal part on the left side and the contextual part on the right side. The focal part shows the entity that is currently selected by the user. It can be a human (actor), an activity (space), an asset or an application (tool). The contextual part consists of the four columns of items, each of which represents one type of entities (actors, assets, activity spaces, tools) linked to the focal entity.

Instead of integrating complicated context framework, we simply define the current learning context as the combination of the focal entity and its related entities. A further explanation of the learning context is described through the following scenario: A group project called "*HCI Project of Group One*" is selected as the focal entity, within which there are four group members ("*Na Li*", "*Evgeny Bogdanov*", "*Sandy El Helou*", "*Andrijana*"), several activity spaces ("*Main Page Design*", "*Assignment Space*", "*Discuss Space*", "*Scenario Development*", and "*Interface Sketching*"), a set of assets ("*HCI Guidelines*", "*Dialog Design*", "*Graphic Design*", "*User Study*", and "*User Centered Design*") created by group members, and a number of tools ("*Color Scheme*", "*Picture Library*", and "*Art Painter*") linked by them. These entities, as a whole, construct the learning context.

B. *Applying Trust in Graaasp*

Different ways to apply and represent trust can be found in various online systems. Rating with the scale of five is the most widely used method to assign trust, which has been applied by eBay, Epinions, Amazon, and so on. Voting as "like" or "dislike" is also used to express trust opinions in a handful of systems including Youtube[5] and Facebook[6]. Additionally, LinkedIn[7] and CouchSurfing[8] adopt a reference or recommendation mechanism to capture the trust relationship between participants. There are other platforms like AllExperts[9] using ranking and FilmTrust[10] using numerals to represent the trustworthiness of users. In *Graaasp*, we use *rating* to capture and manifest global trust, and *ranking* to present contextualized and personalized trust, as *rating* and *ranking* are intuitive ways for average users to express their trust opinions. The trust model proposed in Section III is implemented into the system, where we integrate the four dimensions to construct the trust mechanism.

The detailed illustration of the global rating in Fig. 1 (a) is shown in Fig. 2. When a user selects an entity as the focal one, a rating score can be given to this particular entity in the scale of five. People, activity spaces, assets, and tools can all be rated once selected. Regarding the visibility of the rating score, three alternatives are provided: public, private and anonymous. The user is able to decide whether to make the rating opinion accessible to all the community members (public, anonymous) or restrict it only to a particular group of users (private). For every entity, an average score of all the ratings is computed, which is considered as the reputation score of this specific entity. This reputation score, which is visible in the entire community, represents the global trust perception and thus provides a social metric of trustworthiness associated with the entity.

Figure 2. Three Visibility Types of Rating

As mentioned previously, the learning context in *Graaasp* is defined by the combination of the focal entity and four columns of entities that are related to it. Within a particular context, users can rank the entities in each of the four entity lists by "drag and drop" action. As an example, the original ranking of activity spaces is presented in Fig. 1 (b). How the activity space of "*Interface Sketching*" is ranked at the top of the space list is shown in Fig. 3. This user-defined ranking of entities (generated by a specific user depending on a specific learning context) can then be considered as the personalized and contextualized trust

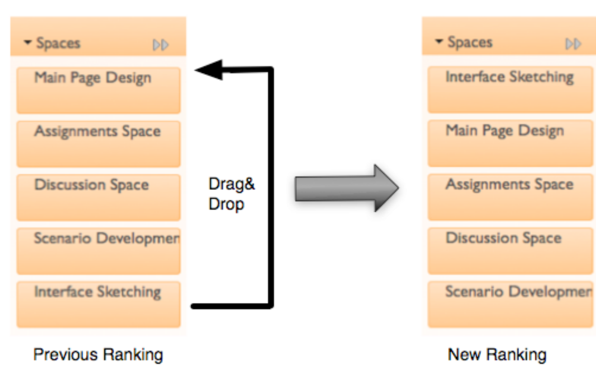


Figure 3. Ranking by "Drag and Drop" Action

[5] Youtube (www.youtube.com): a video sharing website.
[6] Facebook (www.facebook.com): a social network website.
[7] LinkedIn (www.linkedin.com): a business-oriented social network website.
[8] CouchSurfing (www.couchsurfing.org): a social network website for travelling.
[9] AllExperts (www.allexperts.com): a website for asking and answering questions.
[10] FilmTrust(trust.mindswap.org/FilmTrust): a movie review website.

assigned by the user. Explicitly ranking "*Interface Sketching*" at the top of the list suggests that this specific user strongly trusts this activity space within the current context. The customized ranking not only provides users a convenient way to organize content according to their preferences, but also enables the system to capture users' trust opinions for future use including context-dependent recommendation and effective search. As far as the visibility of trust is concerned, the user is able to keep the ranking only to himself/herself or share it with a certain group of community members indicated by the user.

V. PRELIMINARY EVALUATION

To evaluate the usefulness of the trust mechanism in *Graaasp*, a user study is conducted with participants who would be typical users of the system. The evaluation methodology and main findings are discussed in this section.

A. *Evaluation Methodology*

Typically, user studies are used to confirm the design decisions and find any problems that have been overlooked. They can range from closely controlled experimental studies testing specific hypotheses to field studies where the system is deployed for real usage and interviews are used to assess its usefulness and usability [16]. Our study falls into the first category, where the *Graaasp* system is introduced to potential users and in-depth interviews are carried out, aiming at determining whether the design of the trust model is suitable for the intended audience and for its intended purpose.

For each participant, the study consists of two parts. The first part is an introduction of the overall *Graaasp* system and the particular features related to trust mechanisms. The second part is an individual interview with carefully defined user questionnaire. During the study, participants are encouraged to "think aloud", in order to obtain running commentary while they are interacting with the system.

The user questionnaire is composed of Likert-scale questions [17] 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: usefulness of global trust, usefulness of trust visibility specification, usefulness of personalized and contextualized trust, and influence of trustee on trust. The user questionnaire is summarized in Table I.

B. *Data Collection*

The user study is conducted with ten engineering university students, who are the intended audience of the *Graaasp* system. All of them are graduate students between the age of 20 and 30. All participants are frequent Web users who visit the Internet daily and most of them have some experience of using online learning systems. All informants claimed that they were familiar with the rating, voting, ranking or similar features in online systems and used them from time to time.

Data was collected through user questionnaires and interviews. Two experimenters stayed with each participant throughout the session, observing his/her reactions, asking questions, and noting feedback of the participant. Ten individual interviews were carried out and ten questionnaires were successfully completed.

TABLE I. USER QUESTIONNAIRE

I	**Usefulness of global trust**
	I would like to *see* the global average rating score for entities.
	I would like to *provide* rating scores to entities.
II	**Usefulness of trust visibility specification**
	Can you think of a scenario where you would like to give a public rating?
	Can you think of a scenario where you would like to give an anonymous rating?
	Can you think of a scenario where you would like to give a private rating?
	When you provide a private rating, to whom do you think it should be displayed?
III	**Usefulness of personalized and contextualized trust**
	I would like to rank by myself the entity lists (Actors, Activity Spaces, Assets, and Tools).
	I would like to share my personal ranking of entities with the following people. (None, Friends, People in current learning context, Others)
IV	**Influence of trustee on trust**
	I find it useful to rate one or several types of entities. (People, Assets, Activity Spaces, Tools, None)

C. *Results and Discusstion*

All participants stated that the global rating score was useful in providing a general assessment of the quality of entities, which would assist them to select and filter content before going into details of it. They also requested for rating-based search and top-N recommendations. Some indicated that besides rating score, comments would also be valuable to give references of quality. Compared to viewing a global rating score, 9 out of 10 participants were willing to provide rating scores, whereas the remaining one explained that he would just like to view the Web content instead of reviewing it. The results suggest that with regard to the rating system, most users would not only be a viewer or consumer, but also a contributor in the rating system. This facilitates fostering a collaborative assessment environment in PLE, which is the exact objective of the trust mechanism.

Regarding visibility of trust, 7 participants pointed out that they would mostly make ratings public since they wanted to share their rating opinions with other community members. This is consistent with our previous finding that users were shown to be willing to contribute and offer guidance to others in the system. As far as anonymous rating is concerned, 4 students claimed that they would use it for negative rating opinion. We believe that the reason behind this is trying to avoid retaliation from the recipient of the rating. However, there are also other students indicating that anonymous rating would cause unfair cheating rating behaviors, since it is possible for a malicious user to abuse the anonymous rating to sabotage others. As for the private rating, 8 out of 10 participants thought that it was useful for confining the rating opinion within a small range of users instead of spreading it among the entire system. It is worth mentioning that they also preferred to define by themselves the group of users who had access to the private rating scores.

Furthermore, all participants stated that the personalized and contextualized ranking feature was helpful for prioritizing and organizing entities according to their own preferences. They thought that ranking by "drag and drop" action was quite practical and convenient. 7 participants were willing to share their personalized rankings with others, especially with users in the same learning context. While 3 participants preferred to keep the ranking private, because they thought that the personal ranking did not concern others.

Another interesting finding is that 9 out of 10 participants indicated that while they would like to rate assets, activity spaces, and tools, they felt uncomfortable to rate people. The reason behind this was explained to be the fact that they usually did not like to judge others directly since the characteristics of a person were multi-dimensional and hence difficult to assess with only a single numerical score. Instead, they felt more comfortable to evaluate the products or contributions of a person, as the quality and utility of items were easier to assess. This provides us with some insights as to how to improve the trust and reputation metrics for people in the future. A possible approach could be to calculate a person's trust and reputation score based on his/her contributions and behaviors in the system, rather than computing others' rating opinions for him/her.

To sum up, the evaluation results reveals that users not only accept the proposed trust mechanism as a way to get a general assessment of content quality, but also use it for representing personalized and contextualized preferences. Moreover, users are satisfied with being able to define the visibility of their trust opinions. Last but not least, rating people directly is not considered as a suitable approach to assess the trustworthiness of them.

VI. CONCLUSION AND FUTURE WORK

The popularity of Web 2.0 social media has brought about a large amount of user-generated content. To design effective trust and reputation mechanisms that facilitate selecting and recommending trustworthy content becomes the recent research challenge. Particularly, in PLE platforms where multi-disciplinary resources and services are aggregated from heterogeneous sources, there is a clear need for personalized and contextualized trust metrics depending on different learning goals. In this paper, to comply with the specific requirements of PLEs, we explore trust in four dimensions: trustor, trustee, context and visibility. A trust model has been proposed and implemented within a PLE prototype. To evaluate the usefulness of the trust model, a user study has been carried out through questionnaires and interviews. The results show that users are satisfied with the mechanism where trust is tackled not only on a global scale, but also on a personal and contextual scale. Being able to fully control the audience of trust opinions is also accepted as a useful feature. However, suitable trust and reputation metrics for people still need to be investigated in the future.

ACKNOWLEDGMENT

The research work described in this paper is partially funded through the ROLE Integrated Project; part of the Seventh Framework Programme for Research and Technological Development (FP7) of the European Union in Information and Communication Technologies.

REFERENCES

[1] A. Josang, "Trust and reputation systems", Lecture Notes in Computer Science, vol. 4677, pp. 209-245, 2007.

[2] O. Casquero, J. Portillo, R. Ovelar, M. Benito, and J. Romo, "iPLE Network: an integrated eLearning 2.0 architecture from a university's perspecitve", Interactive Learning Environments, vol. 18, pp. 293-308, 2010.

[3] D. Gillet, E.L.-C. Law, and A. Chatterjee, "Personal learning environments in a global higher engineering education Web 2.0 realm", Education Engineering, pp. 897-906, 2010.

[4] G. Heiberger and R. Harper, "Have you facebooked Astin lately? Using technology to increase student involvement", New Directions for Student Services, vol. 124, pp. 19-35, 2008.

[5] A. Josang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision", Decision Support Systems, vol. 43, pp. 618-644, 2007.

[6] L. Page, S. Brin, and R. Motwani, "The PageRank citation ranking: bringing order to the Web", Technical Report Standford Digital Library Technologies Project, 1998.

[7] R. Levien, "Attack resistant trust metrics", PhD thesis, University of California at Berkeley, 1998.

[8] S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina, "The EigenTrust algorithm for reputation management in P2P networks", 12th International World Wide Web Conference, pp. 640-651, 2003.

[9] A. Withby, A. Josang, and J. Indulska, "Filtering out unfair ratings in Bayesian reputation systems", International Workshop on Trust in Agent Societies (at AAMAS'04), ACM, 2004.

[10] A. Josang, "A logic for uncertain probabilities", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 9, pp. 279-311, 2001.

[11] J. Golbeck, "Trust and nuanced profile similarity in online social networks", ACM Transactions on the Web, vol. 3, 2009.

[12] F. Bobillo, M. Delgado, and J. Gomez-Romero, "Representation of context-dependant knowledge in ontologies: a model and an application", Expert Systems with Application, vol. 35, pp. 1899-1908, 2008.

[13] R. Reichle, M. Wagner, M. Ullah Khan, K. Geihs, J. Lorenzo, M. Valla, C. Fra, N. Paspallis, and G. A. Papadopoulos, "A comprehensive context modeling framework for pervasive computing systems", 8th International Conference on Distributed Applications and Interoperable Systems, pp. 281-295, 2008.

[14] S. El Helou, N. Li, and D. Gillet, "The 3A interaction model: towards bridging the gap between formal and informal learning", 3rd International Conference on Advances in Computer-Human Interactions, pp. 179-184, 2010.

[15] Gadgets Specification, http://code.google.com/apis/gadgets/docs/spec.html, retrieved Oct. 2010.

[16] S. Chiasson, P.C. Van Oorschot, and R. Biddle, "A usability study and critique of two password managers", 15th conference on USENIX Security Symposium, vol. 15, pp. 1-16, 2006.

[17] R. Likert, "A technique for the measurement of attitudes", Archives of Psychology, vol. 140, 1932.