Andreas Holzinger (Ed.)

LNCS 5298

HCI and Usability for Education and Work

4th Symposium of the Workgroup
Human-Computer Interaction and Usability Engineering
of the Austrian Computer Society, USAB 2008
Graz, Austria, November 2008, Proceedings

Springer

Lecture Notes in Computer Science 5298

Andreas Holzinger (Ed.)

HCI and Usability for Education and Work

4th Symposium of the Workgroup
Human-Computer Interaction and Usability Engineering
of the Austrian Computer Society, USAB 2008
Graz, Austria, November 20-21, 2008
Proceedings

Springer

Volume Editor

Andreas Holzinger
Medical University Graz (MUG)
Institute for Medical Informatics, Statistics and Documentation (IMI)
Research Unit HCI4MED
Auenbruggerplatz 2/V, 8036, Graz, Austria
E-mail: andreas.holzinger@meduni-graz.at

Library of Congress Control Number: 2008939367

CR Subject Classification (1998): H.5, D.2, J.3, J.4

LNCS Sublibrary: SL 2 – Programming and Software Engineering

ISSN 0302-9743
ISBN-10 3-540-89349-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-89349-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12567461 06/3180 5 4 3 2 1 0

Preface

The Workgroup Human–Computer Interaction & Usability Engineering (HCI&UE) of the Austrian Computer Society (OCG) serves as a platform for interdisciplinary exchange, research and development. While human–computer interaction (HCI) traditionally brings together psychologists and computer scientists, usability engineering (UE) is a software engineering discipline and ensures the appropriate implementation of applications.

Our 2008 topic was Human–Computer Interaction for Education and Work (HCI4EDU), culminating in the 4th annual Usability Symposium USAB 2008 held during November 20–21, 2008 in Graz, Austria (http://usab-symposium.tugraz.at).

As with the field of Human–Computer Interaction in Medicine and Health Care (HCI4MED), which was our annual topic in 2007, technological performance also increases exponentially in the area of education and work.

Learners, teachers and knowledge workers are ubiquitously confronted with new technologies, which are available at constantly lower costs. However, it is obvious that within our e-Society the knowledge acquired at schools and universities – while being an absolutely necessary basis for learning – may prove insufficient to last a whole life time.

Working and learning can be viewed as parallel processes, with the result that lifelong learning (LLL) must be considered as more than just a catch phrase within our society, it is an undisputed necessity. Today, we are facing a tremendous increase in educational technologies of all kinds and, although the influence of these new technologies is enormous, we must never forget that learning is both a basic cognitive and a social process – and cannot be replaced by technology.

Although technology is used at every level of education from K12 through university and offers opportunities for everyone to continue their education regardless of age or mobility, the most efficient technology is of little use when it is not pedagogically and consistently used, or when it fails to comply with basic didactical principles. Teaching and learning technologies are actually developed, almost exclusively, by computer scientists without explicitly seeking co-operation with pedagogues, although real increases in value could be developed in the future at exactly these places: where several disciplines meet.

In my opinion, the challenges lie at the interface of pedagogy, psychology and computer science (informatics). The concept "Life Long Learning – Continuing Education" would thereby offer the possibility of achieving such an interdisciplinary co-operation, with the goal of placing people of all age groups – rather than technological feasibility – into the center of the educational process.

Since *human cognitive evolution does not advance at the same speed as technology* the focus on interaction and communication between humans and computers is of increasing importance in education and work. The daily actions of the end users must be the central concern, surrounding and supporting them with newly available and rapidly emerging technologies. For example, information systems are a central component of modern knowledge-based work services, consequently knowledge management needs to continually be adapted to the needs, requirements and demands of end users within this steadily increasing high-tech environment.

Information processing is of increasing interest, in particular its potential effectiveness in modern education and the optimization of processes and operational sequences. It is particularly important for systems, whether educational information systems or hospital information systems and decision support systems, to be designed with the end users in mind. This year's symposium aimed to benefit end users who are educational professionals and justifiably expect software technology to provide a clear advantage; they expect to be supported efficiently and effectively in their daily activities.

Obviously this view produces specific problems; especially for younger researchers who, being new to their field and not yet firmly anchored in one single discipline, are in danger of "falling between two seats." It is certainly much easier to gain depth and acknowledgement within a narrow scientific community in a single field. However, innovation and new insights often take place at the junction of two or more disciplines; hence, this requires a broader basis of knowledge and – as a matter of fact – much more effort.

Working in any interdisciplinary area necessitates the ability to communicate with professionals in other disciplines and most of all a high willingness to accept and incorporate their points of view. USAB 2008 was organized in order to promote this closer collaboration between software engineers, psychology researchers and educational professionals.

USAB 2008 received a total of 85 submissions. We followed a careful and rigorous two-level, double-blind review, assigning each paper to a minimum of three and maximum of six reviewers. On the basis of the reviewers' results, 7 full papers ($\geq$ 18 pages) and 26 short papers were accepted.

USAB 2008 can be seen as a bridge within the scientific community, between computer science, psychology and education.

Everybody who gathered together to work for this symposium displayed great enthusiasm and dedication.

I cordially thank **each and every person** who contributed toward making USAB 2008 a success, for their participation and commitment: The authors, reviewers, sponsors, organizations, supporters, all the members of the organization team, and all the volunteers; without whose help, this bridge would never have been built.

November 2008 Andreas Holzinger

Organization

Programme Committee

Ray Adams	Middlesex University London, UK
Sheikh Iqbal Ahamed	Marquette University, USA
Henning Andersen	Risoe National Laboratory, Denmark
Keith Andrews	TU Graz, Austria
Sue Bogner	Institute for the Study of Human Error, LLC Bethesda, USA
Noelle Carbonell	Université Henri Poincare Nancy, France
Tiziana Catarci	Universitià di Roma La Sapienza, Italy
Luca Chittaro	University of Udine, Italy
Matjaz Debevc	University of Maribor, Slovenia
Alan Dix	Lancaster University, UK
Martin Ebner	Graz University of Technology, Austria
Judy Edworthy	University of Plymouth, UK
Pier Luigi Emiliani	National Research Council, Florence, Italy
Vlado Glavinic	University of Zagreb, Croatia
Sabine Graf	National Central University Taiwan, Taiwan
Andrina Granic	University of Split, Croatia
Eduard Groeller	TU Wien, Austria
Sissel Guttormsen-Schär	University of Bern, Switzerland
Martin Hitz	University of Klagenfurt, Austria
Andreas Holzinger	Medical University Graz, Austria
Timo Honkela	Helsinki University of Technology, Finland
Kasper Hornbaek	University of Copenhagen, Denmark
Ebba P. Hvannberg	University of Iceland, Reykjavik, Iceland
Julie Jacko	Georgia Institute of Technology, USA
Chris Johnson	University of Glasgow, UK
Anirudha N. Joshi	Indian Institute of Technology, Bombay, India
Erik Liljegren	Chalmers Technical University, Sweden
Zhengjie Liu	Dalian Maritime University, China
Frank Kappe	Graz University of Technology, Austria
Kinshuk	Austriahabaska University, Canada
Hermann Maurer	Graz University of Technology, Austria
Klaus Miesenberger	University of Linz, Austria
Silvia Miksch	Donau University Krems, Austria
Lisa Neal	Tufts University School of Medicine Boston, USA
Alexander Nischelwitzer	FH Joanneum Graz, Austria
Shogo Nishida	Osaka University, Japan
Hiromu Nishitani	University of Tokushima, Japan
Nuno J Nunes	University of Madeira, Portugal

Anne-Sophie Nyssen	Université de Liege, Belgium
Erika Orrick	GE Healthcare, Carrollton, USA
Philipe Palanque	Université Toulouse, France
Helen Petrie	University of York, UK
Margit Pohl	Vienna University of Technology, Austria
Robert W. Proctor	Purdue University, USA
Harald Reiterer	University of Konstanz, Germany
Wendy Rogers	Georgia Institute of Technology, USA
Demetrios Sampson	University of Piraeus, Greece
Anthony Savidis	ICS FORTH, Heraklion, Greece
Nikolai Scerbakov	Graz University of Technology, Austria
Albrecht Schmidt	Fraunhofer IAIS/B-IT, Bonn University, Germany
Lynne Schrum	George Mason University, Fairfax, USA
Andrew Sears	UMBC, Baltimore, USA
Ahmed Seffah	Concordia University, Montreal, Canada
Katie A. Siek	University of Colorado at Boulder, USA
Daniel Simons	University of Illinois at Urbana Champaign, USA
Christian Stary	University of Linz, Austria
Constantine Stephanidis	ICS FORTH, Heraklion, Greece
Zoran Stjepanovic	University of Maribor, Slovenia
Harold Thimbleby	Swansea University, Wales, UK
A Min Tjoa	Vienna University of Technology, Austria
Klaus Tochtermann	Graz University of Technology, Austria
Manfred Tscheligi	University of Salzburg, Austria
Jean Underwood	Nottingham Trent University, UK
Karl-Heinz Weidmann	FHV Dornbirn, Austria
William Wong	Middlesex University, London, UK
Panayiotis Zaphiris	City University London, UK
Jürgen Ziegler	Universität Duisburg Essen, Germany
Ping Zhang	Syracuse University, USA

Organizing Committee

Thomas Billicsich	Graz University of Technology
Martin Ebner	Graz University of Technology (Local Organizing Chair)
Christine Haas	Austrian Computer Society
Andreas Holzinger	Medical University of Graz
Marie-Luise Lampl	Graz University of Technology
Sandra Leitner	Austrian Computer Society
Walther Nagler	Graz University of Technology
Elisabeth Richter	Graz University of Technology
Christian Stickel	Graz University of Technology
Elisabeth Waldbauer	Austrian Computer Society

Members of the WG HCI&UE of the Austrian Computer Society

(In alphabetical order – status as of September 19, 2008, Total: 165)

Albert, Dietrich
Ahlstroem, David
Aigner, Wolfgang
Andrews, Keith
Auinger, Andreas
Baillie, Lynne
Baumann, Konrad
Bärnthaler, Markus
Bechinie, Michael
Bernert, Christa
Biffl, Stefan
Binder, Georg
Bloice, Marcus
Breiteneder, Christian
Burgsteiner, Harald
Dirnbauer, Kurt
Derndorfer, Christoph
Debevc, Mathias
Ebner, Martin
Eckhard, Benedikt
Edelmann, Noelle
Ehrenstrasser, Lisa
Erharter, Dorothea
Errath, Maximilian
Ferro, Bernhard
Flieder, Karl
Freund, Rudi
Frühwirth, Christian
Füricht, Reinhard
Geierhofer, Regina
Gorz, Karl
Graf, Sabine
Graf, Sylvia
Grechenig, Thomas
Grill, Thomas
Groeller, Edi
Haas, Christine
Haas, Rainer
Haberfellner, Tom
Hable, Franz
Hacker, Maria
Hoeller, Martin
Leitner, Gerhard
Leitner, Hubert
Linder, Jörg
Maier, Edith
Maitland, Julie
Makolm, Josef
Mangold, Pascal
Manhartsberger, Martina
Mayr, Stefan
Meisenberger, Matthias
Melcher, Rudolf
Messner, Peter
Miksch, Siliva
Miesenberger, Klaus
Motschnig-Pitrik, Renate
Musil, Sabine
Müller, Regine
Mutz, Uwe
Nemecek, Sascha
Nischelwitzer, Alexander
Nowak, Greta
Oppitz, Marcus
Osterbauer, Christian
Parvu, Andrej
Pellegrini, Tassilo
Pesendorfer, Florian
Pohl, Margit
Purgathofer, Peter
Rauhala, Marjo
Ramkisson, Arun
Reichl, Peter
Richter, Elisabeth
Richter, Helene
Riener, Andreas
Robier, Hannes
Sahanek, Christian
Schaupp, Klaus
Scheugl, Max
Schloegl, Martin
Schreier, Günther

Hackl, Erich Patrick
Hailing, Mario
Haug, Bernd
Hauser, Helwig
Heimgärtner, Rüdiger
Herget, Martin
Hitz, Martin
Holzinger, Andreas
Hruska, Andreas
Huber, Leonhard
Hyna, Irene
Jaquemar, Stefan
Jarz, Thorsten
Kainz, Regina
Keki, Susanne
Kempter, Guido
Kittl, Christian
Kingsbury, Paul
Kleinberger, Thomas
Koller, Andreas
Kotsis, Gabriele
Költringer, Thomas
Kriglstein, Simone
Krieger, Horst
Kriegshaber, Ursula
Kohler, Kirstin
Krümmling, Sabine
Kment, Thomas
Lanyi, Cecilia
Leeb, Christian
Lenhart, Stephan
Lugmayr, Arturri
Loidl, Susanne
Wassertheurer, Sigi
Windlinger, Lukas
Wöber, Willi
Wohlkinger, Bernd
Wolkerstorfer, Peter
Zellhofer, Norbert
Schwaberger, Klaus
Schwantzer, Gerold
Searle, Gig W.
Sefelin, Reinhard
Seibert-Giller, Verena
Seyff, Norbert
Slany, Wolfgang
Spangl, Jürgen
Sproger, Bernd
Stary, Christian
Stenitzer, Michael
Stepanovic, Zoran
Stiebellehner, Johann
Thümer, Herbert
Thurnher, Bettina
Tjoa, A Min
Tscheligi, Manfred
Urlesberger, Berndt
Vecsei, Thomas
Vogler, Robert
Waclick, Olivia
Wagner, Christian
Wagner, Claudia
Wahlmüller, Christine
Weidmann, Karl-Heinz
Weippl, Edgar
Werthner, Hannes
Wimmer, Erhard
Zorn-Pauzi, Gabriele

Sponsors

We are grateful to the following companies and institutions for their support in our aims to bridge science and industry. Their logos are displayed on the next page (status as of September 17, 2008). For a current list please refer to thewebsite:

http://usab-symposium.tugraz.at

OESTERREICHISCHE COMPUTER GESELLSCHAFT
AUSTRIAN COMPUTER SOCIETY

TU Graz
Graz University of Technology

Technische Universität Wien
Vienna University of Technology
TU VIENNA

Medical University of Graz

ifip

ManGold
MEET YOUR SOLUTION

tobii

FERK SYSTEMS

Stadt GRAZ

upa
www.upassoc.org

BMD BUSINESS SOFTWARE

CIRCUA
Collaborative International Research Centre

SEMANTIC WEB SCHOOL
Zentrum für Wissenstransfer

Table of Contents

Cognitive Processes and End User Experience

Social Software and Collaborative Knowledge Development

Knowledge Processing, Assessment and Human Performance

Human Centered Development and Design Accessibility

User Centred Information Visualization and Multimedia

Adaptivity and Personalization in Ubiquitous Learning Systems

Human Centred Design for Safety Critical Systems

Learners, Technology and the Brain

Jean Underwood

Division of Psychology,
Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK
jean.underwood@ntu.ac.uk

Abstract. How do we develop an educational experience fit for twenty-first century learners? There is a perceived malaise at the heart of our education system. The solution to this problem proposed by many governments, although vociferously resisted by some scholars and parent groups, is to use the lure of digital technologies to fix the broken system. Will this work? What evidence is there for the efficacy of digital technologies and particularly video games for learning? A combination of traditional educational evaluations and studies within cognitive neuroscience provide a window on the nature, process and effectiveness of learning with and through digital technologies. Such studies are changing our perception of learning, which is now more brain training than knowledge acquisition. These changes marry both the cognitive and affective aspects of learning.

Keywords: brain training, engagement, effective learning, c-maturity, video gaming.

1 Introduction

The introduction of any new technology into a society is always controversial and the impact of digital technologies on young people, particularly aspects of the Internet and of video games, arouses passionate debate [1], [2]. After more than three decades of educational computing the ferocity of the debate has not diminished. There are those who see digital technologies not only as a waste of young people's time and a drain on our educational system [3], [4], [5] but also as a danger to students' mental and physical health [6], [7]. In contrast the champions see digital technologies as at least a partial solution to the malaise in the UK and also the US educational systems [8], [9], [10]. Where does the truth lie? The role of digital technologies as tools for learning, specifically through video game playing, is used to elucidate this debate.

However, before investigating the role of digital technologies in education we need to identify what it is we want learners to learn. As a starting point I assume that the prime directive for education is to facilitate the development of skilled performance; that is the development of the cognitive and social abilities of the learner, both in terms of low-level procedural competences and also in the range of skills needed to operate effectively within the complex dynamic environment which is our society. Such skills include the ability to problem solve and to interact effectively with others [9]. Educational systems in the developed world are under scrutiny because they are

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 1–18, 2008.

not delivering on either of these core aims of developing basic skills nor the higher order skills. As a result these systems are deemed ill-equipped to prepare learners for an increasingly complex world.

A key focus of educational policy in both the UK and US has been on the basic skills of “at risk” children. One such group in the UK is defined as NEETS, that is 16-18 year olds not in education, employment or training. These are adolescents who are failing to acquire even the basic skills which will allow them to participate as full citizens in tomorrow’s world. The identification of such groups has led to a raft of remedial educational policies linked to standardised assessment to ensure that all children make adequate yearly progress. This is seen as the key to integrating such students into society as active members of the workforce. While this is a laudable aim, Shaffer and Gee [11] argue that the focus on basic-skills-for-all means that young people are being prepared for commodity jobs. These are the jobs that can be done more cheaply and as efficiently in the developing world, but the survival of first world countries, they argue, is dependent on innovation linked to a high skilled workforce. The economic imperative for educational change espoused by Shaffer & Gee is endorsed by the Royal Society of Arts [12] in the UK which argues that young people’s real needs are not being met by a National Curriculum focused on content rather than on the life-skills necessary for the next generation to function efficiently and effectively. This illustrious body also argues that education must change if our society is to become more comfortable with itself.

How then do we meet the need to cover the basic skills, the building blocks for learning, while producing a new generation of flexible problem solvers? We first acknowledge that the development of such skills is not exclusive to formal education [13, 14] and that digital technologies are blurring the boundaries between formal and informal learning. We need then to identify how we can harness the functionalities of digital technologies to support our global educational aims. That this is an imperative is highlighted by the establishment of ‘The National Center for Research in Advanced Information and Digital Technologies’ in the US at a cost of $50M. This Centre has a remit to conduct to explore ways that advanced computer and communication technologies can improve all levels of learning [15].

While there are many ways in which digital technologies can support learners and learning, for example through the e-library or e-discussions, I have chosen to concentrate on video gaming here. Video gaming allows an exploration of the nature of learning through powerfully motivating environments while at the same time confronting one of the most controversial digital activities both within education and the world beyond the classroom. There are a plethora of questions arising from an analysis of the educational use of games. Is video game playing part of the solution for an educational system in crisis or is the passion for game playing exhibited by many young people exacerbating the problem of an ill-prepared citizenry? What, if any, is the educative potential of gaming? How might games be used or adapted to stimulate learning? How do games affect the way we think and why should they have impact? How might schooling need to adapt to accommodate to the use of games? Before looking at the research evidence that might answer some if not all of these questions we need to understand something about the processes by which learning takes place.

2 How Do People Learn?

One of the very real difficulties in establishing the what, the when and the how of learning is that there are so many learning theories each with an entrenched group of supporters who reject other views of learning. The current dominant theoretical position within education marries constructivism with situated action, and emanates from Gestalt psychology. This view of learning has usurped the cognitivists' and before them the behaviourists' model of learning but these and other models do have merit. If we look more closely at the raft of theories we often find there are significant points of overlap and agreement. For example although Caine and Caine [16] argue that all learning is physiological, they recognise the affective aspects of learning. They assert that meaning, and hence learning, occurs through patterning and that emotions are critical to that patterning. The importance of practice also has biological roots as those neurons that repeatedly fire form more and stronger connections [17]. However, this is a more subtle form of learning than mere drill because there is a level of selectivity as neurons have the ability to stop firing when the stimulus turns out to be unimportant. Mere drill and practice models of learning do not take this selectivity into account. Zull [17], as Caine and Caine [16], acknowledges the role of emotion in the learning process. Chemicals, including adrenaline, serotonin, and dopamine released during the act of learning, lead to changes in the neuronal network. So Neuroscience confirms that practice is central to successful learning although Zull adds that practice in a meaningful way only occurs when the student is engaged. For Zull [17] the art of teaching is to find ways that make learning intrinsically rewarding. He agues that this is best achieved through his 4 pillars of learning each of which promotes brain activity and so collectively exercise the whole brain (Table 1), a model that resonates with Ausubel's [18] advanced organizer model of learning.

Table 1. Basic brain functions and Zull's Four Pillars of Learning [17]

Basic brain functions	Brain sites	Pillars of learning
1. Getting information	Sensory cortex	Gathering
2. Making meaning of information	Back integrative cortex	Analyzing
3. Creating new ideas from these meanings	Front integrative cortex	Creating
4. Acting on those ideas	Motor cortex	Acting

The perception of learning as brain exercise rather than as knowledge acquisition is gaining currency outside of education, not least because there is a need to maintain the mental well-being of an ageing population [19]. Brain training is seen as one way to slow down and even reduce the inevitable decline in cognitive functions of the 'silver' generation and results from the recognition that the brain is not immutable and that environmental influences are capable of altering brain structures. This has led both researchers and educators to question the capacity of the brain to respond to enrichment for learners of all ages. Some of the most vivid accounts of the susceptibility of the brain to training, that is its plasticity, occur in Doige's descriptions of brain-damaged individuals [20]. These individuals' who have lost functionality in, for example, perception or motor control, have been shown to recover a level of functionality as activity is routed through non-standard, rather than the damaged standard, pathways in the neural network [20]. While Diamond [21] points out that ascertaining the nature of what constitutes "enrichment" for humans, as opposed to the rats on which most studies have been conducted, is difficult because of individual biological and environmental differences. However, what evidence there is confirms the basic finding that dendritic growth in response to environmental stimulation correlates with learning, suggesting that newness and challenge are as important for the development of the human cortex as that of animals. As Diamond [21] notes, enrichment effects on the brain have consequences for behaviour, and she argues that parents, educators, policy makers, and can all benefit from such knowledge.

Alongside general theories of learning there are also those closely associated with e-learning. Mayer and Moreno [22], [23] have applied the Dual Coding Theory from cognitive psychology [24] to multimodal learning in digital environments. The basic premise of the Dual Coding Theory is that cognition involves two subsystems, a verbal subsystem to process language and a non-verbal imagery subsystem to process non-linguistic information. The theory assumes that visual and auditory information is processed via different verbal and visual systems which can be activated independently but are connected allowing dual pathways allowing more efficient coding of information but that this is limited by the capacity of each pathway [25]. The view that there is limited capacity overlaps with Sweller's [26], [27] Cognitive Load Theory, which states that a learner's attention and working memory is limited. Coupled to this model of processing Mayer [22] argues uncontroversially that learning is an active process thus linking the cognitivist and constructivist descriptions of learning.

The third theory finding currency in the e-learning community is Flow Theory [28]. Flow is described as a mental state that occurs when an individual is fully immersed in an activity. Flow experiences are intrinsically rewarding involving as they do intense involvement, focused attention, clarity of goals leading to a lack of self-consciousness and a feeling of full control over the activity. Athletes often describe this state of consciousness as 'being in the zone.' A state where self and task merge, results in the individual being intrinsically motivated to repeat the activity now deemed to be worth doing for its own sake. Flow Theory has been widely used to explain the feeling of telepresence in the virtual environments, that is the state of consciousness that gives the impression of being physically present in a mediated world [29], [30]. This theory has been extensively used to explain the lure of video games [31], [32].

Although there are many theories of learning at the core we can see that learning leads to physiological changes in the brain, which provide us with new ways of monitoring any learning activity. Learning requires attention and practice but attention and practice require the emotional engagement of the learner. Engagement comes from individuals seeing an activity as relevant to themselves but also achievable by themselves. There are then physiological, cognitive and affective aspects to learning and they are all important and can be demonstrated to occur when video game playing, So video games show promise as learning tools. However, to be of value the learning from video gaming needs to lead to transferable real world skills rather than to parochial skills of the game *per se*. In the next sections of this paper I present the evidence for the efficacy of games as learning tools.

3 Video Games as a Window on Learning

Learning is a move from a state of not knowing or naivety to one of expertise but how do individuals achieve expertise? Can we establish how individual differences and task characteristics impact on knowledge and skills acquisition and how do task and learner variables interact? To what extent can expertise transfer across tasks and domains? Understanding the answers to these questions is essential if we are to produce a 21st century education. The use of Tetris, and other video games, to explore problem-solving is a response to arguments about the artificiality of many earlier laboratory studies of learning, which it is argued fail to capture the superior performance of experts because such fixed tasks constrain the expert performer [33]. On the other hand, sitting in front of a computer screen is the natural environment for the video game-player and so studying game playing not only provides a window on expertise in context it also provides a learning environment in which engagement by the learner is generally ensured.

Learning through game playing, either through serious games [34], games developed with educational intent, or video games designed for leisure may hold the key to a more effective educational system. Prensky [35] argues that the move to student-driven learning will generate a demand for educational games, as they are the natural environment for the digital natives now in our schools. Rather than being part of the problem, he argues that games are a powerful tool in any school's armoury when dealing with adolescents who necessarily pass through a state of personal identity exploration and wandering attention comes to the for. This is a phase of development that is often disruptive to good classroom management. This period of independence seeking, according to Eccles et al. [36] is often met within the school by increasing behavioural control and decreasing complexity of the cognitive tasks required of learners, thus creating a mismatch between the adolescents' developmental stage and the school learning environment. By providing problem-focused learning, games not only motivate but also provide the increasing complexity sought by such learners.

3.1 The Nature of Games and Game Players

Who Plays? There is a clear discontinuity between those in the anti-gaming lobby and people's actual experiences of playing games [37]. The stereotype of the game player as a social inadequate, almost inevitably male, who spends hours in isolation because of an inability to operate in the real world, is a nightmare scenario for many because of the manifest popularity of technology-based games for students-of-all-ages. This stereotype does not match gamers' perceptions of themselves as highly socialised individuals. Nor does it match the research data on game players.

While a 2-year log-file study of adolescents access to the Internet at home showed an average play-time of about 3 hours a week but with some 10% of those surveyed on-line for 16 hours a week [38], this does not mean that game playing is necessarily the dominant use of digital technologies. In a recent survey of young people's use of digital technologies [39] we found that game playing as the main interaction with technology (central interest activity) peaks at around 10 or 11 years of age, an age at which the Gameboy and other consoles dominate (Fig. 1). However, the onset of

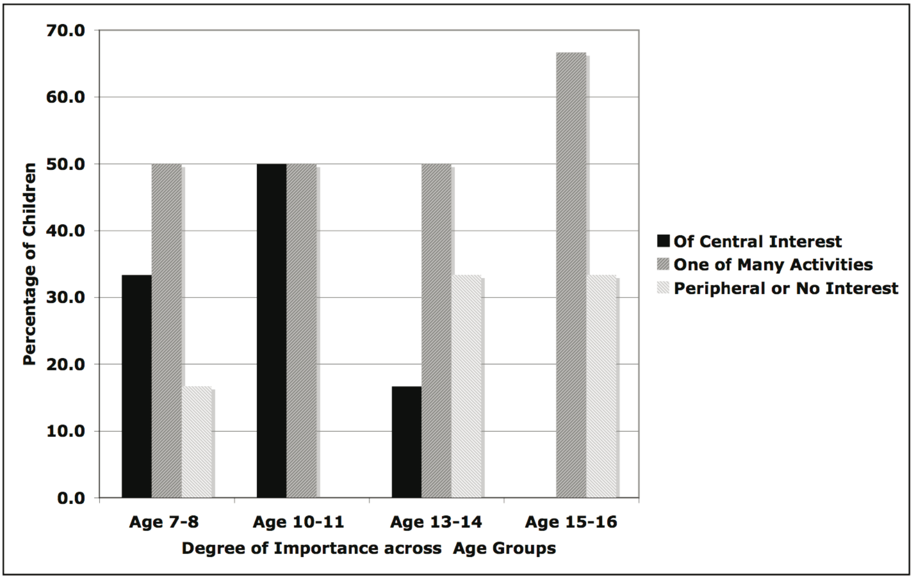


Fig. 1. Importance of Video Game playing compared to other activities using digital technologies by age [39]

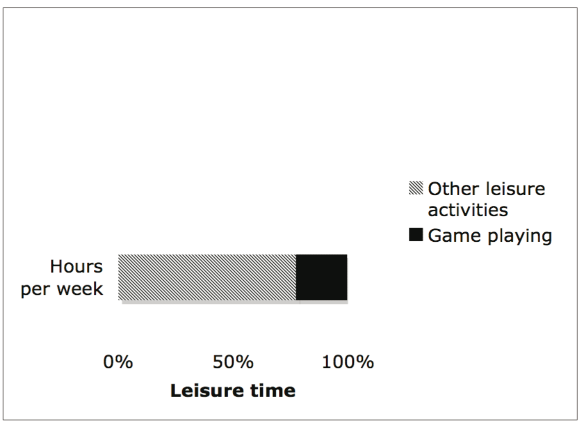


Fig. 2. US Game players distribution of leisure time gaming and non-gaming leisure activities [42]

adolescence coincides with growing social awareness and, while the level of game playing does not necessarily diminish, proportionally other activities such as the use of social networks become as important, if not more important, than playing games. On-line gaming, as opposed to console gaming, is not the prerogative of the young. In the UK Griffiths et al. [40] found that over 60% of on-line gamers were over 19 years of age. In the US, 8% of all Americans played on-line games in 1999, rising to 37%

of the total population by 2003 [41]. The Entertainment Software Association's (ESA) survey showed that the average age of game players was 30 years [42] and contrary to popular perception of a typical game-player, between 43% [42] and 50% [43] were female gamers. The female presence was noticeable in the large multiplayer on-line games. This survey also belies the perception that game players elect for social isolation. (Fig. 2), rather they have a varied leisure experience, which includes involvement in activities such as sport, volunteering, cultural experiences and reading. While Lawry et al. [44] found that many boys who played electronic games had interests in music, programming, reading, and school.

Interestingly, Sandford et al. [45] found a significant majority of teachers (72%) do not play games for leisure. The digital divide between teacher and student appears to be more a life-style choice than a generational gap, that is, many teachers choose not to play games, while peers in other occupations do. Both Sandford's teachers and students agreed that games were motivating, resulting in increased student engagement. However, that motivation was more apparent when leisure rather than educational games were used and when students had a degree of autonomy in playing the game. The educational benefits of gaming were not evident to either students or teachers. For such benefits to be recognised by teachers, Sandford and his colleagues argue that they need to identify the learning opportunities of such games but as the teaching population has a high proportion of game-refusniks, acquiring such an understanding may prove difficult [46].

The Nature of Games: Technology-based games are highly varied and difficult to classify with any precision ranging as they do from Atari's ping-pong, which dominated the scene in the 1970's, to the best selling Mortal Combat, rated the top game in 1993 to the Massive Multi-player Role-playing Games (MMORPGs) such as World of Warcraft and the virtual worlds of Second Life. Many of the early games were mini-games in Prensky's [35] parlance, as are the majority of educational games available on the Internet. In contrast the virtual worlds epitomised by the MMORPGs are complex games. These too have their parallels in the education world for example Econ 201. Prensky argues that mini-games are not 'bad' for learning although they tend to be restricted to single skills learning while the learning from complex games is multi-faceted. Such virtual worlds, it is argued, present exciting opportunities to education [47], [48]. In these rich virtual environments learning is inherently social and embodied by the movement of avatars within vast virtual spaces. Thomas and Brown [48] argue they provide a blueprint for bespoke educational games as they foster communities of practice so beneficial to learning according the Lave and Wenger [49]. However, the experience of playing a game like the Sims or Civilization III, a cerebral blend of planning, building, managing, and competing with other civilizations; is very different from playing games, such as World of Warcraft, which require rhythm and timing as they encourage participants to immerse themselves in complex virtual societies [50].

The theory presented here suggests that games should aid learning but what evidence have we to hand to validate the theory?

3.2 Learning as Brain Training

The research evidence shows that gaming is both a powerful tool to understand learning and also to encourage learning but that the nature of that learning can be constrained and is not always guaranteed. Griffith, Voloschin and Bailey [51] found that video-game users had better eye–hand coordination as measured by a pursuit rotor task. Drew and Waters [52] also showed a relationship between increased video-game playing and improvements of eye–hand coordination, as well as manual dexterity, and reaction time. As Southwell and Doyle [1] point out, however, these studies have not identified a causal link and we are left with the question of whether playing such games leads to skill development or whether players with the requisite skills are drawn to games?

Green and Bavelier [53] have conducted an exciting series of five experiments that provide stronger evidence of both transfer of training of general cognitive skills and an answer to the causal question. Their opening premise is that exposing humans (or any other organism) to an altered visual environment is likely to lead to a modification of the visual system and perceptual learning likely to result in improved performance. They recognize that such perceptual learning, has often been found to be task specific [54] and that education requires to know whether the expertise acquired through gaming is based on automaticity of skill [55] and the use of sub-routines of performance, or whether that expertise is more general and hence can be flexibly used across a range of tasks. They have argued that specific routinised skills that have been acquired implicitly are difficult to transfer but where explicit knowledge has been acquired the transfer of skills to other situations should be enhanced.

Green and Bavelier's [53] five experiments demonstrate that action video game play increases the number items that tracked simultaneously over time with habitual action video game players displaying both a qualitative and quantitative improvement in object tracking when compared with non-players. They also establish a causal effect of action video game play demonstrating that participants who are not self-selecting game players, once trained on an action video game exhibit similar enhancements to those initial participants who are keen players. This series of experiments suggest that action video game play may enhance some aspects of visual working memory.

There is supporting evidence for Green and Bavelier's [53] optimism concerning the efficacy of game playing, much of which has been conducted using Tetris. This real-time interactive video game has been one of the most popular games for over three decades. It is more than 30 years since Alexey Pazhitnov then at the Moscow Academy of Sciences, developed the game and variants of the game are now available on nearly every video game console and computer operating system, as well as on devices such as graphing calculators, mobile phones, portable media players, and PDAs.

Tetris consists of a playing area located off-centre to the left of the screen with an area providing performance feedback to the right of the screen (for example see Fig. 3). The game starts with an empty playing field. On each trial one of seven possible zoids, each uniquely constructed of four squares or bricks, appears at the top of the screen and descends at a regular speed to the bottom. As the game proceeds the speed increases. As soon as one piece comes to rest a new piece begins its descent. The objective of the game is to move and rotate the falling zoids so that they form a complete horizontal line across the playing field. When such a line is completed it will disappear and the zoids above will drop down freeing playing space above and

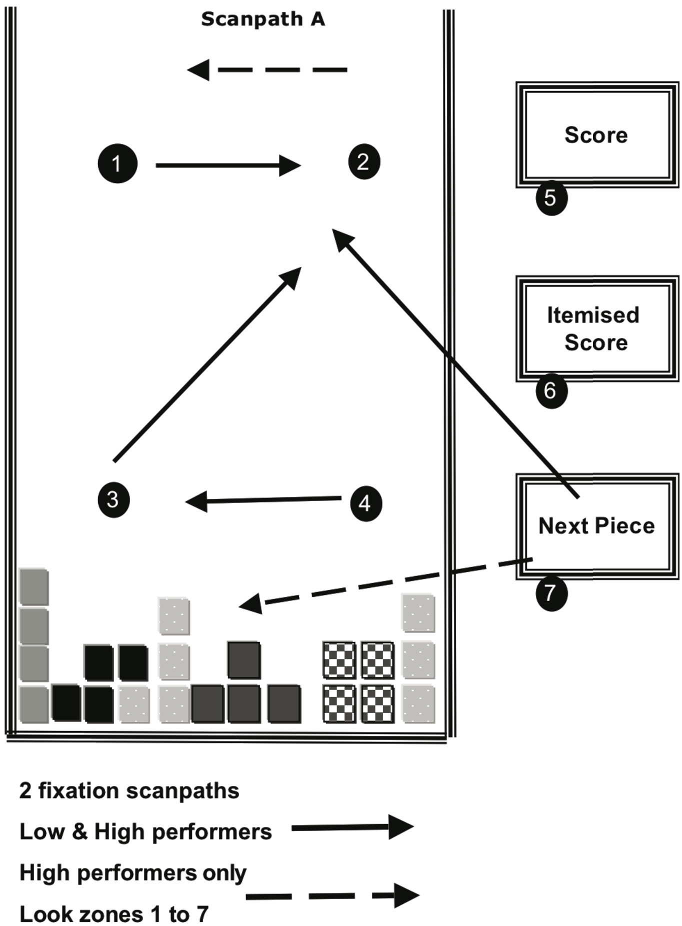


Fig. 3. The Tetris filed with 2-fixation scanpaths exhibited by all players (low and high) and high players only. (Scanpaths are successive eye-fixations: All scanpaths were significant at $p<0.05$ or more) [56].

extending the game playing time. A player loses when the playing space becomes restricted and a new piece is blocked by filled grid squares and cannot descend at all. The player has a limited range of actions which he can take in order to align zoid to build the point-winning line; In the version used by Underwood [56] these were to move the zoid to the left or to the right, to rotate the zoid clockwise and to finally to drop the zoid vertically once the zoid has been aligned with the chosen resting place [Fig. 3]. The player's goal is to maximize his or her score by placing pieces to complete a row of bricks. Playing TETRIS requires the abilities to recognise shapes, to mentally rotate those shapes and to place those shapes in a point-winning position. In addition the player may selectively choose to access peripherally presented information on upcoming objects and on personal performance. All actions are undertaken under increasing time pressure as the presentation of new target zoids is accelerated. Although seemingly simple, Tetris has been defined as highly complex problem [53].

Demaine, Hohenberger and Liben-Nowell [57] argue that Tetris belongs to the class of problems defined as NP-complete problem, which means it is as least as difficult to solve as any other NP "Non-deterministic Polynomial time" problem. In computational complexity theory a NP-complete problem is one that is both NP (verifiable in nondeterministic polynomial time) and NP-hard (any NP-problem can be translated into this problem) of which the traveling salesman problem is a classic example. So in Tetris, although it is relatively easy to check whether a solution to the problem is valid, there is seemingly no efficient way to optimize any of the game's goals such maximising the playing space and thus extending the game time or maximising the number of simultaneously deleted. From the psychology literature Tetris is best defined as an unlimited moves problem, as compared to one-stop and limited moves problems, although it lacks a final definitive solution.

Although fraught with complexity Tetris is an easy-entry activity, that is it is readily accessible to the novice player. DiSessa [58] has argued vigorously that if technological tools are to become effective learning tools then the entry costs to working with the tool must be reduced otherwise the learners will be demotivated. I will return to this issue when discussing complex games, but the popularity of Tetris clearly shows that easy entry followed by a task that stretches the individual is captivating.

One of the concerns raised by the educational anti-technology lobby is that playing video games impacts on the brain. Indeed they do but why should this be a cause for concern? It is well established that performance of complex cognitive tasks, of which video games are an example, are associated with changes in brain chemistry that result in high performers' brains seemingly functioning more efficiently than those of the poor performers. Haier and his colleagues [59] point out that high performers develop well defined and parsimonious brain circuitry requiring fewer neurons to complete complex task, thus requiring minimal glucose use, while poor performers use more circuits and/or neurons, some of which are inessential or detrimental to task performance, and this is reflected in higher overall brain glucose metabolism. They have shown that levels of glucose show such changes with practice and growing expertise when playing Tetris. While there are a number of interesting findings in this study one of the most important here is that the team demonstrated that effective game playing drew not just on procedural skills but also required declarative knowledge.

So playing Tetris does impact on the brain and that impact changes with increasing expertise, however what do individuals learn when they play Tetris? The transfer of problem solving skills can be at a variety of levels: transfer of general skills, transfer of specific skills, and transfer of specific skills in context [57]. Those proposing general-skills transfer argue that learning to problem solve in a domain such as video gaming acts as a mental work-out, improving the mind in general. Many video games have a strong spatial component, as is true for TETRIS the focus of the empirical work presented here, and from the general skills perspective performance on such games should not only be related to tests of spatial ability, but growing expertise on such games should also lead improvement in the overall spatial competence. The argument for the transfer of specific skills suggests that only those skills inherent in the problem solving situation, in the case of TETRIS skills such as mental rotation [60] and visualization [61] will be honed and therefore transferred to other situations. The more limited view of transfer, that of specific skills in context, suggests that component skills such as mental rotation are not altered in any global sense but are used more efficiently when the stimuli are the same or near equivalent because of stored mental representations of those familiar stimuli [52]. There is evidence of both general [61], [62], [63], [64] and specific spatial skills [65], [66]. These results suggest that spatial expertise is highly domain specific and does not transfer generally to other domains.

Kirsh and Maglio [67] through computational modelling and Underwood [56] through eye-tracking of novice and expert game players, have each sought to identify whether successful game-players exhibit different playing strategies than their less successful peers or whether they are simply more efficient in the strategies they used. The evidence from these two studies shows success is built upon strategic actions, encompassed in the use of lateral transitions that allow successful players to increase the reliability of a judgement of the space-time resources needed to compute a solution. Such lateral transitions are shown by the 2 fixation scanpaths in Fig. 3. Scanpaths are sequences of successive fixations, showing the sequence of areas of interest for the game player. It is noticeable that the high level game players exhibit more of these laterals. Theories of action must incorporate not only planning and action themselves but also information gathering. Lateral A, a seemingly meaningless move from right to left, is a focusing move by high level players allowing them a new perspective on the playing area. However, Kirsh and Maglio [67] argue that such epistemic activity encompasses far more than sensor related activity. This top-down processing, from a situated cognition perspective, allows environmentally controlled perceptual properties to guide attention and eye movements in ways that assist in developing problem-solving insights that dramatically improve reasoned action [68].

There is now a growing body of evidence that game playing has benefits in diverse real world situations. Rosser and his colleagues [69] found that video game skills correlate with laparoscopic surgical skills and suturing ability. Surgeons who were proficient gamers performed 47% fewer surgical errors than those who had limited or no gaming experience. They argue that some element of game playing should be incorporated into practical medical training. Research at York University, Toronto on the potential decrement in performance on driving while using a hands-free mobile phone is even more compelling [70]. Testing gamers and non-gamers in a driving simulator, Telner and his colleagues found the standard decrement in driving performance for non-gamers when using a mobile. However, gamers were significantly less impaired by the dual task challenge presented by driving and using the mobile. While these results are promising it should be noted that not all games will promote such learning. There are certain ideal game types that will promote learning. For Green and Bavelier [53] these are action games which force players to juggle a number of varied tasks such as detecting one or more opponents while avoiding getting hurt. Such games require better management and speeded perceptual processing in the central executive. They are in essence muscle-building exercises for the brain. While Gee [47] emphasises the social dimension of gaming and argues that tracking the enemy in a virtual dungeon has an equivalence in skill usage to communicating with 'friends'.

3.3 But Playing Games Can Be Tougher Than Going to Lessons

So games have impacts but do the students always respond positively to them? Conflicting evidence comes from the complex game Civilization III. Squire [70], [71] found that bringing a commercial-quality game such as Civilization III, an historical simulation, into the classroom created as many motivational problems as it solved. Even though his sample of US high school students were all out-of-school game players many questioned the validity of game playing in the classroom. Why was the initial response often so negative? A key reason was that the entry-level skills were too taxing for many of the pupils in stark contrast to a game like Tetris. Students needed 6 to 7 hours of play to understand and be able to operate at a basic level, time they seemed unwilling to donate to the task. Squire found a quarter of his students were highly resistant complaining the game was too complicated, uninteresting and more difficult than anything they encountered in school to date. These students elected to withdraw to participate in reading groups, rather than continue in the game. Conversely he also found that 25% of the students (particularly academic underachievers) were highly motivated to learn history through the game, and considered the experience a highlight of their school year. For these students, who often had issues with authority, replaying history was not only motivating it enhanced performance with the students developing new vocabularies, better understandings of geography, and more robust concepts of world history [70], [71]. The individual differences highlighted by Squire's study are highly pertinent to game design whether for leisure or education.

Learners as game players inevitably start with failure – errors lead to learning – testing the rules of a game system like Civilization III, with tens of thousands of interacting variables results in a deeper level conceptual development. This learning cycle is critical to both intellectually engaging gameplay and academic learning, which illustrates the potential of educational games. Squire offers a warning, however, this testing of the system is in fact a test of the learner also. This act of failing or elicitation of rules as Johnson [8] might put it, was a critical precondition for learning as students found themselves confronting gaps or flaws in their current understandings through cycles of recursive play accommodating new information to old. Students who become comfortable with this very testing form of learning develop a deeper understanding of the subject in hand along with a range of high order cognitive skills. However, for other students the lack of certainty and the demands for intellectual rigour proved too onerous and the reading class became a safe and familiar haven.

There is strong evidence that games give students an opportunity to learn and that that learning is valuable as brain training and problem skills development and in some cases they promote conceptual learning. However, as all the theorists acknowledge that engagement is key, if players are not motivated to get started then the game will not promote learning.

4 Back in the Classroom

Educators have long emphasised the many benefits of games in children's learning processes [73] and the argument that those who have grown up in the era of rapidly advancing technology will prefer, even demand, to learn through games to mitigate against the monotony of traditional course materials has some merit [35], [74]. However, for digital technologies in general and video games in particular to be accepted in main stream education there must be some synergy with current educational trends and it must also show positive impacts on student learning in general and not simply specific skills.

One trend, which digital technologies resonate with, is the move to develop learning environments that focus on student-centred rather than teacher-centred learning activities. This is exemplified at a policy level by the UK Department of Education and Skills (now the DCFS) personalisation of the learning agenda. This policy seeks to reduce the persistent achievement gaps between different social and ethnic groups through an educational experience tailored education to each and every child and young person, that provides basic skills but also stretches their aspirations, and builds their life chances. [75]. To achieve these goals requires a shift in the social dynamics [76] with the catalyst for such change being more extensive and effective use of digital technologies.

Impact 2007, a UK national research project monitoring the effectiveness of the policy to personalise learning, has shown a clear link between technology resource and usage (eMaturity), increased personalisation of learning, individual pupil attributes on nationally standardised performance [77]. Sixty-seven schools, their heads and technology co-ordinators, 425 teachers and 2875 pupils contributed data through a series of on-line surveys, individual and group interviews and classroom observations [77]. While the impact of the move to personalise learning was the key factor under investigation, we found that school achievement, as measured by aggregated learner performance on standardised tests, was predicted by the eMaturity of the school and the level of individual learner's investment in their own learning. Using the dimensions of eMaturity and pupils' investment in their own learning we identified four groups of schools:

1. Low eMaturity / Low Investment in Learning (LE/LIL)
2. High eMaturity / Low Investment in Learning (HE/LIL)
3. Low eMaturity / High Investment in Learning (LE/HIL)
4. High eMaturity / High Investment in Learning (HE/LIL)

Figure 4 shows that students who invest in their own learning, that is self-regulate by setting personal learning targets, and who also record low levels of disaffection from school, unsurprisingly perform well in school. The interplay between the positive goal setting and negative disinfection, which computes a learner's investment in their own learning, is then a key predictor of performance on standardised national tests for children as young as 7 years and as old as 16 years of age. However, the performance of those investing in their own learning is enhanced if they are working in a technology mature environment (High e-Maturity / High Investment in Learning). This is clear evidence of the value of technology for learning but it should be noted that technology by itself does not lead to high performance. High eMaturity coupled with Low Investment in Learning (Fig. 4) can ameliorate the problem of poor performance but, as was the case in Squire's [66] Civilisation III intervention, without engagement of the learners technology can have limited impact.

The data from the Impact 2007 study led us to construct a simple but powerful model of effective learning, which is summarised in the Learning Equation (Fig. 5). In this model the educational institution provides opportunities for learners to learn using resources, which will include good teachers and technology. Although the potential value

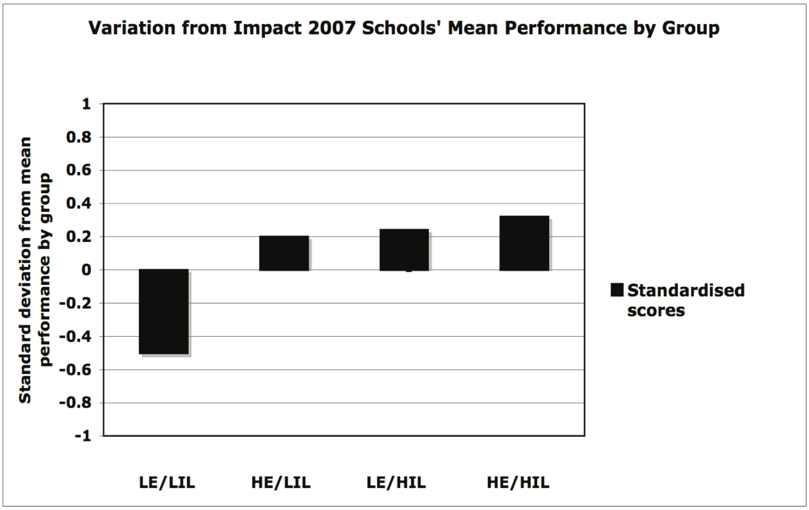


Fig. 4. The Impact of Levels of eMaturity and Pupil Investment in Learning on Whole School Performance: Deviation from Overall Mean Performance by Group [77]

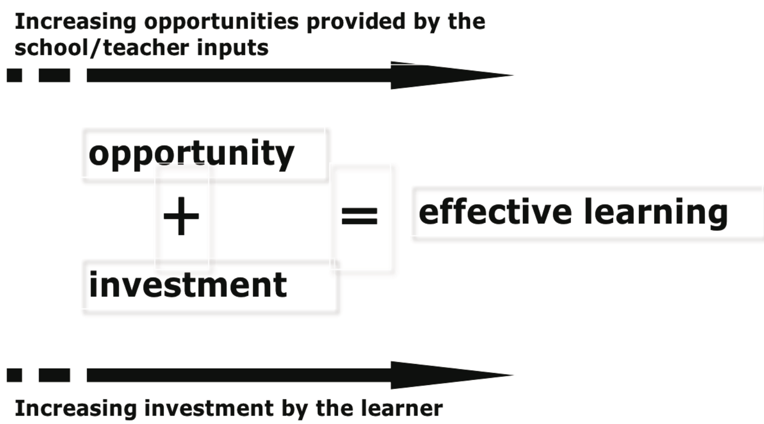


Fig. 5. The Learning Equation developed under Impact 2007 (37)

added by these resources is diminished unless the learner has both a desire to learn and the social and cognitive skills to take advantage of the opportunities on offer. In terms of digital media it must be attuned to learners' experience, expertise, previous knowledge but also their interest [78]. So our current data show that performance on national tests is predicted by learner engagement and by richness of the technology environment within the school. The robustness of this model is undergoing further tests at this time. Gaming as a resource has the advantage that it is motivating with the potential to engage learners. Games also resonate with the current views on what constitutes learning, that of brain training for a fast moving world rather than knowledge acquisition which has reduced value in a world where the Internet provides the external store and we are simply the CPU [19]. However, as Squires [71] has shown, some students' concept of what 'proper' school learning is varies from this brave new world. Many see the acquisition of formal knowledge, rather than the development of process skills, as the core educational goal. This coupled with the fear of failure experienced by many students makes the use of at least complex, although possibly not mini-games, a risky approach with benefits for some but not all of our students. I do not mean to devalue the acquisition of knowledge but it should not be the overriding goal of our educational system. So, we need to redefine what it is to be an educated person and release students from the comfort blanket of mere content. In such an educational milieu the full value of games as educational tools will emerge.

Acknowledgements

Grateful thanks go to Becta who funded the Impact 2007 and Personalising of Learning projects and also to all members of the teams working on those projects.

References

1. Southwell, B.G., Doyle, K.O.: The Good, the Bad, or the Ugly? A Multilevel Perspective on Electronic Game Effects. Am. Behav. Sci. 48, 391–401 (2004)
2. Palmer, S.: Toxic Childhood: How the Modern World is Damaging Our Children and What We Can Do About It. Orion Books, London (2006)
3. Cuban, L.: Oversold and Underused: Computers in the Classroom. Harvard University Press, Cambridge (2001)
4. Cuban, L., Kirkpatrick, H., Peck, C.: High Access and Low Use of Technologies in High School Classrooms: Explaining an Apparent Paradox. Am. Ed. Res J. 38, 813–834 (2001)
5. Oppenheimer, T.: The Flickering Mind: The False Promise of Technology in the Classroom and How Learning can be Saved. Random House, New York (2003)
6. Anderson, C.A.: Violent Video Games: Myths, Facts, and Unanswered Questions. Psych. Sci. Agenda (2003), `http://www.apa.org/science/sb-anderson.pdf`
7. Carnagey, N.L., Anderson, C.A.: The Effects of Reward and Punishment in Violent Video Games on Aggressive Affect, Cognition and Behaviour. Psych. Sci. 16, 882–889 (2005)
8. Johnson, S.: Everything Bad is Good for You: How Popular Culture is Making Us Smarter. Allen Lane/Penguin, London (2005)
9. Green, H., Hannon, C.: Their Space: Education for a Digital Generation. DEMOS, London (2007)
10. Underwood, J.D.M., Banyard, P.E., Davies, M.N.O.: Students in Digital Worlds: Lost in Sin City or Reaching Treasure Island? In: Br. J. Ed. Psych. Monograph Series II, vol. 5, pp. 83–99. British Psychological Society, Leicester (2007)
11. Shaffer, D.W., Gee, J.P.: Epistemic Games as Education for Innovation. In: Underwood, J.D.M., Dockrell, J. (eds.) Learning Through Digital Technologies, Br. J. Ed. Psych. Monograph Series II, vol. 5, pp. 71–82. British Psychological Society, Leicester (2007)
12. Royal Society of Arts.: Opening Minds (2008), `http://www.thersa.org/projects/education/opening-minds`
13. Freire, P., Macedo, D.: Literacy: Reading the Word and the World. Bergin Garvey, New York (1987)

14. Sternberg, R.J., Forsythe, G.B., Hedlund, J., Horvath, J.A., Wagner, R.K., Williams, W.M., Snook, S.A., Grigorenko, E.: Practical Intelligence in Everyday Life. Cambridge University Press, Cambridge (2000)
15. Sec. 802. National Center for Research in Advanced Information and Digital Technologies (2008), http://www.fas.org/press/_docs/Advanced%20Information%20and%20Digital%20Technologies.pdf
16. Caine, R.N., Caine, G.: Understanding a Brain Based Approach to Learning and Teaching. Ed. Leadership 48, 66–70 (1990)
17. Zull, J.E.: The Art of Changing the Brain: Enriching the Practice of Teaching by Exploring the Biology of Learning. Stylus Publishing Sterling, VA (2002), http://www.newhorizons.org/neuro/neu_review_zull.htm
18. Ausubel, D., Novak, J.D., Hanesian, H.: Educational Psychology: A Cognitive View. Holt, Rinehart Winston, New York (1978)
19. Merzenich, M.M.: On the Brain (2008), http://merzenich.positscience.com/
20. Doige, N.: The Brain that Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science, Penguin, New York (2007)
21. Diamond, M.C.: Response of the Brain to Enrichment. New Horizons for Learning (2001), http://www.newhorizons.org/neuro/diamond_brain_response.htm#a
22. Mayer, R.E.: Multimedia Learning. Cambridge University Press, New York (2001)
23. Mayer, R.E., Moreno, R.: Aids to Computer-Based Multimedia Learning. Learning and Instruction 12, 107–119 (2002)
24. Paivio, A.: Mental Representations: A Dual Coding Approach. Oxford University Press, Oxford (1986)
25. Rieber, L.P.: A Historical Review of Visualization in Human Cognition. Ed. l Tech. Res. and Dev. 3, 45–46 (1995)
26. Sweller, J.: Instructional Design in Technological Areas. ACER Press, Camberwell (1999)
27. Sweller, J.: The Worked Example Effect and Human Cognition. Learning and Instruction 16, 165–169 (2006)
28. Csikszentmihalyi, M.: Beyond Boredom and Anxiety: Experiencing Flow an Work and Play, Jossey-Bass, San Fransisco (1975)
29. Minsky, M.: Telepresence. Omni 2, 45–51 (1980)
30. Sadowski, W., Stanney, K.M.: Presence in virtual environments. In: Stanney, K.M. (ed.) Handbook of Virtual Environments- Design Implementations and Applications, pp. 791–806. Erlbaum, Mahwah (2002)
31. Kiili, K.: Evaluations of an Experiential Gaming Model. Human Tech. 2, 187–201 (2006)
32. Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., Groner, R.: Playing Online Games Against Computer- vs. Human-controlled Opponents: Effects on presence, Flow, and Enjoyment. Comp. In Human Behav. 24, 2274–2291 (2008)
33. Ericcson, K.A.: How the Expert Performance Approach Differs from Traditional Approaches in Sport. In: Starkes, J.L., Ericsson, K.A. (eds.) Expert Performance in Sport. Human Kinetics, Champagne, IL, pp. 342–372
34. Ebner, M., Holzinger, A.: Successful Implementation of User-Centered GameBased Learning in H. Ed. – an Example from Civil Engineering. Computers & Edn. 49, 873–890 (2007)
35. Prensky, M.: Students as Designers and Creators of Educational Computer Games: Who Else? Br. J. Ed. Tech (in press, 2008)
36. Eccles, J.S., et al.: Development, During Adolescence: The Impact of Stage Environment Fit on Young Adoloescents Experiences in Schools and Families. Am. Psych. 24, 90–101 (2003)

37. Poole, S.: Trigger Happy: The Inner Life of VideoGames (2004), http://polygonweb.online.fr/apoole.htm
38. Stanger, J.D., Gridina, N.: Media in the Home 1999: The Fourth Annual Survey of Parents and Children (Survey Series No. 5). Annenberg Public Policy Center of the U of Pennsylvania, Norwood, NJ (1999), http://www.appcpenn.org/pubs.htm
39. Underwood, J., Baguley, T., Banyard, P., Dillon, G., Farrington-Flint, L., Hayes, M., Hick, P., LeGeyt, G., Murphy, J., Selwood, I., Wright, M.: Personalising Learning. Becta, Coventry (2008)
40. Griffiths, M.D., Davies, M.N.O., Chappell, D.: Breaking the Stereotype: The Case of Online Gaming. Cyber. Psych. and Behav. 6, 81–91 (2003)
41. Fallows, D.: The Internet and Daily Life. Pew Internet & American Life Project, Washington, DC (2004)
42. ESA Essential Facts About the Computer and Video Game Industry (2005), http://www.theesa.com/files/2005EssentialFacts.pdf
43. Taylor, T.L.: Multiple Pleasures: Women and Online Gaming. Convergence: Int. Res. into New Media Tech. 9, 21–46 (2003)
44. Lawry, J., et al.: Exploring Common Conceptions About Boys and Electronic Games. J. Computers in Math. and Sci. Teaching 14, 439–459 (1995)
45. Sandford, R., Uiksak, M., Facer, K., Rudd, T.: Teaching with Games: Using Commercial Off-The-Shelf Computer Games in Formal Education. Nesta Futurelab, Bristol (2006)
46. Underwood, J.D.M.: Rethinking the Digital Divide: Impacts on Student-Tutor Relationships. European Journal of Education 42, 213–222 (2007)
47. Gee, J.P.: What Video Games have to Teach Us about Learning and Literacy. Palgrave Macmillan, London (2003)
48. Thomas, D., Brown, J.S.: Why Virtual Worlds Can Matter. Working Paper, October, 21 (2007), http://www.johnseelybrown.com/needvirtualworlds.pdf
49. Lave, J., Wenger, E.: Situated Learning: Legitimate Peripheral Participation. Cambridge University Press, Cambridge (1991)
50. Steinkuehler, C.A.: Learning in Massively Multiplayer Online Games. In: Kafai, Y.B., Sandoval, W.A., Enyedy, N., Nixon, A.S., Herrera, F. (eds.) Proceedings of the Sixth International Conference of the Learning Sciences, pp. 521–528. Erlbaum, Mahwah (2004)
51. Griffith, J.L., Voloschin, P., Gibb, G.D., Bailey, J.R.: Differences in Eye-Hand Motor Coordination of Videogame Users and Non-Users. Perceptual and Motor Skills 57, 155–158 (1983)
52. Drew, B., Waters, J.: Video games: Utilization of a Novel Strategy to Improve Perceptual Motor Skills and Cognitive Functioning in the Non-Institutionalized Elderly. Cog. Rehab. 4, 26–31 (1986)
53. Green, C.S., Bavelier, D.: Enumeration Versus Multiple Object Tracking: The Case of Action Video Game Players. Cognition 101, 217–245 (2006)
54. Sims, V.K., Mayer, R.E.: Domain Specificity of Spatial Expertise: The Case of Video Game Players. App. Cog. Psych. 16, 97–115 (2002)
55. Underwood, G., Everatt, J.: Automatic and Controlled Information Processing: The Role of Attention in the Processing of Novelty. In: Neumann, O., Sanders, A.F. (eds.) Handbook of Perception and Action. Attention, vol. 3, pp. 185–227. Academic Press, London (1996)
56. Underwood, J.D.M.: Novice and Expert Performance with a Dynamic Control Task: Scanpaths During a Computer Game. In: Underwood, G. (ed.) Cognitive Processes in Eye-guidance, pp. 30–324. Oxford University Press, Oxford (2005)
57. Demaine, E.D., Hohenberger, S., Liben-Nowell, D.: Tetris is Hard, Even to Approximate. In: Proceedings of the 9th International Computing and Combinatorics Conference, Cocoon 2003, Big Sky, Montana, July 25-28, pp. 351–363 (2003)

58. DiSessa, A.A.: Changing minds: Computers, Learning and Literacy. MIT Press, Cambridge (2000)
59. Haier, R.J., Siegel, B.V., MacLachlan, A., Soderling, E., Lottenberg, S., Buchsbaum, M.S.: Regional Glucose Metabolic Changes After Learning Complex Visuospatial/Motor Task. Brain Research 570, 134–143 (1992)
60. Mayer, R.E., Wittrock, M.C.: Problem-solving Transfer. In: Berliner, D., Calfee, R. (eds.) Handbook of Educational Psychology. Macmillan, New York (1996)
61. Shepard, R.N., Cooper, L.A.: Mental Images and their Transformations. MIT Press, Cambridge (1982)
62. Carroll, J.B.: Human Cognitive Abilities. Cambridge University Press, Cambridge (1993)
63. Okagaki, L., Frensch, P.A.: Effects of Video Game Playing on Measures of Spatial Performance: Gender Effects in Late Adolescence. J. App. Dev. Psych. 15, 33–58 (1994)
64. Subrahmanyam, K., Greenfield, P.M.: Effect of Video Game Practice on Spatial Skills in Girls and Boys. J. App. Dev. Psych. 15, 13–32 (1994)
65. Gagnon, D.: Videogame and spatial skills: an exploratory study. Ed. Comm. and Tech. J. 33, 263–275 (1985)
66. Dorval, M., Pepin, M.: Effect of Playing a Video Game on a Measure of Spatial Visualization. Percept. Motor Skills 62, 159–162 (1986)
67. Kirsh, D.P.M.: On Distinguishing Epistemic from Pragmatic Actions. Cog. Sci. 18, 513–549 (1994)
68. Grant, E.R., Spivey, M.J.: Eye-movements and Problem Solving: Guiding Attention Guides Thought. Psych. Sci. 14, 462–466 (2003)
69. Rosser, J.C., Lynch, P.D., Cuddihy, L., Gentile, D.A., Klonsky, J., Merrell, R.: The Impact of Video Games on Training Surgeons in the 21st Century. Arch Surg. 142, 181–186 (2007)
70. Telner, J.A., Wiesenthal, D.L., Bialystok, E., York, M.: Benefits of Bilingualism and Video Game Proficiency when Driving and Speaking on a Cellular Telephone. Paper read at ICTTP, Washington D.C (2008)
71. Squire, K.D.: Changing the Game: What Happens When Video Games Enter the Classroom. Innovate 1 (2005), http://www.innovateonline.info/index.php?view=article&id=82%20
72. Squire, K.D.: From content to context: Video games as designed experiences. Ed. Researcher 35, 19–29 (2006)
73. Malone, T.W., Lepper, M.R.: Making Learning Fun: A Taxonomy of Intrinsic Motivations for Learning. In: Snow, R.E., Farr, M.J. (eds.) Aptitude, Learning and Instruction: III. Conative and Affective Process Analyses, pp. 223–253. Erlbaum, Hillsdale (1987)
74. Dondi, C., Moretti, M.: A methodological proposal for learning games selection and quality assessment. BJET 38, 502–512 (2007)
75. Department for Education and Skills.: The Primary National Strategy: Personalisation. DFES. London (2006), http://www.standards.dfes.gov.uk/local/ePDs/leading_on_intervention/site/u1/s3/index.html
76. Pollard, A., James, M.: Personalised Learning A Commentary by the Teaching and Learning Research Programme. ESRC, Swindon (2004)
77. Underwood, J., Baguley, T., Banyard, P., Coyne, E., Farrington-Flint, L., Selwood, I.: Impact 2007: Personalising Learning with Technology. Becta, Coventry (2007), http://partners.becta.org.uk/upload-dir/downloads/page_documents/research/impact_July2007.doc
78. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Ed.l Tech. & Soc. 11, 279–290 (2008)

Challenges in the Development and Evaluation of Immersive Digital Educational Games

Effie Lai-Chong Law[1], Michael D. Kickmeier-Rust[2], Dietrich Albert[2], and Andreas Holzinger[3]

[1] University of Leicester, LE1 7RH, Leicester, UK
Department of Computer Science
Elaw@mcs.le.ac.uk
[2] Graz University, A-8010 Graz, Austria
Department of Psychology
Cognitive Science Section
michael.kickmeier@uni-graz.at,
dietrich.albert@uni-graz.at
[3] Medical University Graz, A-8036 Graz, Austria
Institute for Medical Informatics, Statistics & Documentation (IMI)
Research Unit HCI4MED
andreas.holzinger@meduni-graz.at

Abstract. In this paper, we describe a conceptual framework and address the related issues and solutions in the identification of three major challenges for the development and evaluation of Immersive Digital Educational Games (IDEGs). These challenges are (i) advancing adaptive educational technologies to shape learning experience, ensuring the individualization of learning experiences, adaptation to personal aims, needs, abilities and prerequisites; (ii) providing technological approaches to reduce the development costs for IDEGs; by enabling the creation of entirely different stories and games for a variety of different learning domains, each based, more or less, on the same pool of story units; patterns and structures; (iii) developing robust evaluation methodologies for IDEGs by the extension of ISO 9241 to include user satisfaction, motivation and learning progress and other User Experience (UX) attributes. While our research and development is by no means concluded, we believe that we have arrived at a stage where conclusions may be drawn, which will be of considerable use to other researchers in this domain.

Keywords: Immersive digital educational games, Marco adaptivity, Micro adaptivity, User experience, Reusability, User Experience.

1 Introduction and Motivation for Research

Developing Digital Educational Games (DEGs) that can cost-effectively foster learning has been a challenge for researchers and practitioners in the fields of HCI& UE and Technology Enhanced Learning (TEL) for a long time. DEGs offer exciting and dynamic environments, which engage players in meaningful and motivating learning

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 19–30, 2008.

activities, inspiring them, through fun and pleasure, to explore a variety of topics and tasks.

The characteristics of the simulations within DEGs can contribute substantially to individual player's ability to construct knowledge, while the games' social aspects can enhance players' collaborative learning skills; however, in order to realize this vision, three major challenges must still be dealt with:

(i) The large degree of freedom, enabled by digital game environments, renders it extremely difficult to tailor the games to the end users' personal learning experiences and preferences, and to provide end users with purposeful and unobtrusive advice;

(ii) The relatively high competitiveness of the DEGs commercial counterparts, which attract a huge amount of investment and corresponding high quality production has raised the expectations of today's experienced players;

(iii) The difficulty in obtaining support from parents, educators and policymakers for the incorporation of to incorporate DEGs into regular curricula is mostly due to a lack of well established approaches towards yielding persuasive evidence about the educational efficacy of DEGs.

Our suggestions include the improvement of the development methods to ensure intelligent and adaptive educational technologies; to identify technological approaches to reduce the development costs of DEGs; and at the same to classify innovative and robust evaluation methodologies to validate pedagogical models and technological solutions for DEGs. The newly launched R&D project – 80Days (www.eightydays.eu) – aims to address these challenges.

Specifically, our aim is to tackle the problem of *reducing the costs* of developing methods of IDEGs by providing a methodological framework, including technological advancements; adaptive and interactive digital storytelling and *generic adaptive tutoring methodology* (cf. Competence-based Knowledge Space Theory (CbKST), [1]) and by realizing different scenarios within the same base game.

The scenarios implemented in our project are inspired by Jules Verne's appealing novel "Around the World in Eighty Days", which inspired our project's name. The primary topic of the game demonstrator is the discipline geography. The rationale for selecting this discipline is threefold: First, a broad range of sub-topics and knowledge domains provide well-defined internal structures in terms of prerequisites among specific competencies. Second, available curricula ranges from primary education to university level, enabling scalability. Third, attractive 2D and 3D learning resources (e.g. cartographic material, satellite images, etc.) can be perfectly utilized for immersive games with motivating narratives. In this paper, we present our conceptual frameworks and address the related issues and possible solutions.

2 Background and Related Work

Due to the common view that immersive DEGs can make learning engaging, inspiring and presumably effective, enthusiasms and efforts over game-based learning have soared in recent years in various application domains [2], [3], [4], [5], [6], [7].

Many researchers argue that playing computer games provides learners with a mental workout [8], i.e. that the *structure* of activities within computer games can develop cognitive skills due to the fact that end users are faced with decision making, consequently they must plan problem solving strategies in advance, which involve the monitoring of a series of tasks and sub-tasks, so called *Judgement-Behaviour-Feedback* loops [9].

Some of the main strengths of DEGs include [10]:

1) A high level of intrinsic motivation to play and to proceed in the game;
2) clear goals and rules;
3) a meaningful yet rich and appealing learning context;
3) an engaging storyline with random elements of surprise;
4) immediate feedback;
5) a high level of interactivity, challenge and competition.

These characteristics are in line with *Merrill's model for successful learning* [11]. Key factors of this model include *motivation* and *incidental learning* [12], [13], [5], however, memorable educational experiences should not only be enriching but also enjoyable [14], [15], [7].

Nevertheless, DEGs have some drawbacks such as difficulties in providing an appropriate balance between playing and learning activities or between challenge and ability; in aligning the game with national curricula and in affording extensive costs of developing high quality games. Also the lack of sound instructional models, based on pedagogical standards and didactical methods, is seen as the common weaknesses of most educational games, leading to a separation of learning from playing.

Despite these negative issues, commercial computer games are tremendously successful and the game industry constantly increases sales to several billion Euros. A large number of people of all age groups, especially secondary school ages, spend many hours a week playing games. Our observations corroborate the presumption that utilizing gaming activities for educational purposes and exploiting the educational potential of computer games is a highly promising approach to facilitating learning by making it enjoyable and pleasant [2].

The current project *80Days* is based upon the results of its predecessor *Enhanced Learning Experience and Knowledge TRAnsfer – ELEKTRA* [16], which made significant contributions to advancing the state-of-the-art of competitive DEGs in terms of educational game design, integration of pedagogical models and taxonomies, and the possibility of personalization by the use of adaptive technology.

Pedagogically, 80Days is grounded in the framework of *Self Regulated Personalized Learning* (SRPL) [17], [18] which propagates the importance of *Self Regulated Learning* (SRL) through meaningful choice and exploration, reflection and self-personalization in the learning process.

Self regulation [19], [20] can be defined as including an interactive *process* involving both cognitive self-regulation and motivational self-regulation; wherein cognitive self regulation can be taught and that students who use these self-regulatory skills obtain better grades in the content domain to which these skills apply [21].

However, it is argued that self-regulated learning can be domain-specific or domain-transcending and that competent performers in a specific domain rely on different types of previous knowledge related to that domain, consequently to address the previous knowledge is always an important issue [22]. This contributes towards the creation and *sustainability* of intrinsic motivation, which is a key factor of effective game-based learning.

3 The First Challenge: Adaptive Learning Technologies

Individualization of learning experiences, including adaptation to personal aims, needs, abilities and prerequisites, entails in-depth understanding of individual learners and their behaviour with a IDEG. It is critical not to destroy the immersion and gaming experience by intervening knowledge assessments, which are commonly used in traditional approaches to adaptive educational technologies that focus on knowledge and learning progress. IDEGs take into account a broader scope of issues such as individual preferences (e.g. visual styles or gaming genre). Prevailing cognitive models for adaptive educational technologies (which are primarily competence-based) should thus be merged with theories of achievement motivation and models of interactive and adaptive storytelling to establish a comprehensive theoretical framework for combining learning and gaming.

Techniques of adaptation and individualization are essentially adaptive presentation, adaptive navigation support and adaptive problem solving. In the framework of ELEKTRA [16] a new terminology was introduced, basically because of the reason that game-based approaches to learning are substantially different from traditional approaches to e-learning. The new concepts, which are tailored to learning environments with large degrees of freedom, are adaptive on macro and micro levels ([23]; Figure 1).

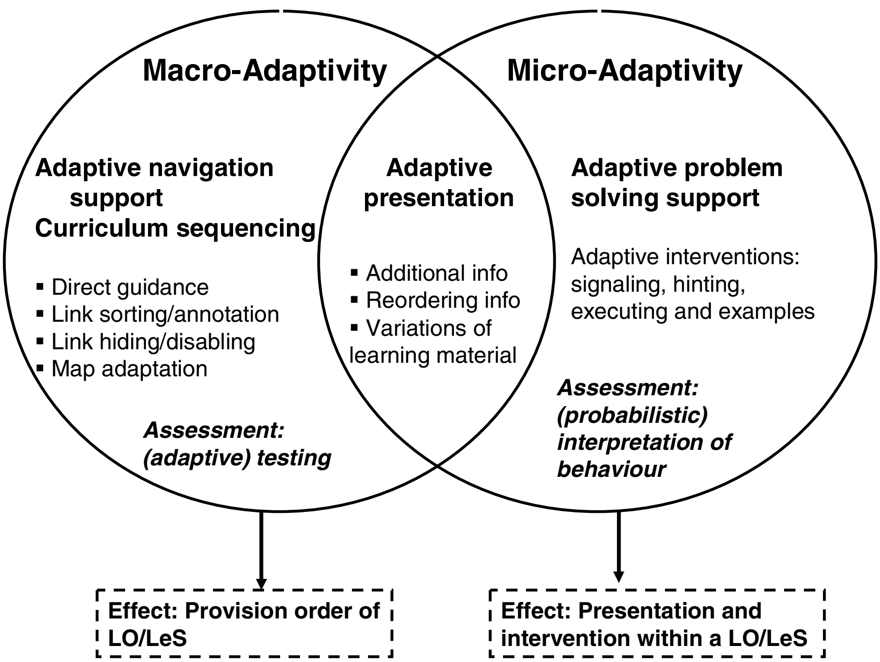


Fig. 1. Macro-adaptive and micro-adaptive techniques (NB: LO = Learning Object; LeS = Learning Situation)

Macro-adaptivity refers to traditional techniques of adaptation, such as adaptive presentation and adaptive navigation on the level of learning objects (LOs) or learning situations (LeSs) in a IDEG [24]. Generally, macro-adaptive interventions are based on a stable learner model (e.g., traits) or an adaptation model (e.g., pedagogical implications) and on typical (knowledge) assessments (e.g. via test items).

Micro-adaptivity or micro adaptive interventions are non-invasive, meaning that an overall narrative is not compromised and only affects the presentation of a specific LO or LeS.

In contrast to existing approaches, which separate learner, domain and adaptation models, this framework is based on a holistic understanding as well as a formal ontological representation of interacting processes involved in active and dynamic learning processes (Figure 2).

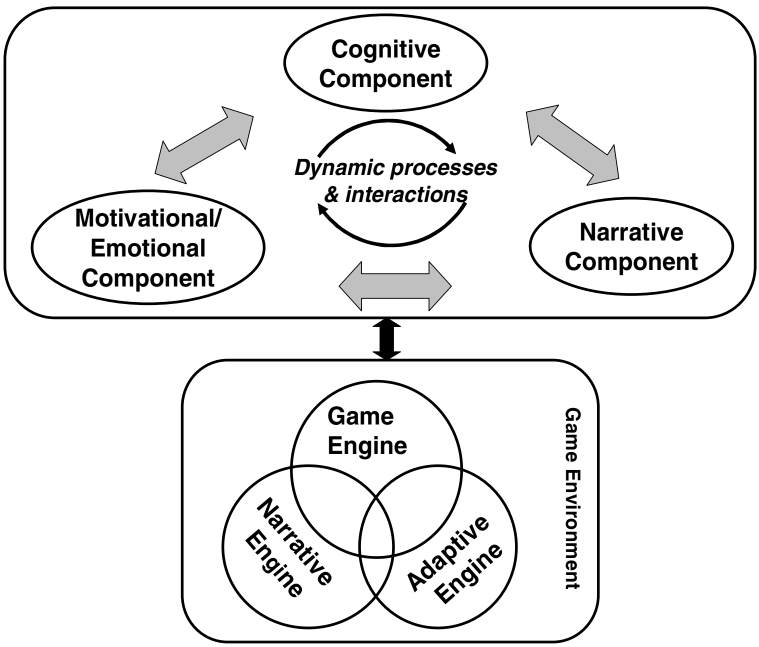


Fig. 2. Macro- and micro-adaptive techniques (NB: LO = Learning Object; LeS = Learning Situation)

4 The Second Challenge: Reusability of Learning Resources

In an IDEG, adaptive and interactive digital storytelling serves two essential purposes: First, it strongly supports a personalized learning experience by adapting the game's story to individual preferences and by providing the possibility of explorative learning processes. Second, it serves the re-usability of learning material by enabling the creation of entirely different stories and games for a variety of different learning domains, each based, more or less, on the same pool of story units; patterns and structures; as well as learning and playing concepts and elements/objects. Furthermore, reusing learning material contributes to minimizing costs, particularly with regard to the design of documents.

The development of competitive DEGs is cost-intensive and, when the topic relates to a limited age group or specific curricula, there is a narrow, albeit specialised, market for the end product.

Thus, the integration of existing learning resources is a crucial aspect of efficient and cost-effective learning design and game development. The integration of external resources (e.g., learning media; websites; web services; 3D material or cartographic material) with a game engine into a coherent and immersive game environment is difficult.

Since the development and application of immersive DEGs is still at an early stage, no appropriate methodologies exist to date, which would enable an effective integration of existing learning resources and their (re)use in DEGs.

Hence, it is deemed critical to analyze the technological and didactic demands and mutual dependencies between learning resources, learning activities, pedagogical models, and narrative/game engines. An approach of resource harmonization; resource symbolization and ontological resource description should also be established.

5 The Third Challenge: Evaluation Framework

Given the complexity of digital game environments, mixed-method evaluations are deemed necessary, which will enable the triangulation of multi-source and multi-perspective quantitative and qualitative data.

Given the complexity of digital game environments, it is deemed necessary to use mixed-method evaluations, which enable the triangulation of multi-source and multi-perspective quantitative and qualitative data.

Essentially, evaluation and validation activities comprise two nested aspects: the software quality of the game environment and the effectiveness of IDEGs in reaching their objectives of facilitating motivated, engaged, and successful learning.

Usability evaluation of digital games can employ a combination of empirical, analytic, and model-based methods [7]. While it is important to measure effectiveness and efficiency (ISO 9241) in terms of task completion rate and time, in the case of IDEGS, we find it extremely important to consider user satisfaction, motivation and learning progress that subsumes a range of User Experience (UX) attributes.

There have been recent attempts to provide theories and frameworks for UX but a unified view on UX is still lacking [25], [26].

Past research on UX acknowledged the demand for a theoretical model to put together cumulative knowledge in research. On the example of an experimentally tested and extended Hassenzahl's model of aesthetic experience, it was demonstrated that beauty was influenced by hedonic attributes (identification and stimulation), but quality by both hedonic and pragmatic (user-perceived usability) attributes as well as task performance and mental effort [27].

Hedonic quality can be measured, for example, by the application of the AttrakDiff questionnaire [28], [29], [30]. This questionnaire is embedded in a theoretical model, which differentiates between the pragmatic quality and the hedonic quality of a user interface [31].

Pragmatic quality describes traditional usability aspects, including efficiency, effectiveness and learnability. It focuses primarily on task related design aspects. Consequently, hedonic quality describes quality aspects, which are not directly related to the tasks the end users want to accomplish with the software, for example originality and beauty [30], [27].

Both qualities are subjective; consequently end users may differ in their evaluation of these aspects.

Moreover, it is deemed important to evaluate the *educational effectiveness* of the digital game environment in specific learning situations and broader learning domains with high conceptual and social content, and to adapt and elaborate existing evaluation methods, considering their potentials and limits (cf. [32]. Since a digital game is a *dynamic* learning tool, the evaluation must be tightly coupled with the actual learning process (i.e., constructivist-situated learning) [33].

It is also indispensable to apply multiple measures of learning and performance along multiple dimensions: technical; orientational; affective; cognitive; pedagogical; social and others.

An evaluation framework, which addresses conceptual; practical and methodological challenges in evaluating IDEGs – an immature and exciting field –, must take the following aspects into account:

(i) Extensibility of conventional usability engineering methods for evaluating both the usability *and* user experience of interacting with IDEGs;

(ii) Multi-dimensionality of the educational efficacy of digital games, including knowledge acquisition, meta-cognition, and social interaction;

(iii) Evaluation criteria for different stakeholders' multiple goals or values (personal, social and economic) to be fulfilled by the gaming environment;

5.1 Extensibility of Existing UEMs

Two types of Usability Engineering Methods (UEM, see [34] for an overview) – heuristic evaluation and thinking-aloud user testing – are widely deployed evaluating the usability of IT products (e.g. [35], [36]) including digital games [37], [7].

Specifically, there has been some recent research on using heuristics to evaluate the *playability* of an entertainment technology such as video games [38], [39], [40], [41]. However, these discount methods, without involving actual users, can only provide an overview of usability and playability of such a game. As the value of a game is primarily determined by end-users' subjective perceptions, assessing through the lens of a user surrogate (i.e. a usability specialist who ideally should also be a game player – a double expert) to review the usability/UX of the game seems inadequate.

Some researchers are skeptical about the applicability of heuristic evaluation for evaluating digital games [42], leaving user-based evaluation as "the" viable option.

Presumably, concurrent think-aloud techniques can yield valid empirical data that will enable researchers to understand users' perceptions of the system in question [43], [44]. However, it is considered detrimental to the flow experience [45] of players if they are asked to verbalize whilst engaging in the game; cognitive loading, as in other non-game situations, is also a compelling concern.

Retrospective thinking-aloud methods [43], by asking the player to interpret video-recordings post-hoc, seem more appropriate. There also exist some schemes for coding users' emotional behaviours (e.g. [46]).

However, the inherent problem of retrospective reporting is memory reconstruction, which is especially acute in a playing situation because players' emotional responses normally change very rapidly. Hence, they may not be able to recall or interpret what actually triggered their specific emotion at a specific time. While such moment-to-moment, or situated feedback, may be relevant for identifying usability problems of a game design, some game developers query, from the pragmatic point of view, why bothers to capture data at such a fine-grained level. Their argument is: If players have an overall positive experience, which overrides its negative counterpart, and are thus willing to play it again in the future, then the game can be seen as a success.

While this argument may be valid, to some extent, for entertainment technologies, it is insufficient to evaluate a IDEG, since it fails to take its educational value into account.

Other retrospective self-reporting methods, such as questionnaire and differential semantic, are commonly employed for evaluating games. However, there exist only a few standardized questionnaires for measuring UX with established psychometric properties (e.g. [6]).

Apart from the issue of standardizing inventories, the timing of administering a questionnaire is also critical because UX fluctuate in the course of system use. It boils down to the above mentioned issue - situated vs. average UX, and to also the basic question: Why do we evaluate?

With the aim to complement, as well as supplement, subjective self-reported data, there is an increasing trend towards employing physiological measures for assessing UX in games [42], such as eye-tracking; galvanic skin response; electrocardiography; electromyography of the face and heart rate [47], [48].

While applying these advanced measuring techniques in HCI is not new, the major issues of reliably interpreting or calibrating the related measures remain. For instance, prolonged viewing at a special location of a user interface, as indicated by certain patterns on a heat map, may imply that the user is cognitively absorbed by the content or that they are simply struggling to figure it out. Similarly, sweaty foreheads/hands may imply negative restlessness or positive excitement; individual and socio-cultural differences in facial expressions make interpreting electromyography data difficult; or a deviation of heart rate can also indicate emotional changes.

Furthermore, while automatic logs enable researchers to capture a huge body of data, the effective and efficient integration of these data, to produce coherent and convincing conclusions about players' behaviour and experience, is still very challenging, despite the advance in data mining techniques. To complicate the picture further, the low ecological validity of lab-based tests, albeit neat and structured, demands their replacement with field tests. Evaluating the usability/UX of IDEGs entails the adaptation and augmentation of existing approaches to address the tradeoffs: retrospective vs. real-time data capture, subjective vs. objective data types, and lab-based vs. field studies.

5.2 Assessing Educational Efficacy

Based on the results of previous research and guidelines for measuring the "success" of technology-based learning material and environments, the project 80Days aims to develop a scientifically sound methodology, grounded in criterion-based designs and, for specific cases, comparison-based designs. Criterion-based designs utilize a-priori specified criteria to measure educational effectiveness, e.g., whether students learn what they were supposed to learn. Comparison-based designs can potentially be applied to evaluating and validating adaptive sequencing of learning material; adaptive story generation, and non-invasive assessment and interventions. Three significant factors [50] to be included in our evaluation framework are:

a) Performance outcomes (e.g., time and efforts required for achieving educational objectives)
b) Attitude outcomes (e.g., attitude towards the learning environment/media; motivation and interest)
c) Programmatic outcomes (e.g., is there a return on investment? Does the learning environment reach its target audience?)

Therefore, while there is a large body of existing methods and research for the evaluation of software, as well as conventional educational technology and learning media, this knowledge cannot be transformed one-to-one to the genre of IDEGs (e.g., [6]. The investigation of the adaptive features in a narrative learning environment with a large degree of freedom challenges the existing evaluation and validation methods. Moreover, aspects of motivation, audio-visual preferences, personality, or individual ability and their interactions with game design, learning design, usability, or narrative are a complex and novel field of research.

The ambition of 80Days is to fill the gaps thus identified and develop a reference framework for the in-depth evaluation of IDEGs at design time (accompanying the design and development stages) and at run time (evaluating interim and final versions of games).

6 Conclusion and Future Outlook

Exploiting up-to-date 3D computer games for educational purposes is a dawning technology but still in its fledgling stage. It is one of the challenges of 80Days, which also aims at augmenting and integrating the related theoretical frameworks in cognitive psychology. Exploiting the desirable features of digital games, to design and develop effective learning tools, entails the well-orchestrated efforts of a highly interdisciplinary team to tackle the three major challenges addressed above.

In contrast to existing approaches, which separate learner, domain and adaptation models, our framework is based on a holistic understanding and a formal ontological representation of interacting processes involved in active and dynamic learning processes. The measurements include performance outcomes (e.g., time and efforts required for achieving educational objectives), attitude outcomes (e.g., attitude towards the learning environment/media; motivation and interest, and Programmatic outcomes (e.g., is there a return on investment? Does the learning environment reach its target audience?).

The range of applications for this technology is not limited to a specific age group, it also provides a special opportunity to use the educational games to familiarize the elderly, who lack previous exposure to technology, to modern live saving communication technology.

Acknowledgements

The research and development introduced in this work is co-funded by the European Commission under the sixth framework programme in the IST research priority, contract number 027986 (ELEKTRA, www.elektra-project.org) as well as under the seventh framework programme in the ICT research priority, contract number 215918 (80Days, www.eightydays.eu).

References

1. Albert, D., Lukas, J.: Knowledge spaces: theories, empirical research, and applications. Lawrence Erlbaum Associates, Mahwah (1999)
2. Holzinger, A., Pichler, A., Almer, W., Maurer, H.: TRIANGLE: A Multi-Media test-bed for examining incidental learning, motivation and the Tamagotchi-Effect within a Game-Show like Computer Based Learning Module Educational Multimedia, Hypermedia and Telecommunication 2001, Association for the Advancement of Computing in Education, Charlottesville (VA), pp. 766–771 (2001)
3. Mann, B.D., Eidelson, B.M., Fukuchi, S.G., Nissman, S.A., Robertson, S., Jardines, L.: The development of an interactive game-based tool for learning surgical management algorithms via computer. The American Journal of Surgery 183(3), 305–308 (2002)
4. Cai, Y., Snel, I., Bharathi, B.S., Klein, C., Klein-Seetharaman, J.: Towards biomedical problem solving in a game environment. In: Sloot, P.M.A., Abramson, D., Bogdanov, A.V., Gorbachev, Y.E., Dongarra, J., Zomaya, A.Y. (eds.) ICCS 2003. LNCS, vol. 2659, pp. 1005–1014. Springer, Heidelberg (2003)
5. Holzinger, A., Pichler, A., Maurer, H.: Multi Media e-Learning Software TRIANGLE Case-Study: Experimental Results and Lessons Learned. Journal of Universal Science and Technology of Learning 0(0), 61–92 (2006), http://www.justl.org/justl_0_0/multi_media_elearning_software/justl_0_0_0061_0092_holzinger.pdf
6. de Freitas, S., Oliver, M.: How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? Computers & Education 46(3), 249–264 (2006)
7. Ebner, M., Holzinger, A.: Successful Implementation of User-Centered Game Based Learning in Higher Education – an Example from Civil Engineering. Computers & Education 49(3), 873–890 (2007)
8. Robertson, J., Howells, C.: Computer game design: Opportunities for successful learning. Computers & Education 50(2), 559–578 (2008)
9. Garris, R., Ahlers, R., Driskell, J.E.: Games, Motivation, and Learning: A Research and Practice Model. Simulation & Gaming 33(4), 441–467 (2002)
10. Prensky, M.: Digital game-based Learning. McGraw Hill, New York (2001)
11. Merrill, M.D.: First principles of instruction. Etr&D-Educational Technology Research and Development 50(3), 43–59 (2002)
12. Cordova, D.I., Lepper, M.R.: Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. Journal of Educational Psychology 88(4), 715–730 (1996)
13. Holzinger, A.: Computer-aided Mathematics Instruction with Mathematica 3.0. Mathematica in Education and Research 6(4), 37–40 (1997)
14. Shneiderman, B.: Relate-Create-Donate: a teaching/learning philosophy for the cyber-generation. Computers & Education 31(1), 25–39 (1998)

15. van Reekum, C.M., Johnstone, T., Banse, R., Etter, A., Wehrle, T., Scherer, K.R.: Psychophysiological responses to appraisal dimensions in a computer game. Cognition & Emotion 18(5), 663–688 (2004)
16. ELEKTRA: ELEKTRA - Enhanced Learning Experience and Knowledge TRAnsfer (last access: 2008-09-06), `http://www.elektra-project.org`
17. Aviram, R., Winer, A., Somech, S., Schellas, Y., Dotan, I., Ronen, Y.: iClass Methodologies: Intelligent Distributed Cognitive-based Open Learning System for Schools (last access: 2008-09-06), `http://www.iclass.info/docs/iClass%20Methodologies.pdf`
18. Wouters, P., Tabbers, H.K., Paas, F.: Interactivity in video-based models. Educational Psychology Review 19(3), 327–342 (2007)
19. Entwistle, N., McCune, V.: The conceptual bases of study strategy inventories. Educational Psychology Review 16(4), 325–345 (2004)
20. Caprara, G.V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G.M., Barbaranelli, C., Bandura, A.: Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. Journal of Educational Psychology 100(3), 525–534 (2008)
21. Boekaerts, M.: Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. Learning and Instruction 7(2), 161–186 (1997)
22. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
23. Kickmeier-Rust, M., Albert, D.: The ELEKTRA ontology model: A learner-centered approach to resource description. LNCS. Springer, Heidelberg (in press)
24. Holzinger, A., Nischelwitzer, A.: Chameleon Learning Objects: Quasi-Intelligente doppelt adaptierende Lernobjekte: Vom Technologiemodell zum Lernmodell. OCG Journal 30(4), 4–6 (2005)
25. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behaviour & Information Technology 25(2), 91–97 (2006)
26. Law, E., Hvannberg, E.T., Hassenzahl, M.: Proceedings of the Workshop Towards a Unified View of UX (last access: 2008-09-10), `http://www.cost294.org/`
27. van Schaik, P., Ling, J.: Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. Interacting with Computers 20(3), 419–432 (2008)
28. Hassenzahl, M., Beu, A., Burmester, M.: Engineering joy. IEEE Software 18(1), 70 (2001)
29. Harbich, S., Auer, S.: Rater bias: The influence of hedonic quality on usability questionnaires. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1129–1133. Springer, Heidelberg (2005)
30. Schrepp, M., Held, T., Laugwitz, B.: The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers 18(5), 1055–1069 (2006)
31. Hassenzahl, M., Platz, A., Burmester, M., Lehner, K.: Hedonic and ergonomic quality aspects determine a software's appeal. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 201–208 (2000)
32. Wideman, H.H., Owston, R.D., Brown, C., Kushniruk, A., Ho, F., Pitts, K.C.: Unpacking the potential of educational gaming. Simulation & Gaming 38(1), 10–30 (2007)
33. Holzinger, A.: Multimedia Basics, Laxmi, New Delhi. Learning. Cognitive Fundamentals of multimedial Information Systems, vol. 2 (2002), `http://www.basiswissen-multimedia.at`

34. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
35. Stephanie, R., Janice Anne, R., Judee, H.: A toolkit for strategic usability: results from workshops, panels, and surveys. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 337–344 (2000)
36. Holzinger, A.: Application of Rapid Prototyping to the User Interface Development for a Virtual Medical Campus. IEEE Software 21(1), 92–99 (2004)
37. Bernhaupt, R., Eckschlager, M., Tscheligi, M.: Methods for evaluating games: how to measure usability and user experience in games? In: Proceedings of the international conference on Advances in computer entertainment technology (2007)
38. Fabricatore, C., Nussbaum, M., Rosas, R.: Playability in action videogames: A qualitative design model. Human-Computer Interaction 17(4), 311–368 (2002)
39. Desurvire, H., Caplan, M., Toth, J.A.: Using heuristics to evaluate the playability of games. In: CHI 2004, pp. 1509–1512 (2004)
40. Korhonen, H., Koivisto, E.M.I.: Playability Heuristics for Mobile Multi-Player Games. In: DIMEA 2007, Second International Conference on Digital Interactive Media in Entertainment and Arts (2007)
41. Röcker, C., Haar, M.: Exploring the usability of videogame heuristics for pervasive game development in smart home enviroments. In: Third International Workshop on Pervasive Game Applications - PerGames 2006, pp. 199–206 (2006)
42. Mandryk, R.L., Atkins, M.S.: A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International Journal of Human-Computer Studies 65(4), 329–347 (2007)
43. van den Haak, M., De Jong, M., Schellens, P.J.: Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & Information Technology 22(5), 339–351 (2003)
44. Van den Haak, M.J., de Jong, M.D.T., Schellens, P.J.: Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: a methodological comparison. Interacting with Computers 16(6), 1153–1170 (2004)
45. Csíkszentmihályi, M.: Finding flow: The Psychology of engagement with everyday life. Basic Books, London (1998)
46. de Lera, E., Garreta-Domingo, M.: Ten emotion heuristics: Guidelines for assessing the user's affective dimension easily and cost effectively. In: Proceedings of HCI 2007 The 21st British HCI Group Annual Conference, University of Lancaster, September 3 - 7 (2007)
47. Stickel, C., Fink, J., Holzinger, A.: Enhancing Universal Access – EEG based Learnability Assessment. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 813–822. Springer, Heidelberg (2007)
48. Stickel, C., Holzinger, A., Ebner, M.: Useful Oblivion Versus Information Overload in e-Learning Examples in the Context of Wiki Systems. In: Luzar-Stiffler, V., Dobric, V.H., Bekic, Z. (eds.) Proceedings of the ITI 2008 30th International Conference on Information Technology Interface, June 23-26, pp. 171–176 (2008)
49. Bernhaupt, R., IJsselsteij, W., Mueller, F., Tscheligi, M., Wixon, D.R.: Evaluating user experience games CHI 2008 extended abstracts on Human factors in computing systems, pp. 3905–3908 (2008)
50. Lockee, B., Moore, M., Burton, J.: Measuring success: Evaluation strategies for distance education. Educause Quaterly 25(1), 20–26 (2002)

Evaluating the Motivational Value of an Augmented Reality System for Learning Chemistry

Costin Pribeanu and Dragoş Daniel Iordache

ICI Bucharest, Bd. Mareşal Averescu No. 8-10
Bucureşti, Romania
iordache@ici.ro, pribeanu@ici.ro

Abstract. The development of Augmented Reality (AR) technologies is creating new opportunities and challenges for the designers of e-learning systems. More recently, AR based educational systems have been developed that are implementing new learning scenarios for primary and secondary schools. An important goal of these novel teaching platforms is to enhance the students' motivation to learn. This paper reports on a user-centered formative usability evaluation of an AR-based learning scenario for Chemistry with a focus on the motivational value. After user testing, quantitative and qualitative data were collected that are measuring the educational and motivational values of the learning scenario. While the comparison between the two types of measure is increasing the confidence in the evaluation results, the qualitative measures are also providing with a detailed description of the user learning experience. In this respect, not only the students found the Chemistry scenario interesting, captivating and enjoyable but they were also delighted to feel in control of the learning process.

Keywords: educational software, augmented reality, user experience, user testing, formative usability evaluation.

1 Introduction

The rapid development of the AR technologies over the last decade is creating new challenges for designers in many application domains such as medicine, education, entertainment and industry. As AR technologies become more wide-spread, there is an increasing interest in their usability. It is widely recognized that designing for usability is not an easy task in the AR field, given the lack of specific user-centered design methods and usability data [1], [2], [8], [18]. A survey of user-based experimentation elaborated three years ago by Swan & Gabbard revealed that only 38 out of 1104 articles (published in four main AR publications during 1992-2004) addressed some aspect of HCI and only 21 (~2%) described a user testing experiment [22]. Although the situation significantly changed over the last years there is still little usability data available for the design of specific AR-based applications.

Augmented Reality systems are creating a new kind user experience (UX), which is based on the integration of the real and virtual into the same interaction space. While the configurations based on head mounted displays (HMD) are integrating

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 31–42, 2008.

specific AR devices into a real life environment, the desktop AR configurations are bringing real life objects into a computing environment. The choice of the viewing configuration depends on several conditions such as cost, task requirements or maintenance problems [12].

The user experience with a specific technology is also influenced by the opportunities created by its application in a particular application domain. In the last years there is a growing interest in using novel technologies in e-learning given their potential to increase the students' motivation to learn. More recently, AR-based educational systems have been developed that are implementing learning scenarios for primary and secondary schools.

Bringing various real life objects into a computing environment is increasing the complexity of the human-computer interaction and requires the development of suitable interaction techniques. In order to prevent usability problems, the interaction components of the AR system have to be tested with users early in the development process. The earlier these problems are identified the less expensive is the development effort to fix them.

Formative usability testing is performed in an iterative development cycle and aims at finding and fixing usability problems as early as possible [23]. This kind of usability evaluation is called "formative" in order to distinguish it from "summative" evaluation which is usually performed after a system or some component was developed [20].

Formative usability evaluation can be carried on by conducting an expert-based usability evaluation (sometimes termed as heuristic evaluation) and / or by conducting user testing with a small number of users. In this last case, the evaluation is said to be user-centered, as opposite to expert-based formative evaluation. As Gabbard et al. [8] pointed out, this kind of user-based statistical evaluation can be especially effective to support the development of novel systems as they are targeted at a specific part of the user interface design.

This work reports on a user-centered formative usability evaluation of an AR-based learning scenario for Chemistry which has been developed in the framework of the ARiSE (Augmented Reality for School Environments) research project. The main objective of the ARiSE project is to test the pedagogical effectiveness of introducing AR in primary and secondary schools and creating remote collaboration between classes around AR display systems. An important research goal is to investigate the extent to which each learning scenario is enhancing the students' motivation to learn.

The project is carried on in a consortium of five research partners and two school partners. ARiSE is developing a new technology, the Augmented Reality Teaching Platform (ARTP) in three stages thus resulting three research prototypes. In order to get a fast feedback from both teachers and students, each prototype is tested with users during an ARiSE Summer School which is held yearly. The second prototype implemented a Chemistry scenario and was being tested during the 2nd ARiSE Summer School in Bucharest.

A usability questionnaire was developed in this project that is targeting several dimensions such as: the general ergonomics of the ARTP, the usability of the target e-learning application, the perceived utility, the attitude towards the system, and the intention to use. The questionnaire provides with both quantitative and qualitative measures of the educational and motivational values of a new learning scenario.

The purpose of this paper is to present and analyze the evaluation results in order to investigate to which extent these are also useful to describe and measure the user learning experience with an AR-based system. In this respect we focus on the items related to attitude towards the system as well as on the positive aspects mentioned by students that are describing their learning experience with the teaching platform.

The rest of this paper is organized as follows. In the next section we will briefly describe some related work with a focus on UX and constructivist learning approaches. In section 3 we will describe the evaluation set-up. Then we will present and compare the quantitative and qualitative evaluation results that are related to the motivational value of the learning scenario. The paper ends with conclusion and future work in section 5.

2 Related Work

As pointed out in [14], [15], UX is an emerging research topic in the area of HCI. In his UX model, Marc Hassenzahl [9] is distinguishing between two kinds of attributes that are giving a product character: pragmatic (related to utility and usability) and hedonic (associated with well being and pleasure). Typical hedonic attributes are attractive, exciting and interesting.

Gilbert Cockton takes a pragmatic approach on the evaluation of user experience and advocates for grounding UX in some form of intended value. A given context of use brings in front specific values of the application domain. Therefore, measures should be interpreted according to the achieved value that attempt to assess [3].

Traditionally, there are two main approaches regarding the instructional process: the curriculum-centered approach and the constructivism. The second approach provides with a valid and reliable basis for a theory in various e-learning environments where the challenge for designers is to implement an active, collaborative and authentic learning [13].

Social constructivism is viewing each learner as a unique individual with unique needs and backgrounds and is encouraging the learner to arrive at her (his) own version of the truth, influenced by the background, culture or embedded world view. In this respect, learning takes place when the student is able to build conceptual models that are consistent with both what they already understand and with the new content. In order to ensure successful adaptation of old knowledge to a new learning experience, a flexible learning direction should be provided [17] as well a possibility to put learners in control of the knowledge transfer process [21].

A crucial assumption regarding the nature of the learner concerns the level and source of motivation for learning. Motivation is an important factor in education and is appreciated that a high level of motivation is often a prerequisite for success. Conversely, there is a high probability that learning will not be successful if there is a lack of motivation [10]. A way to increase learners' motivation is to integrate educational games that are making the learning process more approachable and enjoyable [4], [7].

Since 1981, Malone identified a set of heuristics for designing enjoyable user interfaces: challenge, fantasy and curiosity [17]. These features have been considered as essential characteristics for good educational software by Holzinger et al. [10]. Based on a learner-centered design approach, they extracted a comprehensive list of demands for designing motivating e-learning systems. The main characteristics of the user experience with such a system are: attraction, fun, challenge, fantasy, curiosity, interaction, multi-modality, and ease of use.

3 Experiment and Methodology

3.1 Participants

The 2nd ARiSE Summer School was held on 24-28 October 2007 in Bucharest, Romania. Two groups of students and two teachers from the partner schools and three groups of students accompanied by a total of 4 teachers from 3 general schools in Bucharest were invited to participate.

A total of 20 students from which 10 boys and 10 girls tested the ARTP. 12 students were from 8th class (13-14 years old), 4 from 9th class (14-15 years old) and 4 from 10th class (15-16 years old). Students have different ages because of the differences related to the Chemistry curricula in each country. None of them was familiar with the AR technology.

3.2 Equipment and Tasks

ARTP is a "seated" AR environment featuring a desktop viewing configuration: users are looking to a see-through screen where computer generated images are superimposed over the perceived image of the real objects which are placed on the table [25]. The platform has been registered by Fraunhofer IAIS under the trade mark Spinnstube®.

User testing has been conducted on the ARTP of ICI Bucharest which is equipped with 4 modules organized around a table, as illustrated in Fig.1.

The real objects integrated into the Chemistry scenario are a periodic table and a set of colored balls (4 colors) symbolizing atoms.

The periodic table has two sides: part A, with full notation of chemical elements and part B, with numbered groups and periods. Part B is used to test how students understand the internal structure of atoms. Each workplace has its own periodic table. Once a ball is placed over a chemical element it could be further used as an atom of that element to form molecules. Further on, molecules placed on the tables could be used to simulate chemical reactions. Fig. 1 illustrates how two students are placing a ball onto their periodic table.

A remote controller Wii Nintendo has been used as interaction tool for selecting a menu item.

The participants were assigned 14 tasks: an introduction and 13 exercises related to three lessons. The first lesson is about the chemical structure of the atoms and has 2 exercises. The second lesson is about forming molecules and has 8 simple exercises. The third lesson is about chemical reactions and has 3 exercises.

3.3 The Evaluation Instrument

The ISO/IEC standard 9126 defines usability as the capability of a software system to be understood, learned, used, and liked by the user when used under specified conditions [11].

Fig. 1. Students testing the Chemistry scenario

In this project we took a broader view on usability evaluation by grounding our work in a well known technology acceptance model (TAM) developed by Davis et al. [5]. As Dillon & Morris pointed out, TAM provides with early and useful insights on whether users will or will not accept a new technology [6]. According to Venkatesh et al. [24], TAM has been widely used as an information technology acceptance model able to explain or predict behavioral intention on a variety of information technologies and systems.

According to our knowledge, there is no TAM model available for AR-based educational systems. Therefore, an evaluation questionnaire was developed within the ARiSE project that has 28 closed items (quantitative measures) and 2 open questions, asking users to describe the most 3 positive and most 3 negative aspects (qualitative measures) [19].

The questionnaire is presented in Table 1. The first 24 closed items are targeting various dimensions such as ergonomics of the AR platform (items 1-5), usability of the application (items 6-14), perceived utility (items 15-17), attitude towards the system (items 18-21) and intention to use (items 22-24).

The remainder four closed items are to assess how the students overall perceived the platform as being easy to use, useful for learning, enjoyable to learn with and exciting.

3.4 Procedure

Testing and debriefing with users was done in the morning while the afternoon was dedicated for discussion between research partners. Before testing, a brief introduction to the AR technology and ARiSE project had been done for all students. Each group of students tested ARTP in a session during 75 min.

Table 1. The usability evaluation questionnaire

No.	Item
1	Adjusting the "see-through" screen is easy
2	Adjusting the stereo glasses is easy
3	Adjusting the headphones is easy
4	The work place is comfortable
5	Observing through the screen is clear
6	Understanding how to operate with ARTP is easy
7	The superposition between projection and the real object is clear
8	Learning to operate with ARTP is easy
9	Remembering how to operate with ARTP is easy
10	Understanding the vocal explanations is easy
11	Reading the information on the screen is easy
12	Selecting a menu item is easy
13	Correcting the mistakes is easy
14	Collaborating with colleagues is easy
15	Using ARTP helps to understand the lesson more quickly
16	After using ARTP I will get better results at tests
17	After using ARTP I will know more on this topic
18	The system makes learning more interesting
19	Working in group with colleagues is stimulating
20	I like interacting with real objects
21	Performing the exercises is captivating
22	I would like to have this system in school
23	I intend to use this system for learning
24	I will recommend to other colleagues to use ARTP
25	Overall, I find the system easy to use
26	Overall, I find the system useful for learning
27	Overall, I enjoy learning with the system
28	Overall, I find the system exciting

During testing, effectiveness (binary task completion and number of errors) and efficiency (time on task) measures were collected in a log file. Measures were collected for all exercises performed.

After testing, the students were asked to answer the new usability questionnaire by rating the items on a 5-point Likert scale (1-strongly disagree, 2-disagree, 3-neutral, 4-agree, and 5-strongly agree). Prior to the summer school, the questionnaire was translated in the native language of students.

4 Evaluation Results

4.1 Quantitative Data Analysis

The sample was small (n=20), and problems exist both with respect to generalization to the full population as well as with respect to the choice of statistical techniques to be used for the analysis. Therefore triangulation between quantitative and qualitative data plays an important role. Reliability of the scale (Cronbach's Alpha) was 0.929 which is acceptable.

The analysis of quantitative data revealed four items rated lower than 3.50. These items are associated with the accuracy of visual perception (items 1, 5, and 7) and the difficulty to correct errors (item 13). Other four items were rated between 3.50 and 3.75. These items are associated with other usability problems (items 2, 11, 14 and the general item 25). Five items were rated between 3.75 and 4.00 (agree), eight items between 4.00 and 4.25 and the rest of seven items over 4.25.

Overall, the items associated with the perceived utility, attitude and intention to use dimensions were better rated than the items associated with the usability of the ARTP. The mean values for the items 15-28 are presented in Fig. 2.

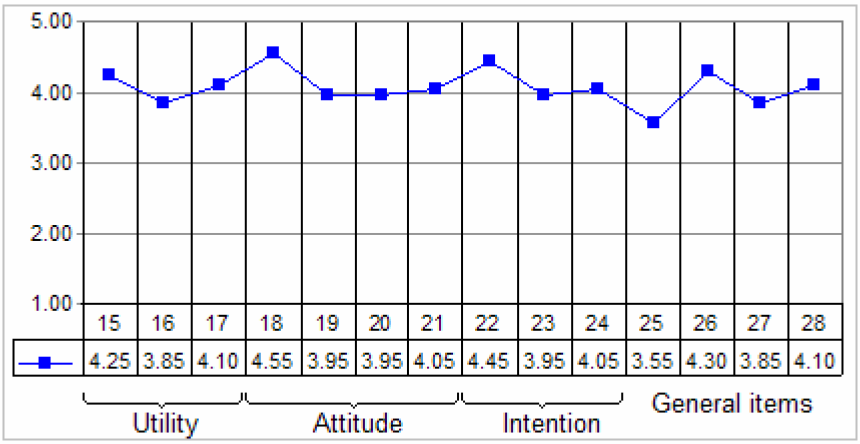


Fig. 2. Answers to the questionnaire (items 15-28) – mean values

Four items have mean values over 4.25 showing that ARTP helps to understand the lesson more quickly (item 15), makes learning more interesting (item 18), the students would like to have it in school (item 22) and overall, they find it useful for learning (item 26). Other four items have mean values between 4.00 and 4.25 showing that ARTP helps to learn more (item 17), the exercises are captivating (item 21), the students will recommend this system to their colleagues (item 24) and overall, the students found the system exciting.

In order to investigate in more detail the educational and motivational value of the ARTP we will compare the answers to items 15-21, and 26-28 with the positive aspects mentioned by students in their answers to the open questions.

4.2 Qualitative Data Analysis

The answers to the open questions were analyzed in order to extract key words (attributes). Then attributes were grouped into categories. Some students only described one or two aspects while others mentioned several aspects in one sentence thus yielding a total number of 70 positive aspects which are summarized in Table 2.

Table 2. Summary of the most positive aspects

Category	Frequency	Percent
Educational support	20	28.57
Exciting and motivating	25	35.71
AR visualization & user guidance	17	24.29
Easy to use and other	8	11.43
Total	70	

Educational support includes various aspects, like: better understanding ("*Everything is more clear now*"), easier understanding ("*I understood the lesson more easy than in school*") and easier learning ("*The system make learning more easy*").

Students liked the 3D visualization ("*the 3D visualization of atoms was very well*", "*The creation of molecules and other chemical processes are visualized very nice and demonstrative.*" as well as the multimodal interface providing with vocal user guidance throughout the interaction process ("*Explanation complete and easy to understand because of images*").

The total of 17 positive aspects mentioned by students is suggesting the idea that the AR technology is a good facilitator for learning. In this respect, the augmentation of a real object with real time 3D computer graphics acts as a moderating factor for both the educational and the motivational values.

Overall, 25 of the positive aspects (35.71%) are directly related to the motivational value of the ARTP. A detailed grouping of these aspects is presented in Table 3. Essentially, these are the hedonic attributes of the ARTP as perceived by students. The related closed item is specified in the left column.

Most of the aspects (10) are related to item 18 which got the highest rating (4.55) and clearly shows that the system makes learning more interesting.

Two students found this learning scenario more interesting than a traditional one: "The lessons will be more interesting by using a device like this". Three students found the tasks (i.e. exercises) very interesting: "It was very interesting to fill in all periodic table", "The tasks are very interesting. Especially the third one". Four students mentioned in a more or less specific way that it was interesting ("It was very interesting", "It is interesting to work and interact", "The system is interesting"). One student mentioned that learning with the ARTP is less boring ("Less boring than the teachers").

This last aspect also reveals a students' preference to work individually which is consistent with the lower rating of item 19 (3.95).

The lower rating of the item 20 (3.95) is closely related with some usability problems. Students complained that it was difficult to distinguish the real balls through the screen because of the augmentation and the superposition of the real and virtual objects.

Nevertheless, the positive aspects mentioned by students are revealing an interesting facet of a user learning experience: mastering of a chemical process. Students were attracted by the interaction in the AR environment which is giving the students the possibility to build something by their own: "*I was surprised that I was able to interact with balls, to move them whatever where you want. It was like my direct influence to something*", "*You can feel your direct influence. It is great.*", and "*You can feel that something can be created by yourself, not by teachers*". These comments are expressing the students' preference for a learning experience based on learning by doing instead of learning by reading.

Table 3. Categories of positive aspects related to motivation

Item	Aspect (attribute)	Frequency	Cumulative	Percent
18	Interesting learning	2	10	40%
	Interesting tasks	3		
	Interesting	4		
	Less boring	1		
19	Stimulating team work	1	1	4%
20	Mastering the learning process	4	5	20%
	AR interaction	1		
21	Captivating exercises	2	2	8%
27	Attractive learning	3	6	24%
	Funny	3		
28	Novel	1	1	4%
	Total	25	25	

Two students found the exercises captivating which shows an intrinsic motivation created by the AR technology and the added value of exploiting the need to play, predominant to the age of students in primary and secondary school. Students liked the idea of learning with ARTP which is attractive (*"It is very attractive to produce atoms and molecules"*), novel (*"It is something new"*) and funny (*"It is funny to touch and move the atoms"*, and *"The different colors make learning more funny"*).

4.3 Age and Gender Analysis

A grouping by age and gender of the positive aspects related to the hedonic attributes of the ARTP is presented in Table 4.

Overall, the 8 students from the 9th and 10th classes mentioned a total of 13 positive aspects while the 12 students from the 8th class only mentioned 12 positive aspects that are related to the hedonic function of the ARTP. This suggests that the younger students were more attracted by the pragmatic aspects. Older students found ARTP more interesting and appreciated the possibility to master the learning process.

Table 4. Positive aspects by age and gender

Category	13-14 years		15-16 years	
	Girls	Boys	Girls	Boys
Interesting	1	2	5	2
Stimulating	0	1	0	0
Mastering	1	0	2	2
Captivating	2	0	0	0
Enjoyable	1	3	0	1
Novel	0	1	0	0
Total	5	7	7	6

The boys found learning with the ARTP more enjoyable and novel than girls. On the other hand, the girls mentioned a greater number of positive aspects related to the interesting and captivating character of the chemistry scenario.

5 Conclusion and Future Work

The usability evaluation results provided with an early feedback on the implemented scenario. ARTP was perceived as a useful teaching aid able to increase the students' motivation to learn: the system is attractive, stimulating and exciting. The students liked the 3D visualization as well as the vocal explanation guiding them throughout the learning process.

Participants to the summer school found the chemistry scenario very attractive. Assigning semantics to a colored ball by placing it onto the periodic table is something new and makes learning more interesting. The students liked the idea of "learning by doing" which provides them with a better understanding of the Chemical reactions.

While item 26 (usefulness for learning) received the highest rating among the general items (4.30), most of the positive aspects mentioned by students were related to motivation. From the one hand, the quantitative data results suggest that they mainly perceived the chemistry lesson as a useful *learning experience*. From another hand, the qualitative data results are revealing the importance of the motivational value for the learning process.

The results of this research are showing that measuring the educational and motivational values is providing with useful UX measures. Nevertheless, there are some inherent limitations since it was difficult to relate some aspects to a specific closed item. In order to get more useful results, some more items might be needed. The qualitative data could also gain in accuracy by interviewing students after filling in the questionnaire.

In this paper we briefly discussed the quantitative and qualitative data and we mainly focused on the measures related to attitude. However, students also mentioned 65 negative aspects. These aspects are actually describing usability problems which are consistent with the relatively low ratings of the items 1-14 and the general item 25. A future research direction is to investigate to which extent UX is negatively affected by specific usability problems.

Acknowledgement

We gratefully acknowledge the support of the ARiSE research project, funded under FP6-027039.

References

1. Bach, C., Scapin, D.: Obstacles and perspectives for evaluating mixed reality systems usability. In: Proceedings of Mixer Workshop - IUI-CADUI 2004, Funchal, pp. 72–79 (2004)
2. Bowman, D., Gabbard, J., Hix, D.: A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. Presence: Teleoperators and Virtual Environments 11(4), 404–424 (2002)
3. Cockton, G.: Valuing User Experience. In: Procedings of UX Workshop NordiCHI 2006, pp. 100–105. ACM Press, Oslo (2006)

4. Costabile, M.F., De Angeli, A., Lanzilotti, R., Ardito, C., Buono, P., Pederson, T.: Explore! Possibilities and Challenges of Mobile Learning. In: Proc. of CHI 2008, pp. 145–154. ACM Press, New York (2008)
5. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35(8), 982–1003 (1989)
6. Dillon, A., Morris, M.: User acceptance of new information technology: theories and models. In: Williams, M. (ed.) Annual Review of Information Science and Technology, vol. 31, pp. 3–32. Information Today, Medford (1996)
7. Ebner, M., Holzinger, A.: Successful Implementation of User-Centered Game Based Learning in Higher education: An Example from Civil Engineering. Computer & Education 49(3), 873–890 (2007)
8. Gabbard, J., Hix, D., Swan, E., Livingston, M., Herer, T., Julier, S., Baillot, Y., Brown, D.: A Cost-Effective Usability Evaluation Progression for Novel Interactive Systems. In: Proc. of Hawaii Intl. Conference on Systems Sciences, Track 9, p. 90276. IEEE, Los Alamitos (2004)
9. Hassenzahl, M.: The Thing and I: Understanding the Relationship Between User and Product. In: Blythe, M., Overbeeke, K., Monk, A., Wright, P. (eds.) Funology: from usability to enjoyment, ch. 3, pp. 31–42. Springer, Heidelberg (2005)
10. Holzinger, A., Pichler, A., Maurer, H.: Multi Media eLearning Software TRIANGLE Case-Study: Experimental Results and Lessons Learned. Journal of Universal Science and Technology of Learning 0(0), 61–92 (2006)
11. ISO 9126-4:2001 Software Engineering - Software product quality. Part 1: Quality model
12. Jeon, S., Shim, H., Kim, G.: Viewpoint Usability for Desktop Augmented Reality. The International Journal of Virtual Reality 5(3), 33–39 (2006)
13. Karagiorgi, Y., Symeou, L.: “Translating Constructivism into Instructional Design: Potential and Limitations. Educational Technology & Society 8(1), 17–27 (2005)
14. Law, E., Hvannberg, E.T., Hassenzahl, M.: User Experience – Towards a unified view. In: Procedings of UX Workshop NordiCHI 2006, pp. 1–3. ACM Press, Oslo (2006)
15. Law, E., Bevan, N., Gristou, G., Springett, M., Larusdottir, M.: Meaningful measures: valid useful user experience measurement - VUUM Workshop 2008, Reykjavik. COST Action 294 – MAUSE (2008)
16. Malone, T.W.: Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games. In: Proceedings of the 1982 conference on Human factors in computing systems, Gaithersburg, Maryland, United States, pp. 63–68 (1982)
17. Mantovani, F.: VR Learning: Potential and Challenges for the Use of 3D Environments in Education and Training. In: Riva, G., Galimberti, C. (eds.) Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age. IOS Press, Amsterdam (2001)
18. Nielsen, S., Johansson, B.: Fun and Usable: Augmented Reality Instruction in a Hospital Setting. In: Proc. of OzCHI 2007, pp. 123–130. ACM Press, New York (2007)
19. Pribeanu, C., Balog, A., Iordache, D.D.: Formative user-centered usability evaluation of an augmented reality educational system. In: Proc. of Interantional Conference on Software and Data Technologies – ICSOFT 2008, pp. 65–72. INSTICC Press (2008)
20. Scriven, M.: Evaluation thesaurus, 4th edn. Sage Publications, Newbury Park (1991)
21. Stary, C., Tooter, A.: On Learner Control in e-Learning. In: Proc. of ECCE 2006 – Trust and Control in complex Socio-technical Systems, pp. 41–48. ACM Press, New York (2006)
22. Swann II J., Gabbard, E.J.: Survey of User-Based Experimentation in Augmented Reality. In: Proceedings of 1st International Conference on Virtual Reality, Las Vegas, Nevada, July 22-27 (2005)

23. Theofanos, M., Quesenbery, W.: Towards the Design of Effective Formative Test Reports. Journal of Usability Studies 1(1), 27–45 (2005)
24. Venkatesh, V., Davis, F.D., Morris, M.G.: Dead Or Alive? The Development, Trajectory And Future Of Technology Adoption Research. Journal of the AIS 8(4), 267–286 (2007)
25. Wind, J., Riege, K., Bogen, M.: Spinnstube®: A Seated Augmented Reality Display System. In: Virtual Environments, IPT-EGVE – EG/ACM Symposium Proceedings, pp. 17–23 (2007)

Cognitive Processes Causing the Relationship between Aesthetics and Usability

Waltraud Ilmberger[1], Martin Schrepp[2], and Theo Held[2]

[1] Ruprecht-Karls-Universität Heidelberg, Grabengasse 1, 69117 Heidelberg, Germany
[2] SAP AG, Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
Waltraud.Ilmberger@stud.uni-heidelberg.de,
Martin.Schrepp@sap.com, Theo.Held@sap.com

Abstract. We conducted an experiment to find out which cognitive processes underlie the high correlations between aesthetic impression and perceived usability of a user interface reported in several studies. Four online shops differing in colour scheme and usability level were created and tested before and after an interaction period to gain support for two differing models of explanation. One stems from Norman who hypothesizes that mood is a mediator between design and perceived usability whereas the other is based on the well known "What is beautiful is good" phenomenon from social psychology and market research. The data provide support for the approach of Norman whereas the high correlations couldn't be replicated. Another interesting finding was that our results militate rather for a "What is usable is beautiful"- relation than the other way round as hypothesized.

Keywords: Usability, Aesthetics, Cognitive Processes, User Experience.

1 Introduction

In the last couple of years there is a clear tendency to extend the concept of usability to a more holistic view on the interaction between humans and systems, which is often referred to as *user experience* [1]. User experience summarizes classic usability criteria like efficiency, effectiveness, or simplicity with additional criteria like fun, aesthetics, or attractiveness. An often used terminological distinction is to refer to the first set of criteria as *pragmatic quality aspects* and to the second set as *hedonic quality aspects*[1] [3].

One research direction in this area is to investigate the relationship between pragmatic quality and hedonic quality. There is, for example, strong evidence from several empirical studies that perceived aesthetics influence the perceived usability of a system.

For example, in an empirical study [4] several variants of user interfaces for automatic teller machines (ATM) were compared. Functionality and interaction behaviour of these variants were identical. The only difference was the arrangement of the controls. Some variants arrange the controls in a way that is perceived as attractive;

[1] Another common terminology is to refer to the first set of criteria as usability goals and to the second set as user experience goals [2].

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 43–54, 2008.

others arrange them in a way which is perceived as unattractive. The results of the study showed that the more attractive variants were perceived to be easier to use. The study reports very high correlations between aesthetic impression and perceived usability of a system. These results were later on replicated in other studies (for example [5]).

A weakness of these studies was that the participants just judged the usability of the user interface without actually using it. But in a further study [6] it was shown that the dependency between perceived aesthetics and usability can also be found in settings where users judged the usability of the system after using it. The authors of this study summarized this result in the statement *What is beautiful is usable*.

It must be noted however, that these results could not be confirmed in all studies concerning this topic. For example, in [1] it was shown that the relation between appeal and perceived usability does not exist for all types of interfaces. But the counterexamples reported were of a quite special type. In other studies [7] the reported correlations between aesthetic appeal and perceived usability were much smaller than the corresponding values reported in [4, 5, 6].

The dependency between pragmatic quality and hedonic quality is not restricted to simple user interfaces like ATM's, web sites or games. A study [8] concerning highly complex user interfaces of business software showed that also in this area both pragmatic and hedonic quality aspects have an impact on the attractiveness of the user interface. In addition, this study demonstrated that the more attractive an interface is, the higher is the preference of subjects for this interface.

Empirical results [9] show that user acceptance of new technologies is affected both by perceived usefulness and perceived fun of the technology. Today user acceptance plays a very important role for the success of many software projects, especially in the area of business software [10]. It is in many cases impossible to introduce a new software solution in a company if the users reject to use it. Thus, the relation of perceived attractiveness and perceived usability has potentially also a high economic impact.

While the existence of a connection between perceived aesthetics respectively hedonic quality aspects and perceived usability is out of question it is still not clearly understood why people connect these two concepts. How can we explain this connection? What cognitive processes are responsible for it? These are the questions we try to address in this paper.

2 Theoretical Background

We will discuss two approaches to explain the connection between perceived aesthetics and perceived usability.

2.1 Influence of the Emotional State

One possible explanation of the dependency between perceived aesthetics and perceived usability is suggested by Norman [11]. He argues that emotions influence the way users handle and evaluate usability problems. Since aesthetics can influence the emotional state of a subject, these findings can explain the mentioned dependency.

This argumentation relies on actual research results [12], which show that a positive emotional state broadens the problem solving processes and facilitates creative thinking. A positive emotional state supports heuristic and explorative problem solving strategies. A negative emotional state on the other hand facilitates a systematic and analytical way to deal with problems[2] [13, 14].

If a user in a negative emotional state will run in a problem on the user interface he or she will thus typically focus on this problem and try to analyse it. If a user in a positive emotional state will run in the same problem he or she will most likely explore the situation and will try to get around the problem based on some heuristics. Users in a positive emotional state will therefore find alternative solutions for usability problems more easily. In addition, they will be tolerant against minor difficulties in a user interface [16]. Users in a negative emotional state will on the other hand focus on problematic details and will therefore be less tolerant against observed difficulties in a user interface.

According to Norman [11] this influence of the emotional state of a subject on his or her problem solving strategies explains the dependency between perceived aesthetics and perceived usability.

2.2 Stereotypes and Evaluative Consistency

A second often used (for example, [7]) explanation for the dependency between perceived aesthetics and usability is the *what is beautiful is good stereotype* known from research in social psychology [17, 18]. Results show that people relate a person's beauty positively to attributes like social competence, adjustment, intellectual ability or general goodness. In a similar way the aesthetic impression of a user interface might also indicate good usability.

Similar mechanisms are also reported from studies in market research. For example, when information about certain attributes of a product are missing, then consumers evaluated these missing attributes in conjunction with the brand's overall evaluation (*evaluative consistency*) [19]. Another example is that if the prize of a product is high, then people often infer a high quality from this fact by applying a price-quality heuristic [20].

If we apply this to the area of user interface design similar stereotypes or inference heuristics can cause the relation between aesthetic impression and perceived usability. Thus, especially in cases where users have no or little interaction experience with a user interface [21] these stereotypes can explain that the missing information about the actual usability of the user interface is inferred from known information, for example the visual appeal of the user interface.

2.3 Research Hypotheses

Please note that the two processes described above do not exclude each other. It may be possible that both explanations contribute to some extent to the relation between aesthetics and perceived usability.

---

[2] A possible explanation for this behavioural pattern is provided in [15]. It is argued that negative emotions indicate a problematic situation. Thus, there is a motivation for the person to change this situation, which requires a detailed analysis of the problem. On the other hand, positive emotions indicate an unproblematic situation. There is no motivation for the person to change the current situation and thus a heuristic and explorative problem solving behaviour is sufficient.

If the relationship between perceived aesthetics and perceived usability results from an influence of the cognitive state or mood of the subject on problem solving processes, then we would assume that:

- The perceived aesthetics of a user interface should have an influence on the mood of subjects (*Hypothesis 1*).
- The mood of a subject should influence the perceived usability of a user interface (*Hypothesis 2*). More detailed:
  - Subjects in a good mood should rate the usability of a user interface better than subjects in a bad mood.
  - Subjects in a good mood should report fewer problems in a user interface than subjects in a bad mood.

If the relationship between perceived aesthetics and perceived usability results from the application of choice heuristics, like for example *What is beautiful is usable,* we would expect that:

- If the subjects have not interacted with a user interface (for example, if they only have seen screen shots or a demo of the interaction behaviour), then their usability judgement should be based mainly on the aesthetic impression. In addition, the observed correlations between ratings of hedonic quality aspects and pragmatic quality aspects should be high (*Hypothesis 3*).
- If subjects have already interacted with a user interface, then their usability judgement will be based on their interaction experience. Thus, the experienced usability of an interface should have a strong impact on the usability judgement by the subjects. In addition, the observed correlations between hedonic quality aspects and pragmatic quality aspects should be lower than in the condition where no interaction took place (*Hypothesis 4*).

3 Method

To test the four hypotheses stated above, an experimental design varying aesthetic quality and usability of a web-shop was used.

3.1 Participants

72 subjects (51 female, 21 male), mainly psychology students of the Ruprecht-Karls-University of Heidelberg, took part in the experiment. They were between 19 and 31 years old (average 22.21 years) and received a certificate for participation in the study. Almost all participants (n= 68) had online shopping experience and most of them (n=61) had used the Internet for more than five years.

3.2 Material

We intended to vary usability and one aesthetic dimension -namely the colour scheme- of the user interface independently. Online shops were chosen as the domain of study, because almost everyone has some domain knowledge there. Based on the template of a shop interface at STRATO [22] (see Figure 1) two different versions only varying in colour were designed.

Aesthetic Variation by Colour Manipulation. Two pilot studies were conducted to identify suitable manipulations of the colour dimensions of the web-shop.

The goal of the first study was to create two pleasant variants of the visual design for the web shop. As a basis for the visual design four of the ten highest rated colour schemes from Adobe® kuler[3] [23] were used. The colour schemes were used to define the colours for the user interface elements (background, menus, etc.) of the web shop.

To decide which of the four variants of the web shop is rated most aesthetic, 18 subjects (6 employees and 12 students) at the SAP AG were contacted via e-mail to do a complete paired comparison with screenshots of the four interfaces. The Bradley-Terry-Luce (BTL) model [24, 25] was used to analyze the paired comparison data. This method is particularly appropriate for decisions concerning subjective preferences between alternatives.

There were no significant differences between the BTL scale values of the four alternatives. For further investigation we selected two out of the four alternatives. Therefore, we took the interface that was rated highest in the complete sample and the interface that was preferred in the students sample, because the later sample was also planned to consist of students.

Thereafter we designed three shops with comparable colour contrasts, but (assumed) unpleasing colour combinations. The reason for achieving similar contrast proportions was to avoid a confounding of usability and aesthetic dimensions. To obtain unfavourable colour combinations, the results of a preceding analysis of the kuler [23] colour combinations were used. This analysis showed that concerning the HSV colour space, that defines all colours by means of hue, saturation and value, almost no colours in the hue ranges 130-150 and 230-320 appeared among the highest rated combinations. For the construction of the unpleasant colour combinations colours in these hue ranges were chosen.

A second paired comparison with 17 subjects (6 employees and 11 students of SAP AG) was conducted to find the two most differing interfaces. BTL values were .370 for the best rated interface and .038 for the worst rated interface. Since the BTL scale is a ratio scale we can conclude that the best rated interface is rated as almost ten times more aesthetic then the worst rated interface. These two web shops were chosen to be used in the experiment.

Usability Manipulation. To produce two different usability levels, the following system features were varied according to general usability principles (e.g. [26]):

- visualization and position of the shopping cart
- existence of an overall product search
- link effects (e.g. hover effects of links)
- depth and hierarchy of the menu
- naming of menu items (for example *transport* instead of *rucksacks*)
- the possibility to change the number of ordered products directly before adding it to the shopping cart

[3] Adobe® kuler [22] is a web-hosted application from Adobe Labs for generating, downloading or rating colour themes.

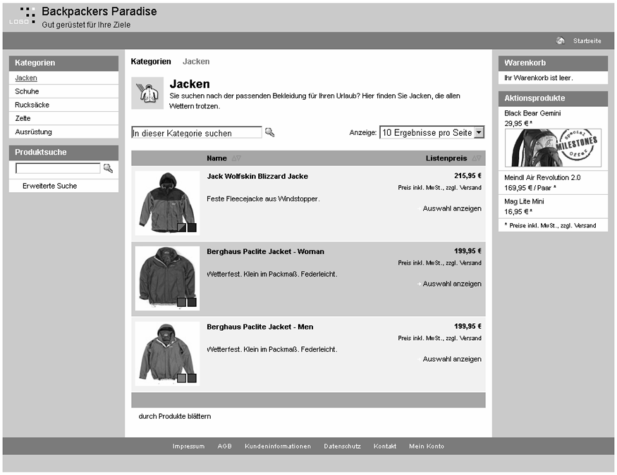


Fig. 1. Screenshot of the shop interface with good usability

- the naming of the products so that it is more or less difficult to find them directly over the search function

Since each of the variations of 'usability' and 'colour' had two levels, four online shops were created:

- high-usability and pleasant-colour
- high-usability and unpleasant-colour
- low-usability and pleasant colour
- low-usability and unpleasant colour.

3.3 Measures

First, there were two variables about how the participants liked the interface visually as a whole and the colour combination in particular to ensure that the two versions differed in colour evaluation as assumed. The User Experience Questionnaire (UEQ) [27] served to capture the user's perception of usability and hedonic dimensions. The questionnaire consists of twenty-six 7-point items with bipolar verbal anchors that measure six dimensions, thereof three usability dimensions that are called *perspicuity*, *efficiency* and *dependability*, two hedonic dimensions called *stimulation* and *novelty* plus one overall dimension that is called *attractiveness*. The sub-dimension scores were calculated by averaging the respective item values per participant. Internal consistency (measured by Cronbach's alpha coefficient) of the six dimensions was high (attractiveness, $\alpha = .92$; perspicuity, $\alpha = .88$; efficiency, $\alpha = .78$; dependability, $\alpha = .83$; stimulation, $\alpha = .84$ and novelty, $\alpha = .81$). This confirmed the assumption that users perceive these dimensions consistently. The questionnaire was conducted before and after interacting with the shop.

We collected also possible confounding variables such as Internet and online shopping experience by asking the subjects "How much Internet experience (in years) do you have" and "How often did you shop online yet". Furthermore we assessed subjective emotional data via the Self-Assessment-Manikin (SAM) [28] which retrieves the valence and intensity of emotions.

3.4 Procedure

Participants were led one by one into the laboratory (at Psychologisches Institut, University Heidelberg) and given a booklet comprising a description of the experimental procedure and all questionnaires. They were asked to fill in some demographic data and answer three questions concerning Internet and online shopping experience. After that they had to rate their subjective feelings on the SAM. Depending on the experimental condition to which they were assigned randomly, the participants saw a demo version of the relevant shop in form of a screen video with tone. In this video they could observe how someone conducts the task of finding a certain jacket and adding it to the shopping cart. They filled in the UEQ-questionnaire for a first impression of the shop.

After that, the participants started to accomplish five typical tasks in the web-shop, where the first one was identical with the task seen in the demo. The interaction took about 3 to 11 minutes (means of 3.60 for the good and 5.83 minutes for the bad usability condition). After they had finished all tasks, the participants were again asked to fill in the SAM. They rated the shop on the UEQ and how they liked the interface visually as a whole and the colour combination in particular. Finally there were two open questions about problems and ideas for improvement regarding the shop they interacted with. Each participant worked through only one shop version. The experiment took roughly 40 minutes on average.

4 Results

4.1 Manipulation Check

A t-Test between the pleasant and the unpleasant condition confirmed that the colour manipulation was successful $t(70) = 6.99$, $p < .01$(one-tailed). A 2x2 ANOVA for 'usability' and 'colour' was performed assessing task-completion times. There was a significant main effect for 'usability' only, indicating that usability was also varied successfully $F(1, 68) = 74.98$, $p < .01$.

4.2 The Hypothesis of Norman

We first analyzed if the participants in the pleasant colour condition were in a better mood after interacting (*Hypothesis 1*) and found that the valence appeared to be slightly better (means of 5.39 and 5.14, respectively), but this difference was not significant.

A two way analysis of variance revealed that the valence-dimension of the SAM had a significant effect on both, the number of problems reported $F(4, 67) = 2.93$, $p < .05$ and the usability dimensions of the UEQ such as perspicuity $F(4, 71) = 4.47$, $p < .01$, efficiency $F(4, 71) = 3.61$, $p < .05$ and dependability $F(4, 71) = 2.68$, $p < .05$ (*Hypothesis 2*).

The better the participants rated their valence of mood, the higher they evaluated the shops' usability and the fewer problems were reported. These findings support that Norman could be right with his statement that "attractive things work better" [11]. This could especially be true if an interface is aesthetic enough to improve the mood of a user.

While Hypothesis 1 has to be rejected, Hypothesis 2 was confirmed.

4.3 The "What Is Usable Is Beautiful" Hypothesis

We hypothesized that the participants had too few usability information after watching the demo. An ANOVA with the data of the first survey (questionnaire filled in directly after the demo) revealed that the usability condition had no influence on usability dimensions of the UEQ questionnaire that are efficiency $F(1, 70) = 1.16$, $p > .05$, perspicuity $F(1, 70) = 0.41$, $p > .05$ and dependability $F(1, 70) = 0.24$, $p > .05$. So it wouldn't be surprising if the users apply a heuristic to judge the usability of the shop.

If the heuristic 'What is beautiful is usable' [6] is also valid for user interfaces, one would expect high correlations between hedonic and usability ratings after only watching the demo (*Hypothesis 3*). After interacting with the system, the real usability of the interface should have a strong impact on the usability judgement by the subjects and as a consequence these correlations should be low (*Hypothesis 4*) because then the user has more information about the real usability and is no more that reliant on using heuristics. After the demo, all correlations were at most moderate (maximum correlation between efficiency and stimulation of $r=.310$, see **Table 1**) and far away from the level found by Kurosu [5] and Tractinsky [6]. This result doesn't fit with Hypothesis 3, but supports findings of Hassenzahl [7], who discusses possible weaknesses of the stimulus material in the studies of the last-mentioned authors.

After the interaction took place, the usability condition had significant influence on all rated usability dimensions after the interaction (perspicuity $F(1, 70) = 21.61$, $p < .01$, efficiency $F(1, 70) = 17.94$, $p < .01$ and dependability $F(1, 70) = 18.90$, $p < .01$). The second part of Hypothesis 4 could not be supported, because all correlations increased against our supposition (maximum correlation between perspicuity and stimulation of $r = .555$, see Table 2). Therefore the fourth hypothesis could be only partially confirmed.

The pattern of correlations doesn't support the hypothesis of users applying similar heuristics as the well known phenomenon in social psychology and market research.

Further results rather argue for a reverse process. Results of the analysis of variance showed that the usability condition had significant influence on ratings of hedonic dimensions of the UEQ questionnaire, such as stimulation $F(1, 70) = 13.39$, $p < .01$, and novelty $F(1, 70) = 3.97$, $p < .05$. The usability condition had also a significant influence on the overall rating of visual design $F(1, 69) = 5.09$, $p < .05$.

The rating of visual design was better in the good usability condition at both measuring times, $t(70) = 1.77$, $p < .05$ and $t(70) = 1.71$, $p < .05$ (both one-tailed). In addition, the visual design judgement decreased in the condition with bad usability and remained equal in the good usability condition $t(35) = 1.98$, $p < .05$ and $t(34) = .49$, $p < .05$ (both one tailed).

The fact that the correlations between hedonic dimensions and usability increase after the interaction is also in line with a theory assuming that perceived usability influences the perception of aesthetic factors. Not until interacting with the online shop the users had information about the real usability, so the higher correlations could have been the result of participants rating aesthetics on base of perceived usability after the interaction took place.

Table 1. Correlations of the relevant UEQ scales after the demo

	perspicuity	dependability	efficiency	stimulation	novelty
perspicuity	1	.396**	.561**	.145	-.286*
dependability		1	.410**	.173	-.128
Efficiency			1	.310**	-.027
stimulation				1	.418**
Novelty					1

* Significant with $p < .05$.
* * Significant with $p < .01$.

Table 2. Correlations of the relevant UEQ scales after the interaction

	perspicuity	dependability	efficiency	stimulation	novelty
perspicuity	1	.827**	.828**	.555**	.165
dependability		1	.731**	.526**	.148
efficiency			1	.463**	.158
stimulation				1	.507**
novelty					1

* Significant with $p < .05$.
* * Significant with $p < .01$.

5 Discussion

The reported investigation was a first step to approach the question for the cognitive processes underlying the relationship between aesthetic properties and the experienced usability of user interfaces.

We selected two prominent and well-established explanatory approaches to derive testable hypotheses. According to the results of our experiment, at least the assumption that the mood of a user influences perceived usability seems to hold. Given that the users' mood can initially be influenced by evident aesthetic qualities of a user interface, this result may have quite strong implications. While it may be impossible to "compensate" bad experienced usability by a good look, it is quite likely that the experience of good usability will be amplified by a good aesthetic impression.

The direct relationship between variations of user interface aesthetics or factors influencing usability and experienced usability of a software product could be uncovered only partially by our investigation. We assume that our operationalization of aesthetics may constitute a problem, because colour is not able to cover all aspects of the construct. Further research should also investigate the influence of other aesthetic dimensions.

A deeper elaboration both of the theoretical background and the experimental design should be the focus of further investigations.

Theoretically interesting is the observation that usability might have an impact on the perception of aesthetic qualities. This indicates that the "what is good is beautiful" relationship that has recently been found in the area of face perception [29] might as well be relevant in the area of usability of user interfaces.

In conclusion our results argue rather for Normans proposal of the emotional state influencing information processing and the evaluation of a product. The "what is beautiful is usable" approach supposed for example by Tractinsky [6] is not able to explain the found pattern of results at all. Our investigation surprisingly rather supports a converse relationship of "what is good is beautiful" between usability and aesthetic evaluations. This means that - putting the trend of emphasizing design aspects into perspective - the perceived usability of an online shop is the striking variable that is even able to influence the aesthetic impression.

It has to be tested if this kind of relationship holds also in other contexts than online shops.

What are the practical implications of our results? First of all, when conducting usability tests or observations in their natural working environment, one should be aware that the mood of the participants (induced by the interface or set by the working environment) influences the usability evaluation. Therefore, the participants in good mood tend to rate systems more favourable and report fewer concrete problems as those in bad mood.

Another implication of our results can be drawn from the fact, that after presenting a non-interactive demonstration of the system, the participants were not able to diversify different levels of usability. According to this, one can not expect reasonable feedback from users only having watched a demonstration. This is especially important in the area of online usability research.

Finally our findings indicate that the software engineer should in each case strive for good usability and good visual design, because they benefit each other. But caution is advised here: bad usability can not be compensated by a good visual design.

References

1. Lindgaard, G., Dudek, C.: What is this evasive beast we call user satisfaction? Interacting with Computers 15, 429–452 (2003)
2. Preece, J., Rogers, Y., Sharp, H.: Interaction design: Beyond human-computer interaction. Wiley, New York (2002)
3. Hassenzahl, M.: The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction 13, 481–499 (2001)
4. Kurosu, M., Kashimura, K.: Apparent usability vs. inherent usability: experimental analysis of the determinants of the apparent usability. In: Conference Companion of human factors in computing systems, Denver, Colorado, pp. 292–293 (1995)
5. Tractinsky, N.: Aesthetics and Apparent Usability: Empirical Assessing Cultural and Methodological Issues. In: CHI 1997, Electronic Publications (1997), `http://www.acm.org/sigchi/chi97/proceedings/paper/nt.htm`
6. Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. Interacting with Computers 13, 127–145 (2000)

7. Hassenzahl, M.: The Interplay of Beauty, Goodness and Usability in Interactive Products. Human-Computer Interaction 19, 319–349 (2004)
8. Schrepp, M., Held, T., Laugwitz, B.: On the influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers 18(5), 1055–1069 (2006)
9. Igbaria, M., Schiffman, S.J., Wieckowski, T.J.: The respective roles of perceived usefulness and perceived fun in the acceptance of microcomputer technology. Behaviour & Information Technology 13, 349–361 (1994)
10. Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
11. Norman, D.: Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books, Boulder Colorado (2003)
12. Isen, A.M.: Positive affect and decision making. In: Lewis, M., Haviland, J.M. (eds.) Handbook of emotions, 2nd edn., pp. 417–435. Guilford Press, New York (2000)
13. Petty, R., Cacioppo, J.: Communication and Persuasion. Central and Peripheral Routes to attitude change. Springer, New York (1986)
14. Eagly, A.H., Chaiken, S.: The psychology of attitudes. Harcourt Brace, Forth Worth (1993)
15. Schwarz, N.: Situated cognition and the wisdom of feelings. In: Feldman-Barrett, L., Salovey, P. (eds.) The wisdom of feeling: Psychological processes in emotional intelligence, pp. 144–166. Guilford Press, New York (2002)
16. Fiedler, K.: Affective states trigger processes of assimilation and accomodation. In: Martin, L.L., Clore, G.L. (eds.) Theories of Mood and Cognition: A User's Guidebook, pp. 85–98. Erlbaum, Mahwah (2001)
17. Dion, K.K., Berscheid, E., Walster, E.: What is beautiful is good. Journal of Personality and Social Psychology 24, 285–290 (1972)
18. Eagly, A.H., Ashmore, R.D., Makhijani, M.G., Longo, L.C.: What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. Psychological Bulletin 110, 109–128 (1991)
19. Dick, A., Dipankar, C., Gabriel, B.: Memory-Based Inference During Consumer Choice. Journal of Consumer Research 17, 82–93 (1990)
20. Ford, G.T., Smith, R.A.: Inferential Beliefs in Consumer Evaluations: An Assessment of Alternative Processing Strategies. Journal of Consumer Research 14, 363–371 (1987)
21. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11, 419–432 (2008)
22. STRATO AG Online Shops, `http://www.strato.de/shops/index.html`
23. Adobe Labs Kuler, `http://kuler.adobe.com/#`
24. Bradley, R.A., Terry, M.E.: Rank analysis of incomplete block designs, I. the method of paired comparisons. Biometrika 39, 324–345 (1952)
25. Luce, R.D.: Individual Choice Behaviours: A Theoretical Analysis. Wiley, New York (1959)
26. Nielsen, J.: Enhancing the explanatory power of usability heuristics. In: Proceedings of the ACM CHI 1994 Conference on Human Factors in Computing Systems, pp. 152–158. ACM Press, New York (1994)

27. Laugwitz, B., Schrepp, M., Held, T.: Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In: Heinecke, A.M., Paul, H. (eds.) Mensch & Computer 2006: Mensch und Computer im StrukturWandel, pp. 125–134. Oldenbourg Verlag, München (2006)
28. Bradley, M., Lang, P.: Measuring emotion: The Self-Assessment Manikin and the semantic differential. Journal of Behavioral Therapy & Experimental Psychiatry 1, 49–59 (1994)
29. Little, A.C., Burt, D.M., Perrett, D.I.: What is good is beautiful: Face preference reflects desired personality. Personality and Individual Differences 41, 1107–1118 (2006)

Educational Uses of the e-Book: An Experience in a Virtual University Context

Eva Patrícia Gil-Rodríguez[1] and Jordi Planella-Ribera[2]

[1] Universitat Oberta de Catalunya
Learning Technologies Office
Av. Tibidabo 45
08036 Barcelona - Spain
[2] Universitat Oberta de Catalunya
Psychology and Sciences Education Department
Rambla del Poble Nou 156
08018 Barcelona - Spain

Abstract. Despite the fact that the Digital Ink Book (DIB) has been developing over the past 10 years in different contexts, it has not made its presence felt within university education. This paper focuses on experimental educational use of the e-book in the context of a virtual university. Based on the available technologies, this research for technological innovation in education aims to carry out a pilot test with the students on the Digital Literacy and Exclusion course from a user centred design (UCD) perspective. Each student received a Digital Ink Book (DIB) for use in both the reading and working notes necessary for learning the different topics that make up the course. The data gathering was divided into a field diary kept by students participating in the DIB pilot test, a final report from each student about their own process of acquiring digital literacy in relation to the DIB, a virtual forum for the course, where the students commented on their difficulties and discoveries in using the DIB, and individual interviews with the students participating in the project once the experiment was finished. Through a content analysis of collected data we mapped the processes used by the students as strategies for the study of the university material using the DIB from a UCD perspective.

Keywords: e-book, m-learning, mobile HCI, digital ink book, human aspects of e-learning.

1 Introduction

At the Open University of Catalonia (UOC) all didactic material used is developed in digital format (XML, PDF and HTML), but it must be read from a computer (with the resultant effects of reading from a screen) or at best, as a printed version of the digital material (with corresponding paper wastage and ecological impact). Since the creation of e-Book e-Ink devices, it is possible to imagine accessing digital documentation through these highly accessible technologies and interfaces for pleasant reading. For many years now, there have been experimental prototypes and commercialised

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 55–62, 2008.

devices. For a number of reasons their implementation has been very slow and relatively unsuccessful (costs of the reading devices, availability of electronic books, high levels of security, size of the screens etc.). Following advances made by MIT Media Lab, so-called "Molecular Machines" have been created, that allow the implantation of that kind of digital reading technology. The research has been aimed at the protection of eyesight from the effects of reading digital texts (an aspect that was not taken into account in other digital e-Book technology prototypes) (Fuller et. al. 2002). This new micro-electromechanical technology enables highly “readable” reading.

Research projects centred on digital reading technologies (Hyper Book, Visual book, EXLIBRIS, Lecvtrice, Web book, EBONI, etc..), along with various commercial projects for the editing, sale and distribution of digital books (OZONeBooks, Alantic Bridge Publishing, BOOKS ONSCREEN TM, Ebook Express, Ebook Mall and Hypertech Media Inc, VanGoach Books, eBooks and N'Bytes, etc.) as well as advances made by e-Book reader manufacturers (particularly Amazon Kindle and Irex) have made it possible for reading a digital document to be little different from reading a printed one. Parallel to the technological research into the effects of digital reading on the reader’s vision, and the many possibilities for use and innovation, there have been studies into its application in a university context. The work of Shiratuddin (2005, 2003, 2001) into the use of e-Books in higher education has been a pioneer. Based on a comparison of the different e-Books found on the market, they study student's attitudes to the technology and the possibilities for its use, offering extensive examples of the advantages of the e-Book over the conventional book. From the multiple possibilities for the application of e-Books in university contexts, the author places special emphasis on the e-Books collaborative dimension and its application in distance learning, based on an online questionnaire that aimed to characterise some of the standard tasks that distance students develop when working with the e-Book. However, the actual implementation of e-Ink e-Books in university teaching is far from being a reality; that is why, with the current work, we want to go deeper into the processes of appropriation of the electronic ink book with a group of real students on the “Digital literacy and exclusion” course, looking into its educational uses.

2 Objectives

The objectives that follow from this research are:

1) To understand the processes undergone by the university students to organise their work using an electronic book.

2) To study the possibilities for use of electronic books in a virtual university context: annotations, underlining of text, management of annotations, uses of the dictionary, audio version, sensations in digital reading, portability of the device, management of the digital library, spaces for digital reading, notebook and collaborative work.

3 Methodology

Based on the available technologies, this project for technological innovation in education aimed to carry out a pilot test with the students on the Digital Literacy and

Fig. 1. Digital Ink Book used in the pilot test

Exclusion course, a part of the Educational Psychology (Psychopedagogy) degree. Each student received a Digital Ink Book (DIB) (with 768 x 1024 pixels resolution, 256MB of memory, 217mm x 155mm x 16mm dimensions and 389 grams weight) for use in both the reading and working notes necessary for learning the different topics that make up the course.

Some of the elements that made up the experiment were: ways of annotating documents, ways of underlining teaching texts, management of annotations and text markers, uses of the dictionary, uses of the other documents different from teaching materials, sensations, ease and difficulties of digital reading, how portable is the device, storage and management of the digital library, the spaces in which the reading takes place, uses of the e-Book notebook, possibilities for collaborative work with the DIB.

Monitoring of the pilot test into uses of the DIB took place on different levels and the focus was qualitative. The small number of students on the course (seven students) allows a precise and continuous monitoring of the students' relationship with the Digital Ink Book. The methodology of the test was divided into:

1) Field diary kept by students participating in the DIB pilot test. Each student noted down the most relevant information about the relationship (appropriation, rejection, etc.) that they maintained with the technological device we were experimenting with. The field diary recorded everything they considered worth communicating about their relationship to the device.

2) Virtual forum for the course, where the students commented on their difficulties and discoveries in using the DIB. This forum was facilitated by participants from the innovation project, and acted as a support service in the student's process of acquiring digital literacy in relation to the DIB.

3) Individual interviews with the students participating in the project once the experiment was finished – including a user test of the most basic functionalities of DIB – to collect and contrast the final data.

In order to understand the processes undergone by the university students to organise their work using an electronic book, and through a content analysis of collected data (field diaries, messages from the virtual forum and transcription of individual interviews) we mapped the uses of the DIB by students in the pilot test, and we list and study other categories developed by the participating students that were not initially foreseen.

4 Results

In this section we show the processes used by the students as strategies for study and work using the DIB, including other categories developed by them that were not initially foreseen.

In general the device has not been used much; the low levels of participation in the forum dealing with questions relating to its use – the activities of the virtual classroom were as in previous semesters focused on the content and concepts of the course at a purely theoretical level – as well as the field diaries and the interviews demonstrate this (clearly, due to the aims of the investigation, which was to study the "natural" appropriation of the device by the students, the teacher did not encourage nor oblige the students to use it at any point: or at least not directly, since the field diary on the use of the device formed part of the final evaluation of the course).

Nevertheless, and despite the fact that in the beginning all the students expressed enthusiasm on receiving the device, later all abandoned its use; principally they say it was due to "lack of time". Other motives mentioned were the slowness of the device, the fact that it is yet another unnecessary gadget with another charger, more cables, etc… and the lack of habit and/or need for its use:

"For me the device is difficult, the cables, the battery, the pencils, from the very beginning it was complicated for me, and furthermore, I didn't find a reason for it. I already have a tool that enables me to do things and now I am used to it. Changing to another without good reason..." (Interview 2)

Nevertheless, a more in-depth look at the body of analysis reveals that the principal motive for the low level of use of the device is its lack of interactivity. The device is only conceived as a device for reading documents and does not allow tasks needed for learning. We will return to this idea as it is key to the process of appropriation of new technologies applied to learning and to the conclusions of the current study. First we will map those processes of appropriation and use of the device, the advantages and disadvantages perceived, and the aspects related to its interaction, design and usability, and finally we will lay out the principal conclusions and recommendations of the research.

4. 1 Processes of Appropriation and Use of the Device

The mapping of the processes of appropriation and use of the device was carried out using the following axes of analysis: *making contact with the device, uses of the information, interactivity, use of the devices functions and educational uses.*

- *Making contact:* for all those interviewed it was a novelty to receive the device and the device itself was also new to them. All managed to turn it on and investigate and experiment with what they could do with it.
- *Uses of the information:* Of the seven people interviewed, five connected the device to their computer and passed or downloaded information, documents or books. Two did not get so far as connecting the device to the computer, so they did no pass any information. Of the five who connected the device to the computer, three downloaded the teaching materials for the course adapted for the device. One only downloaded the teaching materials and one only downloaded books. Of the two participants who did not connect the device to the computer, one only used it to make a few notes and the other did not do anything, except try out using it a bit.
- *Interactivity:* three of those interviewed tried unsuccessfully to connect to the internet. They had configuration problems they were unable to solve. Only three of those interviewed tried the USB option to read files they had stored.
- *Uses of the device's functions:* none of them had downloaded notes to the computer.
- *Teaching uses of the device:* of the seven people interviewed only three downloaded the teaching materials from the Digital Literacy course to the device. Only one did the first learning activity on the device and then abandoned it for the following activities. They used the notes for fun: as a way of playing, trying out, experimenting… The students did not perceive the DIB as a device that could help them study and work, and all of them described it as a reading device:

"I don't think it improves the quality of the studies, nor the performance of the student, I don't believe that" (Interview 3)

4.2 Perceived Advantages and Disadvantages

The principal disadvantages cited were the slowness of the device in getting started, the lack of colour (highly valued by the pupils for studying), the lack of a word search function (also very necessary for study). Logically, they also mentioned the problems with connecting to the Internet.

Positive aspects are: device comfortable to read, appropriated size and weight, the mobility, the capacity to store information and the fact that the use of the device could mean a notable reduction in the use of paper, with the resultant ecological consequences. The quality and comfort of reading the device compared to digital screens was mentioned. Finally, they highlighted the immediacy with which all the students in a class could receive information if it were all digitized.

4.3 Interaction, Design and Usability

A range of issues to be improved emerge in the results of the user test carried out. In terms of interaction with the device and its usability, the passing of pages in documents or books was basically done with the pencil that comes with the device, therefore without using the available buttons. Nevertheless, comments were made about difficulties in calibrating the pencil. Also, during the tasks three of the participants found that it blocked when they started to use it, although they did not say that this was an habitual occurrence. Only one of them knew where the "reset" button was. None of they knew how to view the notes passed to the computer, or that they appear in a specific folder. They also mention that the device is slow to start up, although it takes practically the same time as a computer to start.

The rest of the tasks proposed were easily carried out, which demonstrates that, in general, the device is well thought out in terms of usability; however there were criticisms – which emerge very clearly in the field diaries – of the lack of interactivity in the device, as it does not allow interaction with other functions or applications on laptops or mobiles that are central parts of virtual learning, such as connection to the Internet, the notes in word processor format, the same processors for the creation of activities, reports, etc…

This means that the DIB continues to be a too literally digital transposition of printed books, with the disadvantage – given the lack of calibration of the pencil that comes with the device, the lack of colour, ignorance of and difficulty in downloading notes to the computer, which, on the other hand, would allow them to be shared and contributed to by the teacher and even the other students – of not being perceived as useful for taking notes, a basic issue if the information contained in a book is not just to be read, but also transformed into knowledge for the student:

"When I work or write, I use colours to distinguish titles and subtitles, so that is lacking for me, right?" (Interview 4)

Effectively, it is the new forms of interaction with the rest of the students and with the teaching staff, made possible by new technologies, that shapes the new concepts of technologically mediated learning. As happened with the advent of the Internet and mobile technologies, these new forms of relationship and learning generate new needs when it comes to teaching and studying in a virtual learning environment; however, the literal way in which the Electronic Ink Book simulates the conventional printed book means that it does not generate a new need for its use among students:

"Really it is a question of necessity, if you really need to use something, I would have found or looked for a way of using it, used my initiative" (Interview 5)

For all the above reasons, we will now go on to offer strong recommendations for the design of functionalities and interactivity in Digital Ink Books, in order to adapt this technology to a virtual higher education context.

5 Conclusions

Although the device creates a similar initial impression to any new technological device, its use is later abandoned, principally because the students believe that this type of device does not improve the quality of the study or the performance of the student. They did not find it useful or practical. They see it as a gadget, a novelty that does not really meet a need; they don't see any clear use for it, apart from reading books. That is the image they have formed of the DIB: a reading device.

They therefore consider the device attractive and useful for specific moments and in specific places (on holiday, travelling, public transport, etc) where the computer is not practical, and, we are adamant in this, always as a device for reading, not studying. The students do not perceive the e-book as an adequate tool for study; therefore, to consider the device useful in the field of university learning, the following aspects should be improved on:

- Work should be done on ways of including or improving other functions necessary to the university student's learning process, principally in terms of being able to make notes on the documents. This would facilitate the reading of the different learning resources – such as teaching materials, articles or other books – using the device. Furthermore, the possibility to share annotations would be a considerable contribution to the field of collaborative learning.
- Other functions that should be included in order to make the e-book a study tool include a search engine, a word processor that goes beyond the notes documents that the device can currently generate, some kind of markers that could substitute colour (which currently is not possible in this kind of technology) and a full connection to the Internet.
- On the other hand, given the novelty and the limited penetration of this type of device in society there is a need for studied institutional communication when it comes to using them.

To finish, the device's lack of interactivity with other functions and applications lead to it being more an addition to the student's tasks than a way of facilitating them; that is why it was perceived as unnecessary for the learning process, and that made its appropriation unviable.

References

1. Bellaver, R.F., Gillette, J.: The Usability of eBook technology: Practical issues of an application of electronic textbooks in a learning environment (2002) (February 4, 2008), http://publish.bsu.edu/cics/ebook_final_result.asp
2. Crawford, W.: Why Aren't Ebooks More Successful? EContent 29(8), 44–54 (2006)
3. Díaz, P.: Usability of hypermedia educational e-Books D-Lib Magazine, vol. 9(3) (2003) (February 14, 2008), http://www.dlib.org/dlib/march03/diaz/03diaz.html
4. Fuller, S.B., Wilhem, E.J., Jacobson, J.M.: Ink-Jet Printed Nanopaticle Microlelectromechanical. Journal of Microelectromechanical 11(1), 54–60 (2002)
5. Garbhi, Z.: Analyse des pratiques de lecture sur livres électroniques chez étudiants universitaires. Tesi Doctoral, Montreal: Université de Montreal, Facultat d'Arts i Ciències (2006)
6. Gibbons, S.: Ebooks: Some concerns and surprises. Portal: Libraries and the Academy 1(1), 71–75 (2001)
7. Greenman, B.: e-Books: search & Downland in the Live Sciences (2007) (September 15, 2008), http://units.sla.org/division/dst/Annual%20Conference%20Contributed%20Papers/2007papers/1
8. Jacobnson, J., Comiskey, B., Turner, C., Albert, J., Tsao, P.: The Last Book. IBM Systems Journal 36(3), 457–463 (1997)
9. Hawkins, D.: Electronic books: Reports of their death have been exaggerated. Online 26(4) (2002)

10. Henke, H.: Electronic Books and ePublishing: a practical guide for authors. Londres. Springer, Heidelberg (2001)
11. Henke, H.: An empirical design for eBooks. Chartula Press, Niwot (2003)
12. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
13. Ojala, M.: Searching by and for the Book. Online 31(2), 49–51 (2007)
14. Peters, T.A.: Academic libraries take an E-look at E-Books. Spoon River College and Eureka College (2001) (February 4, 2008), http://www.geocities.com/lbell927/eBkFinal
15. Shiratuddin, N., Hassan, S., Landoni, M.: A Usability tudy for Promoting eContent in Higher Education. Educational Technology and Society 6(4), 112–124 (2003)
16. Shiratuddin, N.: e-Books ikn highder education:technology, e-marketing prospects and pricing strategy. Journal of Electronic Commerce in Organitzations 3(2), 1–16 (2005)
17. Thomas, S.: Another Side of the e-book Puzzle. Indiana Libraries 26(1), 39–45 (2007)

Construction and Evaluation of a User Experience Questionnaire

Bettina Laugwitz, Theo Held, and Martin Schrepp

SAP AG
Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
bettina.laugwitz@sap.com,
theo.held@sap.com,
martin.schrepp@sap.com

Abstract. An end-user questionnaire to measure user experience quickly in a simple and immediate way while covering a preferably comprehensive impression of the product user experience was the goal of the reported construction process. An empirical approach for the item selection was used to ensure practical relevance of items. Usability experts collected terms and statements on user experience and usability, including 'hard' as well as 'soft' aspects. These statements were consolidated and transformed into a first questionnaire version containing 80 bipolar items. It was used to measure the user experience of software products in several empirical studies. Data were subjected to a factor analysis which resulted in the construction of a 26 item questionnaire including the six factors Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Studies conducted for the original German questionnaire and an English version indicate a satisfactory level of reliability and construct validity.

Keywords: User experience; Software evaluation; User satisfaction; Questionnaire; Usability assessment; Perceived usability.

1 Introduction

Questionnaires are a commonly used tool for the user-driven assessment of software quality and usability. They allow an efficient quantitative measurement of product features.

Some questionnaires can under certain circumstances be used as a stand-alone evaluation method, for example the IsoMetrics questionnaire [1]. But in general, user questionnaires have to be combined with other quality assessment methods to achieve interpretable results (see e. g. [2]). In such a context, some usability questionnaires provide rough indicators for certain product features [3], while others are designed to discover specific usability problems (e. g. SUMI, see [4]).

In any case, the results have to be interpreted by a trained usability expert, taking into account also the results from other assessment methods that have been used.

The quantitative data from an assessment done by the users of a product can be a useful addition to methods that allow a sophisticated assessment of the strengths and weaknesses of interactive products, like for example usability tests or heuristic evaluation methods [5].

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 63–76, 2008.

A very effective way to get helpful feedback by end-users is to allow them to assess what concerns them most immediately: How did the interaction with the product feel, how was the use experience? This does not only include usability aspects as they are described by ISO 9241-10 [6] or the criteria of effectiveness or efficiency according to ISO 9241-11 [7]. The more fuzzy criteria that are subsumed under the concept of *user experience goals* [8] are an even more promising subject to a questionnaire assessment done by the users themselves. These criteria are for example reflected in the concepts of hedonic quality [9] or user satisfaction according to ISO 9241-11 [7] (for a deeper discussion on user satisfaction see e. g. [10]).

The objective of the construction process described below was to develop a questionnaire that allows a *quick assessment* done by end users covering a preferably *comprehensive impression of user experience*. It should allow the users in a very *simple and immediate way* to express feelings, impressions, and attitudes that arise when experiencing the product under investigation.

The available questionnaires lay emphasis on one or two of the mentioned criteria but none meets all three requirements. This paper contains an overview over the objectives, theoretical assumptions, and procedure of the construction process as well as the results of some validation studies investigating the quality of the questionnaire.

2 Construction of the Questionnaire

2.1 Objectives

Quick assessment: Generally, questionnaires are a particularly efficient method to apply and analyze. The application of some questionnaires may nevertheless be rather time consuming when the absolute amount of time is considered. With the SUMI questionnaire [4] the users have to decide on their level of agreement with 50 statements on usability. The long version of IsoMetrics [1] requires ratings for 75 different items. In these cases, the goal is to achieve a comprehensive usability evaluation including detailed descriptions of particular usability problems, on the sole basis of the questionnaire data. This is not what our questionnaire aims at. Rather, it is supposed to be an efficient tool to enhance the results from expert evaluations or usability testings.

Comprehensive impression of user experience: Traditional methods often focus on usability criteria in a narrower sense, which correspond roughly to the concepts of usability goals [8] or pragmatic quality [9]. More recent approaches increasingly give attention to the subjective reactions, also including emotional aspects of the user's experience, which can be subsumed under the concept of user satisfaction as outlined in ISO 9241-11 [7]. These criteria are also referred to as user experience goals [8], or as hedonic quality aspects [9]. A discussion of relevant usability criteria for special user groups, for example elderly persons, can be found in [11].

According to Norman [12] product design affects users on three levels of information processing, namely on a visceral level, on a behavioral level, and on a reflective level. This implies that usability criteria do not cover all aspects relevant for the user experience. This is also supported by studies (for example [13]) which show that there is a dependency between aesthetic impression of a user interface and its perceived usability.

It could be shown that semantic differentials for assessing the pragmatic and hedonic quality (e. g. [9]) are applicable not only to the evaluation of websites or games but also for business software [14]. However, this particular questionnaire (AttrakDiff2) lays a greater emphasis on the hedonic aspects of product quality than on the pragmatic aspects. This may not be perfectly appropriate for a comprehensive evaluation of professional software. A contrary perspective is represented by the SUMI questionnaire [4]. Here only one of six scales aims at the measurement of emotional aspects.

An overall picture has to include as many product aspects and features as possible that are of relevance for the user. For the new questionnaire no potential (hedonic or pragmatic) criteria should be excluded or favored a priori. The initial item pool should include a range of criteria as wide as possible, reduction and selection taking place on the basis of empirical data using an explorative factor analysis.

Simple and immediate: How does the interaction with the product feel? Which were the most striking features of the product and of the interaction? The user should be enabled to give his rating about the product as immediately and spontaneously as possible. A deeper rational analysis should be avoided.

The questionnaire should not force the user to make abstract statements about the interaction experience or remember details that are likely to be forgotten or had been overlooked in the first place. An explicit evaluation demanded by the user retrospectively is not always reliable (see e.g. [15]). This is supported by results [16] where differently colored UIs affected users' feelings differently (e. g. as measured with a mood questionnaire), while this difference was not reflected by users' answers on questions regarding the UI quality.

Experts are able to evaluate user interfaces in detail. Detailed data can also be gained from the observation of a user when interacting with the product.

Thus, a user questionnaire can lay its emphasis on criteria which are accessible immediately: the user's subjective perception of product features and their immediate impact on the user him/herself.

2.2 Theoretical Background

For the construction of our questionnaire we rely on a theoretical framework of user experience [3]. This research framework distinguishes between perceived ergonomic quality, perceived hedonic quality and perceived attractiveness of a product. The framework assumes that perceived ergonomic quality and perceived hedonic quality describe independent dimensions of the user experience.

Ergonomic quality and hedonic quality are categories that summarize different quality aspects. The focus of ergonomic quality is on the goal oriented or task oriented aspects of product design. High ergonomic quality enables the user to reach his or her goals with efficiency and effectiveness. The focus of hedonic quality is on the non-task oriented quality aspects of a software product, for example the originality of the design or the beauty of the user interface.

Thus, it is assumed that persons perceive several distinct aspects when they evaluate a software product. The perceived attractiveness of the product is then a result of an averaging process from the perceived quality of the software concerning the relevant aspects in a given usage scenario.

According to this assumption the constructed questionnaire should contain two classes of items:

- items, which measure the perceived attractiveness directly,
- items, which measure the quality of the product on the relevant aspects.

2.3 Generation of the Item Pool

Two brainstorming sessions (each lasting about one and a half hours) with fifteen SAP usability experts were conducted. The experts were asked to propose terms they suppose to be characteristic for the assessment of user experience. A moderator took down the proposed terms. The experts were asked the following questions:

- To which properties of products are users particularly responsive?
- Which feelings or attitudes of users are caused by products?
- What are the typical reactions of users during or after usability studies?

All redundant answers were removed from the list of the initial 229 expert proposals. All proposals that were not formulated as adjective were replaced by the corresponding best fitting adjective. The consolidated cleaned up list consisted of 221 adjectives.

Seven usability experts then individually extracted a "top 25" list out of the whole set of terms. In addition, they marked terms they considered to be inappropriate with a "veto" (unlimited number). Adjectives that received more than one veto or occurred less than twice in the top 25 lists were removed.

After this procedure a set of 80 adjectives remained. Since the target format of the questionnaire is a semantic differential, the best fitting antonym for each of the 80 adjectives had to be identified. The sequence of adjective pairs and the polarity of each pair was then determined randomly. In addition, a second version of the list with complementary order and polarities was prepared.

Both lists had the format of a seven stage semantic differential (another example of an application of semantic differentials in product design can be found in [17]).

We use a seven stage scale to reduce the well-known central tendency bias for such types of items. An example of an item is:

attractive ① ② ③ ④ ⑤ ⑥ ⑦ unattractive

2.4 Data Collection

In order to examine the specific properties of the adjective pairs concerning the assessment of software products, the eighty items raw-version of the questionnaire was used in six investigations. In the following, each of the six investigations is briefly explained.

- SYSTAT (number of participants *N*=27; location: University of Mannheim; paper-pencil version of the questionnaire): The participants of an introductory course for the statistics software package Systat were asked to perform a given task with the software or to observe a person that works on the task. After that the participants completed the questionnaire in order to assess the software quality.

- Cell Phone (*N*=48; University of Mannheim; paper-pencil): The participants of a psychology class were asked to add an entry to the address book of their cell phone and then to delete this entry. This application should then be evaluated with the questionnaire.
- BSCW (*N*=14; University of Mannheim; paper-pencil): Students rated the online-collaboration software BSCW that had been used during a lecture. Each of the participants had worked actively with the software before completing the questionnaire.
- Selection (*N*=26; University of Mannheim; paper-pencil): The participants of a computer-science course had the choice to assess one of the following products: Eclipse Development Workbench, Borland JBuilder, Microsoft Visual Studio, Mozilla 1.7 Browser, Microsoft Internet Explorer 6, and Firefox 1.0. Ratings were provided for Firefox 1.0, Microsoft Internet Explorer 6, and the Eclipse Workbench.
- CRM Mobile (*N*=15; SAP AG, Walldorf; paper-pencil): During a regular meeting of SAP usability experts, a user interface variant of the SAP Customer Relationship Management (CRM) software was demonstrated. The experts filled out the questionnaire after the demonstration.
- CRM PC (*N*=23; SAP AG, Walldorf, online version of the questionnaire): An online investigation consisting of a short demonstration of a further variant of SAP CRM and the electronic version of the questionnaire was conducted with SAP usability experts.

All in all, 153 participants provided complete datasets. 76 of the participants had completed the first version of the questionnaire while 77 had completed the second version (see above). Those data were used for the process of item reduction as described in the following section.

2.5 Reduction of the Item Pool

As described above the questionnaire should contain items that measure the perceived attractiveness directly and items that measure the quality of the product on the relevant aspects.

For this reason the item set was split into two subsets. The first subset contains 14 items that represent an emotional reaction on a pure acceptance/rejection dimension. These items of valence do not provide any information concerning the reason for the acceptance or rejection of the product. Examples for items from the first subset are *good/bad* or *pleasant/unpleasant*. The second subset contains the remaining 66 items from the item pool.

A factor analysis (principal components, varimax rotation) of the first subset of items extracted one factor concerning the Kaiser-Guttman criterion[1]. This factor explained 60% of the observed variance in the data. This factor is called *Attractiveness*. To represent this factor in the questionnaire we picked the six items with the highest

---

[1] If we apply the scree test [18] as a decision criterion to determine the number of factors also only a single factor results from the analysis.

loading on the factor. The original German items and their English translations can be found in Appendix 1 (for details on the translation procedure see chapter 2.6).

A factor analysis (principal components, varimax rotation) of the second subset of items extracted five factors. The scree test was used to determine the number of factors[2]. These five extracted factors explain 53% of the observed variance in the data[3]. We named these factors according to the items that showed the highest factor loadings as *Perspicuity* (examples for items: easy to learn, easy to understand), *Dependability* (predictable, secure), *Efficiency* (fast, organized), *Novelty* (creative, innovative) and *Stimulation* (exiting, interesting).

Per factor, we chose four items to represent this factor in the questionnaire. Those items were selected that had high loadings on the respective factor and low loadings on all other factors. The original German items and their English translations can also be found in Appendix 1.

All items that were not selected to represent one of these five factors were eliminated from the data matrix. The reduced data set was now again analyzed by a factor analysis (principal components, varimax rotation).

This analysis extracted again five factors according to the scree test. These five factors explained 70% of the variance in the reduced data set. The table containing the loadings of the items of the second subset[4] on these factors can be found in Appendix 2.

For the final questionnaire we randomized the order of the remaining 26 items. In addition the polarity of the items (i.e. the order of the positive or negative term per item) was randomized.

The final questionnaire contains thus the scales Attractiveness (six items), Perspicuity, Dependability, Efficiency, Novelty and Stimulation (four items each). We call this questionnaire in the following *User Experience Questionnaire* (UEQ).

To guarantee an efficient handling of data a tool (based on Excel) was developed that calculates the scale means and basic statistics from collected questionnaires.

2.6 Creation of an English Version

The basic version of the questionnaire was prepared in German language. In order to develop an equivalent English version, the following procedure was applied.

In a first step, the German version was translated by a native English speaker. The results of this first translation were checked by a group of native English speakers.

According to this feedback, a reworked version was created. The new version was translated back to German language by a professional translator (native German speaker). The differences between the re-translated German version and the original German version were examined and discussed with the translator as well as the native English speakers. Based on this last consolidation, the final English version was created. For first empirical data on the quality of the English version see 3.3.

---

[2] We choose the scree test since the Kaiser-Guttman criterion tends to extract too many factors in item sets that contain a large number of items. For our data set the Kaiser-Guttman criterion would lead to a solution with 13 factors.

[3] The variance explained by each factor is 28.7% for the first, 11.1% for the second, 5.3% for the third, 4.5% for the fourth, and 3.3% for the fifth extracted factor.

[4] The items representing the factor Attractiveness are not contained in the table. These items show, as expected, high loadings on all factors.

3 Validity of the Questionnaire

Concerning the validity of the questionnaire we are currently able to report data from two usability studies.

3.1 Validation Study 1

As described above the design of the UEQ fits perfectly into an existing research framework on user experience [3]. Perspicuity, Efficiency and Dependability represent ergonomic quality aspects. Stimulation and Novelty represent hedonic quality aspects.

The task oriented aspects Perspicuity, Efficiency and Dependability should show a strongly negative correlation with task completion time. The faster a user can solve his or her tasks with a software product the higher should be his or her rating concerning these ergonomic quality aspects.

On the other hand we expect no substantial correlation of the non-task related aspects Stimulation and Novelty with task completion time. We tested these two hypotheses in a usability test.

Participants. The 13 participants were recruited during the 2005 annual conference of the German SAP User Group (DSAG). They were not paid for their participation. All had high experience using computers, and experience with SAP software.

Procedure. The participants had to walk through a scenario that contained typical tasks of a sales representative. The scenario for the test was described to the participants in a step-by-step instructional document. The scenario contained a number of typical tasks a sales representative has to perform frequently during his or her daily job (plan customer visits, search for contact persons, find the last customer interactions, etc.). Each task was motivated by a little story, which explained the context of the task and why the task is performed.

Each test session was conducted as follows:

1. The participant was greeted and guided to the test station.
2. The moderators introduced themselves and collected basic demographic data.
3. The participant was given an overview of the test session and about the intention of the test.
4. The participant was then asked to solve the described tasks. The tasks description was available on paper during the whole session. The participant was instructed to think aloud during his or her attempt to solve the tasks.

After the participant finished the last task, the screen was turned off and the participant filled out the User Experience Questionnaire.

The screen was turned on again and the participant had the chance to discuss usability problems of the software and to ask questions.

The moderators asked follow-up questions related to the usability problems they observed during the test .

The total time required by participants to solve all tasks varied between 33 and 65 minutes (M = 41.62 minutes, SD = 9.64 minutes).

Results. Table 1 shows the correlations of the observed task completion times and the observed values of the UEQ scales. As a measure of scale reliability we give in addition Cronbach's alpha coefficient per scale.

Table 1. Correlation of the UEQ scales with the observed task completion times and Cronbach's alpha per scale

UEQ Scale	Correlation with task completion time	Cronbach's Alpha
Attractiveness	-.54	.89
Perspicuity	-.66 *	.82
Efficiency	-.73 *	.73
Dependability	-.65 *	.65
Stimulation	.10	.76
Novelty	.29	.83

* Significant with $p < .05$.

The correlations show the expected pattern. Perspicuity, Efficiency and Dependability show a significant correlation ($p < .05$) with task completion time. Novelty and Stimulation show only a weak correlation with task completion time.

Thus, our hypotheses do not have to be rejected. This can be seen as a first indicator for the validity of the questionnaire. The values of Cronbach's Alpha coefficient are an indicator for a sufficient reliability, but here we have to consider that the number of test participants was only small.

3.2 Validation Study 2

In a second validation study we investigated the relation of the UEQ scales to the scales of the AttrakDiff2 questionnaire [9]. This questionnaire was developed inside the above mentioned research framework from Hassenzahl [3]. It contains the scales Pragmatic Quality, Hedonic Quality (which is here split into the two sub-aspects Identity and Stimulation) and Attractiveness.

The concept behind the Attractiveness scales is nearly identical in both questionnaires. These scales should thus show a high positive correlation. In addition we can expect that the UEQ scales Perspicuity, Efficiency and Dependability show a high positive correlation to the AttrakDiff2 scale Pragmatic Quality. The UEQ scales Novelty and Stimulation should show a high positive correlation with the AttrakDiff2 scale Stimulation.

The concept behind the AttrakDiff2 scale Identity is quite different to the concept of any of the UEQ scales. For this scale we can thus not formulate any hypothesis concerning its dependency to the UEQ scales. We tested our hypothesis again in a usability test.

Participants. 16 students of the University of Cooperative Education in Mannheim, Germany, participated in this test. All had sufficient experience using computers. The participants were not paid for their participation in the study.

Procedure. The participants had to walk through a scenario which contained typical tasks in a CRM system (create a new account, create activities with the account, search for data of already existing accounts, etc.). The scenario for the test was described to the participants in a step-by-step instructional document. Each task was motivated by a little story, which explained the context of the task and why the task is performed.

The procedure for the test sessions was identical to the one for validation study 1 including the task completion step (step 4, see 3.1). After that, the sessions proceeded as follows:

5. Immediately after the participant finished the last task, the screen was turned off. Eight of the participants filled the UEQ and eight of the participants filled out the AttrakDiff2 at this point in time. It was randomly determined per participant to which of these two groups he or she was assigned.
6. The screen was turned on again and for around 30 minutes the participant and the moderator discussed about usability problems which were observed during the test session.
7. The participants that had already filled out the UEQ were now asked to fill the AttrakDiff2 and vice versa. Thus, each participant evaluated the tested user interface with the UEQ and with the AttrakDiff2 questionnaire. Since some of the items in both questionnaires are similar the delay introduced by step 6 is intended to reduce dependencies between the two evaluations.

Results. Table 2 shows the correlations of the UEQ scales with the AttrakDiff2 scales. The results show the expected pattern. The UEQ scales Perspicuity, Efficiency and Dependability show a significant correlation with the AttrakDiff2 scale Pragmatic Quality. The AttrakDiff2 scale Stimulation shows a high correlation with the UEQ scales Novelty and Stimulation.

The AttrakDiff2 scale Identity shows a high positive correlation with the UEQ scale Dependability, but no significant correlation with the UEQ scales Novelty and Stimulation.

Table 2. Correlations of the single scales from the User Experience Questionnaire and the scales of the AttrakDiff2 questionnaire

		User Experience Questionnaire (UEQ)					
		Attrac-tiveness	Perspi-cuity	Efficien-cy	Depen-dability	Stimula-tion	Novelty
AttrakDiff2	Attract-iveness	.72 *	.56 *	.30	.51 *	.51 *	.40
	Pragmatic Quality	.33	.73 *	.59 *	.54 *	.31	.07
	Identity	.45	.45	.29	.62 *	.30	.32
	Stimula-tion	.42	-.17	-.40	-.14	.72 *	.64 *

* Significant with $p < .05$.

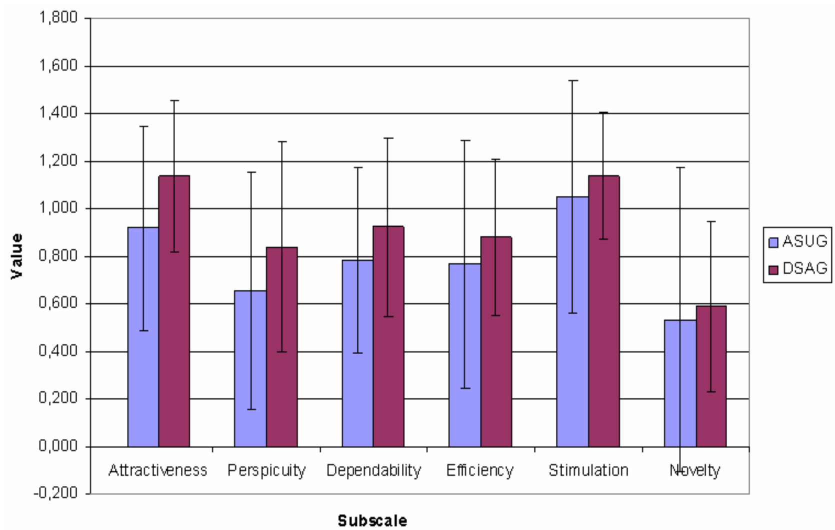


Fig. 1. Questionnaire scores from two parallel investigations. Investigation "ASUG" has been conducted at a conference of the American SAP User Group, while "DSAG" ran at a conference of the German SAP User Group. The raw data have been transformed so that the final data may range from -3 to +3. The error bars represent the 95% confidence interval for each arithmetic mean.

Thus, our hypothesis does not have to be rejected. This is again an indicator concerning the validity of the UEQ questionnaire. But again we have to mention that the number of participants in the study was small, so these results need to be confirmed in bigger validation studies.

3.3 First Data on an English Version

Though this has not yet been tested systematically, there are indicators that the language versions are sufficiently equivalent. For instance, two parallel investigations, one conducted in Germany and one in the US with the respective questionnaire versions delivered questionnaire scores as shown in Figure 1.

The one investigation was conducted at the 2005 fall conference of the American SAP User Group (ASUG), while the other investigation ran at the annual conference of the German SAP User Group (DSAG). The scenario and the SAP system were the same in both investigations; the only difference was the user interface language. The differences of the average scores on the different dimensions appear to be only marginal.

In another investigation, only the English version of the UEQ was used. This investigation was conducted as an online study with 21 participants who had tested a new software product for about one week. Each of the participants filled out the UEQ at the end of the testing period. In order to get an indicator for the reliability of the questionnaire, the Cronbach's Alpha coefficient was calculated for each of the subscales. Table 3 displays those values.

Table 3. Cronbach Alpha values for an investigation conducted with the English version of the UEQ

UEQ Scala	Cronbach's Alpha
Attractivity	.86
Perspicuity	.71
Efficiency	.79
Dependability	.69
Stimulation	.88
Novelty	.84

Except for the subscale Dependability, in each of the other cases the Alpha value exceeds the threshold of .7. According to this result, it may be assumed that the reliability of the English version of the questionnaire is sufficiently high.

4 Conclusions

For the construction of the user experience questionnaire UEQ the process should ensure that as many relevant product features as possible were taken into account. The factors revealed by the factor analysis support the assumption that ‚soft' (user experience) criteria and 'hard' (usability) criteria are of similar relevance for the end user (two scales and three scales, respectively). This fact is not reflected adequately by the structure of other user feedback questionnaires.

Studies reported here indicate a satisfactory level of reliability and construct validity. Data from the English and the German version of the questionnaire that have been collected in parallel studies confirm a good congruence of both language versions.

The user experience questionnaire UEQ in its current form appears to be an easy to apply, reliable and valid measure for user experience that can be used to complement data from other evaluation methods with subjective quality ratings. Nevertheless, further research will be done to provide a more detailed and extensive picture of UEQ's features from a methodical as well as from a practical point of view. In particular, the overall factor structure and the relative weakness of the "Dependability" scale will be in the focus of future studies.

References

[1] Gediga, G., Hamborg, K.-C., Düntsch, I.: The IsoMetrics Usability Inventory: An operationalisation of ISO 9241-10. Behaviour and Information Technology 18, 151–164 (1999)

[2] Dzida, W., Hofmann, B., Freitag, R., Redtenbacher, W., Baggen, R., Geis, T., Beimel, J., Zurheiden, C., Hampe-Neteler, W., Hartwig, R., Peters, H.: Gebrauchstauglichkeit von Software: ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11, Schriftenreihe der Bundesanstalt für Arbeitschutz und Arbeitsmedizin [Usability of Software: ErgoNorm: A method to check software conformity on the basis of DIN EN ISO 9241 parts 10 and 11, Institute Report Series of the BAuA]. Bundesanstalt für Arbeitschutz und Arbeitsmedizin, Dortmund, Germany (2000)

[3] Hassenzahl, M.: The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction 13, 481–499 (2001)
[4] Kirakowski, J., Corbett, M.: SUMI: The Software Usability Measurement Inventory. British Journal of Educational Technology 24, 210–212 (1993)
[5] Nielsen, J.: Heuristic Evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, pp. 25–62. Wiley, New York (1994)
[6] ISO 9241-10: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 10: Dialogue principles. Beuth, Berlin, Germany (1996)
[7] ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. Beuth, Berlin, Germany (1998)
[8] Preece, J., Rogers, Y., Sharpe, H.: Interaction design: Beyond human-computer interaction. Wiley, New York (2002)
[9] Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. [AttrakDiff: A questionnaire for the measurement of perceived hedonic and pragmatic quality]. In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer 2003: Interaktion in Bewegung, Teubner, Stuttgart, Germany, pp. 187–196 (2003)
[10] Lindgaard, G., Dudek, C.: What is this evasive beast we call user satisfaction? Interacting with Computers 15, 429–452 (2003)
[11] Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
[12] Norman, D.: Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books, New York (2004)
[13] Tractinsky, N.: Aesthetics and Apparent Usability: Empirical Assessing Cultural and Methodological Issues. In: CHI 1997. Electronic Publications (1997), http://www.acm.org/sigchi/chi97/proceedings/paper/nt.htm
[14] Schrepp, M., Held, T., Laugwitz, B.: The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers 18, 1055–1069 (2006)
[15] Nielsen, J.: Jakob Nielsen's Alertbox August 5, 2001: First rule of usability: Don't listen to users (2001), http://www.useit.com/alertbox/20010805.html
[16] Laugwitz, B.: Experimentelle Untersuchung von Regeln der Ästhetik von Farbkombinationen und von Effekten auf den Benutzer bei ihrer Anwendung im Benutzungsoberflächendesign. [Experimental investigation of the aesthetics of colour combinations and of its impact on users when applied to graphical user interface design]. dissertation.de-Verlag im Internet, Berlin (2001)
[17] Komine, K., Sawahata, Y., Uratani, N., Yoshida, Y., Inoue, T.: Evaluation of a prototype remote control for digital broadcasting receivers by using semantic differential method. IEEE Transactions on Consumer Electronics 53(2), 561–568 (2007)
[18] Catell, R.B.: The scree test for the number of factors. Multivariate Behavioural Research 1, 245–276 (1966)

Appendix 1: Original German Items and Their English Translation

Scale	Original German items		English translation	
Attractiveness	unerfreulich	erfreulich	annoying	enjoyable
Perspicuity	unverständlich	verständlich	not under-standable	understandable
Novelty	kreativ	phantasielos	creative	dull
Perspicuity	leicht zu lernen	schwer zu lernen	easy to learn	difficult to learn
Stimulation	wertvoll	minderwertig	valuable	inferior
Stimulation	langweilig	spannend	boring	exiting
Stimulation	uninteressant	interessant	not interesting	interesting
Dependability	unberechenbar	voraussagbar	unpredictable	predictable
Efficiency	schnell	langsam	fast	slow
Novelty	originell	konventionell	inventive	conventional
Dependability	behindernd	unterstützend	obstructive	supportive
Attractiveness	gut	schlecht	good	bad
Perspicuity	kompliziert	einfach	complicated	easy
Attractiveness	abstoßend	anziehend	unlikable	pleasing
Novelty	herkömmlich	neuartig	usual	leading edge
Attractiveness	unangenehm	angenehm	unpleasant	pleasant
Dependability	sicher	unsicher	secure	not secure
Stimulation	aktivierend	einschläfernd	motivating	demotivating
Dependability	erwartungs-konform	nicht erwar-tungskonform	meets expecta-tions	does not meet expectations
Efficiency	ineffizient	effizient	inefficient	efficient
Perspicuity	übersichtlich	verwirrend	clear	confusing
Efficiency	unpragmatisch	pragmatisch	impractical	practical
Efficiency	aufgeräumt	überladen	organized	cluttered
Attractiveness	attraktiv	unattraktiv	attractive	unattractive
Attractiveness	sympathisch	unsympathisch	friendly	unfriendly
Novelty	konservativ	innovativ	conservative	innovative

Appendix 2: Loadings of the Final Questionnaire Items on the Extracted 5 Factors

Items	**Factors**				
	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
confusing / clear	.661				
easy to learn / difficult to learn	.856				
complicated / easy	.851				
not understandable / understandable	.857				
usual / leading edge		.849			
dull / creative		.785			
conservative / innovative		.772			
conventional / inventive		.790			
demotivating / motivating			.601		
boring / exiting			.661		
inferior / valuable			.725	.422	
not interesting / interesting			.838		
obstructive / supportive				.505	
does not meet expectations / meets expectations	.438			.549	
unpredictable / predictable				.791	
not secure / secure				.740	
inefficient / efficient					.722
slow / fast					.723
cluttered / organized					.650
impractical / practical				.419	.635

Only loadings > .4 are shown in the table.

Integration of a Wiki for Collaborative Knowledge Development in an E-Learning Context for University Teaching

Birka Jaksch[2], Saskia-Janina Kepp[1], and Christa Womser-Hacker[1]

Hildesheim University, 31141 Hildesheim, Germany
Institute of Applied Linguistics (IfAS)
Information Science
[1] {keppsa,womser}@uni-hildesheim.de,
[2] birka_jaksch@hotmail.com

Abstract. E-Learning, which is technology-enhanced learning proved to have potential for adding value to education. Recently, with the shift of the World Wide Web to a more interactive web through the introduction of social software, the term E-Learning 2.0 was coined, indicating that the potential influence of social software on E-Learning has been recognized. This paper investigates the potential of integrating a wiki into an E-Learning module and the impact this integration can have on self-directed, collaborative and responsible learning in an E-Learning setting. Therefore, the focus is set on the integration of usability aspects and the surrounding didactic conditions that need to be considered when social software is to be embedded into learning scenarios. First evaluation results of the wiki-based glossary demonstrating this potential will conclude this paper.

Keywords: Social Software, Web 2.0, Wiki, E-Learning, Collaborative Knowledge Development.

1 Introduction

The explosive appearance of new web services and applications such as weblogs, wikis or other social software technologies shifted the World Wide Web "[…] from [a] static content-based web to a dynamic communication-based web." [1]. This trend, which is often referred to as the Web 2.0, is characterized amongst others by participation and user-generated content. As a result of this, in contrast to the Web 1.0 where users would only consume, social software applications enable anyone on the internet to create content, share information and connect with other users. This is expressed by [2] by characterizing the Web 1.0 as following a top-down approach while "web 2.0 takes a more bottom-up approach" where content creation doesn't require the use of code and new applications support "greater social interaction and collaboration".

This shift of the World Wide Web to a more interactive web through the introduction of social software has also led to discussions about E-Learning 2.0 indicating that the potential influence of social software on E-Learning has been recognized. This

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 77–96, 2008.

paper will investigate the potential of integrating social software into an E-Learning module and to what extent this integration can have an influence on self-directed, collaborative and responsible learning in an E-Learning setting. The focus of this investigation lies on the integration of usability aspects and the surrounding didactic conditions for the embedment of a wiki into a collaborative learning scenario.

Section 2 describes social software in general and wikis in particular along with their potential for E-Learning. Section 3 gives a detailed overview of the instructional design process framework for the integration of a wiki into a course about North American Culture conducted at the University of Hildesheim that participates in the CELEB project [3;4;5]. The empirical evaluation of the deployment with special focus on usability issues will be outlined in section 4 before an outlook concludes this paper.

2 Social Software in Education

Social software applications are web-based and therefore are accessible from anywhere at any time [6]. Hence, users have the option of working synchronously and asynchronously on the same content, which enhances online collaboration. Additionally, for the purpose of this paper, "[…] social software is defined as a tool that must meet at least two of the three following conditions:

1. It allows people to communicate, collaborate, and build communities online.
2. It can be syndicated, shared, reused, or remixed, or it facilitates syndication.
3. It lets people learn easily from and capitalize on the behavior or knowledge of others. " [6]

Furthermore, social software is characterized by their concentration on data that leads to rather simplistic interfaces with reduced functionality that lower the threshold for using the software [8]. As a consequence of these characteristics, the usability of this software is enhanced, because users are able to quickly understand and become capable of using the applications [6].

The following section describes wikis as one particular social software application and its potential for usage in educational scenarios.

2.1 Wikis

Wikis seem very helpful if there is a need for a platform that enables online collaboration and the creation of a knowledge base that can be accessed and amended by anyone on the World Wide Web. In fact, one of the first thoughts that probably come to mind in association with wikis is the free online encyclopedia Wikipedia[1], which has developed into a well-known online knowledge base based on wiki technology. Even though wikis were already in use long before the appearance of this free online encyclopedia, Wikipedia increased their popularity and made media pay attention to them [10]. Articles composed in more than 250 languages can be found on Wikipedia and as of June 2008 the English-language Wikipedia alone has accumulated 2,425,913 articles since it was launched in 2001 [11]. The concept of a wiki, openness as one of

[1] http://www.wikipedia.org

its biggest challenges, and the usage of wikis in education will be described in the following subsections.

2.1.1 Concept of a Wiki

The meaning of the word wiki refers to the main characteristic of this software application. Wikiwiki is Hawaiian and stands for quick. Ward Cunningham[2] coined this term and chose it because of the quick and easy changes that can be made to wiki contents [12]. While the online encyclopedia Wikipedia is well known to Internet users the concept of a wiki deserves some explanation and discussion. In general, one could say that a wiki enables creating a website where every user can get involved. [12] define wikis as "[…] web-based software that allows all viewers of a page to change the content by editing the page online in a browser." This means that every user can obtain author and editor privileges, which makes the wiki truly democratic [14]. Therefore, in contrast to weblogs where one author or a predefined group of authors owns its contents, "[i]n a wiki, no one person owns the content – and yet, everyone owns the content." [6].

A wiki consists of several interlinked web pages that can incorporate text, sound, images and videos. Apart from that, each wiki can look different, because a wiki has no predefined structure. Wikis therefore represent a multi-facetted repository, which may at the same time serve as a space for asynchronous as well as synchronous collaboration or collaborative activities. The changes that are being made are shown instantly. Since every user can post, edit or delete contents a version control is usually implemented so that malicious or incorrect changes can be made reversible.

The next paragraph will discuss openness as the main challenge in connection with wikis.

2.1.2 Openness of Wikis as a Challenge

The number one reason why wikis are so popular in the first place is also the number one reason for others to react critically to the concept of wikis: openness. The freedom of author and editor privileges, which also includes the right to delete or edit someone else's content, seems to be a major issue. For newcomers, the ingrained norm of authorship creates objections to wikis. "The notion of private property is so deeply embedded in our society that it's difficult to imagine going onto someone else's Web site and changing its content, even when you are invited to do so." [6]. This concern appears mostly with regard to public wikis that are open to all users on the World Wide Web. Nevertheless, also people who are working collaboratively toward a specific goal in closed groups need to overcome this barrier. More urgently, this openness creates a security gap for vandalism. Spammers are given the opportunity to maliciously delete contents or publish false information. One attempt to minimize this vandalism is some kind of version control in which all edits are made visible. This way, malicious changes can be made reversible. However, in communities that show strong commitment to their wiki, vandalism is usually removed quickly. "The community enforces behavioral norms so that the wiki doesn't become a free-for-all, an example of self-organizing group behavior in action." [6]. Still, if lack of interest and commitment prevail, the wiki concept will not work.

---

[2] Ward Cunningham developed the first wiki in 1995 [13].

Another concern in connection to a wiki's openness regards the quality of the contributed articles in wikis. Since there is no authority that approves and double-checks the content with respect to credibility and objectivity, "[m]istrust is directed at the question as to whether or not the masses even have enough knowledge to produce something trustworthy." [12]. Until now it seems that there is not a satisfying solution for this problem. However, the lack of a single authority that approves or rejects contents is exactly what allows a topic to be described from many different angles. It "[…] reduce[s] regional and cultural bias found in many other publications." [15]. Also, it enforces the quick appearance of articles that reflect thoughts about current events. In the long process of many changes, articles tend to represent the collective perspective in the end [14].

The last critical aspect of a wiki's openness is its lack of a predefined structure imposed from an authority that makes keeping track of the wiki's contents challenging. "The search function and the 'Recent Changes' page are useful, but without some sort of organization - just as with the World Wide Web as a whole - some pages may never be found, except by their author(s)." [6]. However, this does not mean that there is no possibility of creating structure in a wiki. It is the users' decision to add some sort of structure such as hierarchies, table of contents or assigned categories. Even though it might be challenging to maintain this structure since users can add or edit contents constantly, many existing wikis still incorporate structure for better navigation.

All in all, trust seems to be a possible solution to the challenges that arise with a wiki's openness. Restricting a wiki's openness leads to a reduction of its advantages like the possibility to describe certain aspects from different perspectives and the independence from imposed structures from a higher authority. Instead, the concept of trust should be embraced. In order for the wiki concept to succeed, the participating users need to overcome their ingrained norm of ownership and instead trust their peers in creating correct entries.

2.2 Wikis in Education

Despite the challenges, wikis are convenient tools for collaboration, collection and reflection and therefore are applied in numerous different ways at educational institutions [25]. They are considered a tool that "[…] facilitate[s] collaborative finding, shaping, and sharing of knowledge, all of which are essential properties in an educational context." [23]. Whether for assembling a syllabus, for collaboratively building an annotated bibliography or a knowledge base, for group authoring of documents and presentations [9] or for collaboration on projects [7] there is a variety of applications for the use of a wiki in education. In the literature many more examples can be found that offer simple guidelines on how and in which educational scenarios wikis can be applied [9; 35]. Other authors investigate further into the didactic aspects of using wikis in a collaborative learning setting, e.g. [7]. Then again, since wikis are information systems, there is also a need to consider usability aspects [25; 36; 37]. However, the didactics and usability of a learning tool or setting have not yet been analyzed in an integrated fashion.

The purpose of this paper is to integrate didactic as well as usability aspects in order to design a wiki as well as the appropriate learning scenario it is embedded into.

But first of all, this section will discuss why wikis have become more and more popular in academic settings but it will also point out the challenges that still remain to make wikis productive in that specific domain. The following section will then introduce the process design framework that is based on the ADDIE model and considers usability as well as didactic aspects.

One of the reasons why wikis are applied more frequently in educational institutions is the ease of use and the low cost. It is fairly easy to learn how to use a wiki, it is generally free to use for it is an open source application and does not take a lot of time to create new contents. Therefore the focus remains on the contents and not on the software itself. Wikis provide everyone with direct access, "[…] which is crucial in group editing or other collaborative project activities." [14]. As a result, all participants are equal in the sense that anyone can add or edit the contents at any time. However, it is possible to assign different levels of access and control to participating users in order to influence the teaching and learning experience. "A wiki's versioning capability can show the evolution of thought processes as students interact with the site and its contents." [14]. It therefore provides a visible state of the changes in dealing with a certain topic for students and teachers. Nevertheless, there is the possibility of creating different levels of access both to content in general, and to content editing functions. Sometimes these different levels of access are required, for instance, when the educators feel the need for a more controlled environment. However, [16] opposes that "[t]o really use a wiki, the participants need to be in control of the content – you have to give it over fully." In literature, it is emphasized that too much control in a wiki diminishes the effectiveness of this tool. counter

An invaluable implication for teaching and learning with wikis besides the opportunity of developing and enhancing general writing skills is "[…] teaching the rhetoric of emergent technologies." [17]. [18], a hypertext theorist, refers to this as "[…] network literacy: writing in a distributed, collaborative environment." Wikis provide an opportunity to enhance writing for public consumption. Working within a wiki enables anybody to review the contents and therefore the awareness of the authors is increased.

In addition, the wiki technology also enables the integration of multimedia such as podcasts. In this respect, podcasts (provided by news services such as CNN[3], etc.) could be embedded and aggregated in the wiki via RSS for discussions about current events in politics viewed from different perspectives. The wiki can be accessed at any time from anywhere, which ensures that the content can be updated constantly by the participating students and therefore guarantees topicality as well as authenticity. While the content of a course is emerging during the semester the students can expand the wiki's content autonomously. Over time a repository of shared knowledge develops, which "[…] facilitates writing as a process rather than a product." [19].

Added value also results from teaching information literacy, which is described as the "[…] ability to locate, manage, critically evaluate, and use information for problem solving, research, decision making, and continued professional development." [20]. In connection with the concept of wikis, which was discussed in section 2.1.1, information literacy plays an important role. Reviewing critically the peers' contributions that were added to a wiki, for example, improves the quality of content. [21] underlines that teaching information literacy especially in conjunction with what

[3] http://www.cnn.com

students learn in a course, motivates the students to learn these skills and enables them to become independent thinkers and lifelong learners.

Furthermore, the wiki can be used as a repository for learning resources, which would also enable the interdisciplinary addition, exchange and commenting of learning resources across institutions in a university context. This way the sustainability is ensured and the wiki can be used in different contexts. Thus, the sustainability of the wiki's content can serve as a source of information and knowledge. The wiki can be used continuously in the following semesters or by other institutions at a later time. The perspective of not only writing for the teacher but for a wider audience promotes collective authoring, which involves peer editing [19]. Especially with a wider audience in mind the students are required to critically read their peers' contributions in order to ensure authentic information. All in all, a wiki hosts an environment for collaborative and cooperative knowledge development since it enables all participants to work together on its content. The wiki does not only allow its users to absorb but also to develop knowledge actively and collaboratively.

One major challenge that remains for educators who work with wikis is tracking their students' work. Depending on the size of the course this might "[…] become a logistical nightmare." [17]. Anybody can edit at any time and as a result educators might face many changes in content. In addition, the attribution of individual work becomes very challenging as the contents are continuously changing. Changes made by individual students are again reviewed and edited by other students. Therefore, the end product will be the work of many different students. From there it might be difficult to assess which work was done by whom. "If people need to take credit for the things they write, a wiki is probably not the best tool." [6]. However, an approach towards the problem of attributing individual work is by having the students register. In that way the educator can identify the author of the recent changes being made [17]. Additional to the difficulty of taking credit for the users' individual writing is the issue of intellectual property. The anonymity of users and the lack of references turn this to an even more complicated matter [17]. Several policies have been introduced to address this issue, yet no approach has proven to be satisfying in all circumstances. Another issue often reported when introducing new technologies (in this case a wiki) to courses is the students' participation. Evaluations showed that the students' contributions made in course-based wikis are not sufficient if there are no incentives given [22]. This issue can be addressed, however, by creating incentives such as giving students credit points towards their exam for active participation.

As a consequence of these advantages and challenges, usability and HCI aspects like interface design, navigational and content structure as well as the embedment into the existing E-Learning infrastructure have to be considered in addition to didactic conditions like the different roles in learning contexts as well as the motivation of learners and teachers. The next section describes these facets as part of the instructional design process framework for the implementation of a wiki-based glossary in higher education.

3 Instructional Design Process Framework

According to the ADDIE model, which is a systematic approach to the instructional design process [24], there are five different phases that have to be considered in creating

E-Learning content. The first phase is the analysis phase that is followed by the design phase. The designed product will then be developed and implemented before it will be evaluated. The following subsections describe the essential steps of these different phases, beginning with the analysis of the basic conditions and how they have been carried out at the University of Hildesheim for developing a wiki-based glossary for integration in blended learning scenarios.

3.1 Basic Conditions

3.1.1 Target Group

The primary target group consists of students who are in their first year of a teacher training course for the subject of English. The number of students attending this course is estimated to be 20 and up. The students are approximately between 19 and 24 years of age and it is assumed that their previously obtained computer literacy skills are only basic. Through surveys carried out at the beginning of each semester these assumptions could be verified to a large extend. While all of the students possess a computer with access to the internet and use both nearly every day their usage habits are concentrated on basic applications like word processing software and communication tools like instant messaging and social networking platforms. Services like wikis or weblogs are mostly only consumed or not known very well and therefore not used. This information was obtained through a survey that was conducted at the beginning of the course.

3.1.2 The Course and Teaching Content

The course where the glossary is applied is based on a blended learning concept and consists of an introduction to the culture of North America focusing on the USA. The objective is to give an overview of North America (Canada, USA, and Mexico) relating to history, politics, culture, literature, educational as well as media systems. A focus is set especially on how these aspects collude within American culture. Current texts, audio and video files will therefore be used in order to discuss the consequential effects. As a result of the assumed little computer literacy on the part of the students, no advanced computer skills are required as prerequisites.

3.1.3 Objectives of the Application of a Wiki-Based Glossary

A general objective of all E-Learning modules created within the CELEB project is to improve the quality of teaching and to enhance the individual study as an additional support to the present teaching. In this respect, special attention is paid to the content development of the E-Learning modules in consideration of adding value facilitated by the integration of interactive multimedia. The E-Learning modules shall provide students with access to the learning resources independent of time and place and facilitate the process of learning from each other through knowledge sharing and knowledge creation.

The field of Cultural Studies is especially relevant in the training of students who want to become prospective English teachers. Constant changes in politics, social and cultural processes, which sometimes call for daily updates of the learning and teaching resources, are characteristic in the field of Cultural Studies. As a consequence, a learning environment, in which the content can be easily kept up-to-date, needs to be provided in order to ensure authenticity and topicality. By accommodating the changeability of learning and teaching resources the learning and teaching quality can be enhanced. In the context of the E-Learning Module *Cultural Studies*, the introduction of social software components and multimedia-based elements is intended to accommodate these requirements. Authenticity can be guaranteed, for example, by providing students with access to authentic audio recordings, which will give them the opportunity to listen to the pronunciation of native English speakers and therefore enhance the students' auditory understanding. Additionally, the benefit of learning independent of time and place creates a more learner centered approach. In respect to audio recordings, for example, the students can choose to repeat listening to them as many times as they need to fully understand. In addition, this will also contribute to their training in articulation. Topicality can be ensured by social software components, because they enable easy updating. In consequence, the content can be adapted to current events and supply the students with interesting newsworthy information. Expert interviews that have been recorded and show new developments in the students' field of study, for example, can be made available to them online in the E-Learning module. The module therefore allocates learning and teaching resources that are more current than those in textbooks.

Other objectives regard the development of certain competencies like social and communicative competence as well as information and media literacy in addition to the active knowledge construction about North American culture. The development and training of these competencies shall equip students with the necessary skills for the knowledge society that calls for life-long learning.

3.2 Usability Issues and Wiki Design

Due to the importance of authenticity and topicality as well as for the training of the above mentioned competencies it is important that the students don't just receive information but actively develop their knowledge. Therefore, collaboration between students is facilitated through social software according to the constructivist learning paradigm. With collaboration playing a key part in the learning process, the use of this wiki is based on the social constructivist learning model, which postulates that "[…] in order to learn, students have to create." [26]. This constructivist approach to the acquisition of knowledge emphasizes besides other interaction, reflection and exploration of knowledge, which is reflected in the five principles for the integration of social software in formal educational settings formulated by [27]:

- the process of learning is predominantly social,
- knowledge is generated through practice,
- learning needs active participation (responsibility),
- content is generated in cooperative learning situations,
- above all, communication structures need to be designed instead of content.

Based on these principles, the students are encouraged to interactively deal with problems, reflect their own and others' work processes and link information to a knowledge network.

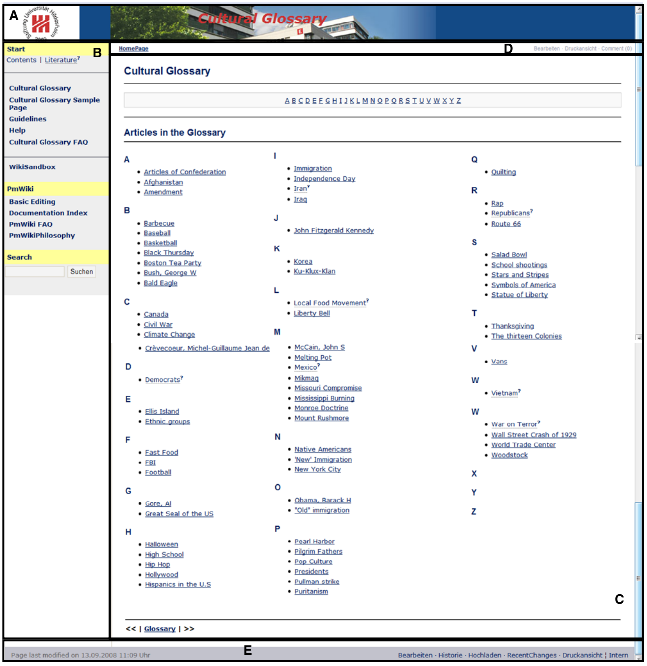


Fig. 1. Screenshot of the wiki-based glossary

Besides these design issues, which mainly describe course design, the design of the wiki interface also needs to be considered in this phase. In order to provide orientation to the users – students and teachers alike – a consistent interface is offered that consists of five central areas (Fig. 1). The header (A) is automatically filled with the title of the glossary that the educator has chosen. The functionality of the navigation bar (B) on the left is static according to the needs of the course whereas the menu items can be edited by the teacher at any time. The students only have editing rights in the main area (C) in the center of the screen although editing restrictions can be set here by the teacher or an administrator. Editing restrictions are employed by setting passwords for those parts of the wiki students should not edit. In addition, conditional markups facilitate adaptive hiding of links for the role of students, so confusion about non-executable functionality is prevented (Fig. 2 and 3). This is for example the case on a wiki field's start page that gives an overview of the field's contents and can only be changed by the administrator or the instructors of the classes a module is applied to. The areas (D, E) surrounding the main area contain this functionality to manipulate the main area like editing, printing, commenting or viewing the page's history.

By implementing a basic structure in the wiki the instructor provides additional orientation. For the wiki-based glossary the content structure is represented in alphabetical order in the first phase, but can be enhanced by categorizing the entries as the wiki becomes bigger through the student activity. Discussion sections regarding the different topics are tied to the created presentation pages in order to provide room for feedback, reviews or for coordinating the collaborative work.

Fig. 2. Wiki-based glossary without edit privileges

Fig. 3. Wiki-based glossary with edit privileges

According to the basic conditions and the design issues just discussed the wiki-based glossary is developed on the basis of the open-source software PmWiki[4]. The following section focuses on the implementation of this wiki into the course and the didactic factors like different roles and motivation as well as the technological infrastructure.

3.3 Implementation of the Wiki

Introducing a wiki as a new technology in educational settings does not necessarily guarantee to be a success by just implementing it. Section 2.1.2 already discussed objections towards the use of wikis in general and concerns referring to their use in academic settings. Therefore, for the successful use of a wiki in education certain prerequisites need to be considered before starting to actually work with it. These will be discussed in the following sections.

The students of the *Cultural Studies I – North America* course are requested to develop a Cultural Glossary throughout the semester, which is to result in an encyclopedic collection of information on different areas of North American cultural studies. The contents of this Cultural Glossary are oriented towards the topics addressed during the semester. Twice during the semester the students' task is to choose an aspect of the current topic (American history and American holidays and traditions) and write a glossary entry about it. The presentation of a certain topic can be a mixture of text, picture, audio or video files.

A detailed definition of the students' and the instructor's role will be given in the following before the issue of their motivation and the technological infrastructure that wiki is integrated into will be addressed.

3.3.1 Definition of Roles and Their Responsibilities

First of all it needs to be acknowledged that the students and the instructors embody different roles within the wiki. Whereas the role of the students will be that of active creators, authors and reviewers, the role of the instructor will be within the range of an observer, mediator, reviewer and coach. Within the wiki the students will create articles, review their peers' contributions and at the same time provide learning resources. [27] points out that in some respects while working with a wiki the students are required to become more autonomous and self-organized and as a result of this educators lose control over the content creation. However, in order to guide this

[4] http://www.pmwiki.org

learning process the educator will provide the students amongst others with specific learning resources and create initial page templates, which present simple guidelines[5] on working with the wiki referring to the course's content.

Since the target group constitutes of first year students of an introductory course and it is assumed that they are neither familiar with the contents nor have they developed collaborative skills yet, the educator needs to enable guided learning by giving structured instructions. In addition, it is advisable that the instructor points out the technical equipment that is made available to the students. In combination with using a wiki, which is web-based, this means pointing out the availability of computers as well as wireless access to the World Wide Web at the university.

3.3.2 Learner Motivation

A wiki is only successful under the condition that people actively participate. Several studies have dealt with the phenomenon of lurking, which means that most users of wikis don't actively contribute contents but instead only passively consume it [34]. Especially for the use of a wiki in an educational setting the participation of students is important and therefore needs to be motivated. That is why incentives need to be created in order to increase the willingness to work with the wiki and enforce a more active participation. According to [22], the instructor has to encourage the students' work in the wiki by being an active reviewer and leaving feedback so that the students feel the instructor's presence. This is of special importance for this target group since the wiki scenario is supposedly entirely new to them. In the case of assessment-driven students [22], an effective incentive is to make the contributions to the wiki part of the requirements for the course or even to reward the students' active contributions in form of additional credits towards the oncoming exam. Besides incentives that increase extrinsic motivation it is even more important to address the learners' intrinsic motivation by pointing out the usefulness of the wiki as preparation material for the final exam.

However, the difficulty of motivation is not only limited to students. Applying a new technology in a course also results in extra work and effort for educators, because functionality as well as the syntax of a wiki need to be learned. Furthermore, the educator's traditional role is expanded in the way that "[p]lanning lessons, a traditional hallmark of teacher expertise, need to be extended to designs." [28]. This means that the educator needs to design activities specifically for interaction within the wiki. Therefore essential conditions need to be created in order to point out the benefits of adopting a wiki to the educators. It is essential to present a wiki tutorial to the educators in order to motivate them to work with this technology. It needs to be discussed why technology can enhance their teaching and what the specific benefits are for the educators. Therefore it should be highlighted that the application of technology within a course serves "[…] to create opportunities for new objectives that may not be possible without them." [29]. One of these new objectives is possible because of the fact that wikis enable teachers to communicate easily and asynchronously with their students on course topics and thereby facilitate to "[…] quickly dispel misconceptions and correct errors […]" that occurred in class [30]. Furthermore, collaboration of

[5] Example for a guideline: "Add your entries in alphabetical order using the reference style specified in *How to work with this wiki*".

students is mostly a black box[6] process to instructors. In most cases the instructor is not able to review the actual collaboration process but only the output in form of a final product. Wiki technology reveals this collaboration process and makes it visible to the educator. Its version control features enable to track and assess the knowledge development of the students as well as monitor the content development in order to determine problem areas for students. In addition the educator can use the wiki to present course information and have the basics compiled by the students in it. This way, the focus of the traditional classroom sessions can shift to more topical issues. Concluding, it is important that the educator is supporting the concept of wikis because after all, the educators can only motivate the students to work with a wiki if they believe in the benefits of the wiki concept themselves.

An integral part of the implementation of the wiki-based glossary as well as other wiki-based applications in the CELEB project is the training of the instructors in using a wiki in order to get to know this kind of software and become aware of its advantages and limitations. The goal of this introduction is to familiarize the teachers with the technology and to enable them to train, support and encourage their students in using the wiki, because it can be assumed that educators are more encouraged to use a software in their teaching practice and are capable of encouraging their students to also actively use it, if they are familiar with it. In addition to the introduction, further support is supplied by written tutorials about introductory topics, article templates, FAQs and contact persons in case of technical difficulties. These support channels are accessible by instructors and students alike and can be supplemented by forum threads or chats if necessary. Further support through video tutorials by way of screencasts is planned. Because they enable and encourage instructors and students to use the wiki effectively, these support channels constitute further incentives.

3.3.3 Guidelines for Working with Wikis

Guidelines for working with wikis can further enhance the effective usage of wikis. Deriving from the objections towards wikis discussed in section 2.1.2 the following guidelines are addressed particularly.

First of all, an emphasis has to be put on the main characteristic of collaborative work. As discussed in section 2.1.2 the ingrained norm of authorship raises objections towards the work with wikis. It is essential to emphasize from the start that there will not be individual ownership of contributions and that the students need to be aware of that. Each contribution can be edited by every other participant and everyone is welcome to do so. To avoid contributions of lesser quality it is advisable to announce that even though it will not be possible to take individual credit for single contributions, the students' participation will still be monitored in order to give feedback on the development process of the content and in order to assess whether the students are eligible to earn credit points towards their exam [22]. Secondly, students are taught that the wiki concept depends on the constant changes made to its content. Therefore the students should be encouraged to contribute to a wiki page even though the presentation might not be the final version yet. A wiki enables the successive development of content. Thirdly, the participants are requested to review their peers'

[6] Black Box: "Black box is a technical term for a device or system or object when it is viewed primarily in terms of its input and output characteristics." (cf. Wikipedia: Black Box 2008).

contributions critically in order to improve the content quality. This means that in consideration of spelling mistakes, formal mistakes and mistakes as regards content students are invited to read through and edit their peers' presentations. Especially for the means of a glossary, a neutral style with preferably no bias needs to be established. Guidelines and techniques on objective writing are therefore of special importance. [12] advert to this as a process, in which "[…] striving towards objectivity is a form of self-education." Hence, an important pre-condition for collaborative editing is that guidelines with examples for objective language are given in order for the students to be able to express the "neutral point of view" [22].

The next section will give an overview of the technological infrastructure in which the wiki will be embedded.

3.3.4 Technological Infrastructure

The E-Learning modules created within the CELEB project are embedded in the E-Learning infrastructure of Stud.IP[7]. In the context of the project, the Stud.IP system presents a central access point for the administration of the courses and therefore for the E-Learning modules. As a consequence, access to the wiki-based glossary, which is part of the E-Learning module Cultural Studies, can be obtained via the user interface of Stud.IP. First of all, the students need to sign in to the Stud.IP system. They then are automatically directed to the interface, which lists their courses. By clicking on the link for the *Cultural Studies – North America* course they enter the course's website (Fig. 4).

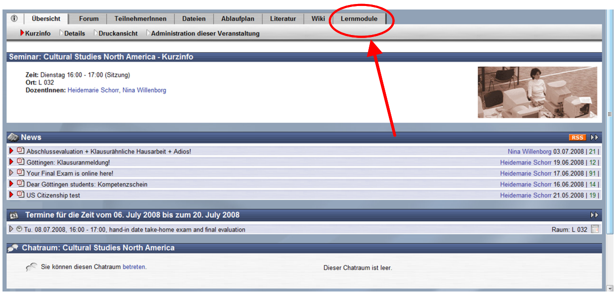


Fig. 4. Stud.IP Cultural Studies North America website

By choosing one of the learning modules listed under the tab *Lernmodule* the user is redirected to the wiki-based learning environment, which uses PmWiki software. PmWiki includes the basic wiki features and is easy to install but at the same time it is quite extensible and customizable. The software facilitates the inclusion of images in the wiki pages, the attachment of documents as well as the creation of links and therefore allow for "[…] efficient collaboration, knowledge-sharing, and resource-tracking." [31]. In addition, it enables the assignment of different levels of permission for different users and wiki pages by establishing passwords. This is especially useful in the context of the wiki-based glossary, since some of the pages, such as the start

[7] http://www.studip.de

page or pages that contain statistics about activity in the wiki, are not intended to be edited by students. Therefore, password protection for certain wiki pages is suitable. The decision for using PmWiki is based on the fact that it can easily be incorporated into the Stud.IP system.

In contrast to public wikis the wiki-based glossary will only be open for editing to students participating in this particular course in the first phase of the project. For a later phase of the project the wiki-based glossary can be involved in additional courses at the same time. This enables comprehensive collaboration over different courses as well as collaboration over more than one semester. When working within the wiki the students' contributions will not be signed with their name but instead with their identification issued by the university's electronic data processing center. This permits to overcome the inhibitions of public writing especially in the case of first year students. It guarantees a high degree of anonymity yet the contributions can still be attributed to the individual students.

3.3.5 Summary

The described design and implementation should help conquer the main objections that arise with a wiki's openness and were mentioned in section 2.1.2. One of these common objections to wikis concerned vandalism. However, the mechanism of version control facilitates monitoring the changing content of the wiki pages and easy restoration of deleted content. Furthermore, in the context of the Cultural Glossary wiki this issue does not fully apply since it is not open to all users. As of now, only students who are registered in the course of *Cultural Studies I - North America* are able to access and edit the wiki. Therefore, the number of users editing the wiki pages is limited. It can be assumed that the students are interested in a positive process while developing wiki pages since they are motivated by the credits they can earn for the final exam when actively taking part in the wiki (cf. section 3.3.2). Additionally, vandalism in the wiki-based glossary is restricted by having the participants sign in before accessing it. In this respect all contributions can be traced to individuals. Last but not least the awareness of the instructor's presence in the form of feedback will limit vandalism.

Another challenge in connection with the wiki philosophy concerns the traditional epistemology of individual ownership that students need to overcome. Research has found that students are hesitant editing each other's pages and that they primarily edit the contents they have created themselves since they feel individual ownership for these contents [19]. Another aspect is that according to the wiki principle users create links to non-existing topics indicating that there is still need for information. These links invite others to fill the articles with content depending on their knowledge. However, [32] points out that it will be difficult for new users to get used to this kind of collaboration as well as to use premature ideas as a suggestion for their peers. It can be assumed that this will also apply to the students using the wiki-based glossary and that it will remain challenging for them to get used to the process of collaborative authoring and open access. Therefore the challenge in respect to the educator's role is to constantly and actively encourage the students and to scaffold productive interactions among students in order to facilitate collaborative authoring. This involves amongst others commenting existing entries, designing activities such as peer editing and referring the students to the guidelines page, which contains principles on working with the wiki.

Another issue often reported is the motivation for using a wiki. Educators as well as students have to deal with a new technology within a small time frame and as a result of this have to put up with extra work. As a consequence this might cause reluctance to work with a new system. Although the technical hurdles of a wiki are few its use still needs to be learned. Especially first time users such as the students of the wiki-based glossary not only have to get acquainted with the wiki philosophy but also with the PmWiki syntax. [12] underline that "[t]he acceptance of wikis depends on the degree to which […] [the user] can truly benefit personally from using them." As a consequence, it is not only necessary to point out the wiki syntax, technical background and philosophy of the wiki in an early tutorial but also the beneficial areas, for which the wiki can be valuable. Therefore, training is provided to the educators who use the PmWiki software in their courses regarding technical and didactic aspects for the use of the software. In addition, technical assistance to support the wiki implementation serves as an incentive. Motivated educators who introduce new technologies in their courses then need to take over the role of a guide. In the context of the wiki-based glossary the educator gives an introduction to the wiki and its advantages such as the fact that the contents of the wiki pages can serve as a repository of learning materials and help the students prepare for the final exam and other incentives for contributing to the wiki (cf. section 3.3.2).

High relevance in regards to the issue of quality assurance exists for the context of the wiki-based glossary. Therefore it is necessary to introduce mechanisms such as peer reviews and feedback from the educator. This way closer attention is paid to the formal aspect of the wiki articles and its contents. The prospective that the wiki will be used, extended and promoted in different contexts in successive semesters and that it will be used by the students' peers as a starting point for their learning might also have influence on the quality of form and content. Under these circumstances it can be assumed that the students will work more thoroughly since the wiki will be made available to future students. However, making the wiki accessible to other users in a later stage of the project increases the exigency of adherence relating to copyright. A neutral style of the articles as well as academic research and writing needs to be postulated.

4 Evaluation Arrangements

Deriving from the objectives discussed in section 3.1.3 the focus of evaluation in the context of the wiki-based glossary is based on the question whether and in what way a wiki has an influence on collaborative learning and the active development of knowledge and competencies.

For this type of process evaluation online surveys and logging data are analyzed during and at the end of the semester. The online survey encompasses the overall opinion towards E-Learning in general and the wiki usage in particular, the quality of the technical design and technical problems as well as the motivation towards interacting with the E-Learning platform and with peers while working on the wiki-based glossary. The results of the survey that has been filled out by 22 out of 29 participating students, are compared with the logging data which show the actual overall participation. Participation data is distinguished between activity in the wiki itself and in discussions about wiki entries that take place in the discussion section of the wiki. Above that, the number of active students is examined. Other sources of information are the different support channels offered to the students in case of technical problems. Critical design issues as well as the willingness of the students to work with the wiki can be inferred from the number and kinds of problems that occurred while using the software. While students fill out surveys, instructors are interviewed in an unstructured fashion on similar aspects. These arrangements constitute a combination of formative and summative evaluation, since logging data and the usage of the support channels is analyzed during the semester while the survey is conducted at the end of it.

First results have shown that students are eager to try out this new concept and after some time they produce good results. After initial reluctance the activity as well as the number of authors has increased noticeably in creating wiki entries as well as commenting on peers' entries in the discussion section (Fig. 5).

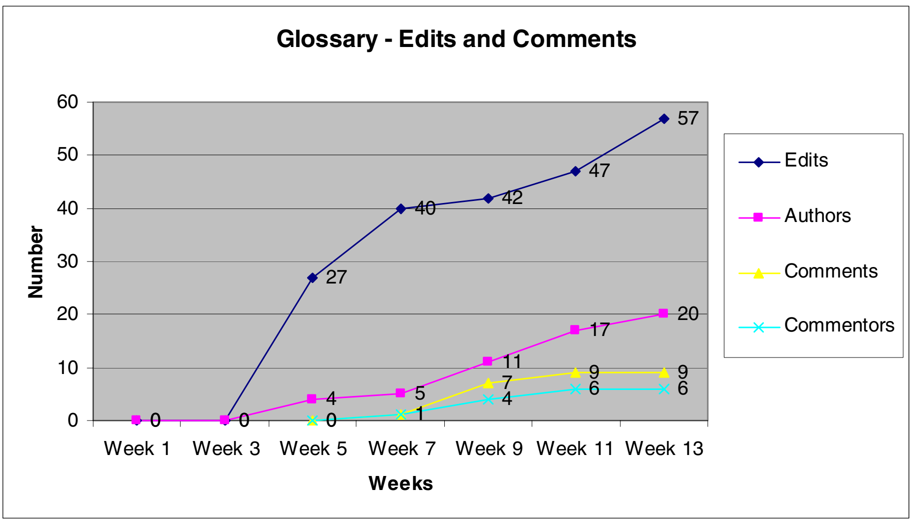


Fig. 5. Number of edits and comments in the wiki-based glossary

These numbers have to be put in relation with the minimum course requirements of two wiki entries per student which is accumulated to 58 required entries, because of 29 participating students. Comparing these figures with the actual number of 74 entries produced, about 28% extra work has been done. Therefore, it can be inferred that by creating a rather informal scenario in addition to the formal classroom phases, it is possible to enhance intrinsic motivation, which is necessary in order to be able to learn and develop skills [33]. From the survey results we know that students wished to work on the wiki more, but didn't have time to do so, because they were also required to prepare weekly tasks and meeting minutes for the course. This statement confirms the inference that intrinsic motivation has been addressed successfully.

Despite these positive results the logging data also revealed that students mostly worked on an entry by themselves and didn't collaborate with their peers. In another scenario where a wiki was incorporated into the same course in order to collaboratively write meeting minutes groups were build in advance. As a result the average number of authors in these group scenarios was 3.5 whereas in the glossary only 1.6 authors averagely worked together on an entry.

The usability of the wiki undergoes an iterative process that is accompanied by user tests. The first iteration occurred in parallel to the productive use of the glossary and has been supplemented by an expert review. Whereas the logging data and the survey results reveal the user acceptance of the wiki-based glossary the expert review pointed out a range of possible improvements to the interface and functionality of the wiki, which are attended to in the current second iteration. The main improvements concern the structuring of the wiki by setting passwords and separating open areas from those that are closed for students. Other aspects concern navigational support, a more visible offering of help documents as well as the handling of multimedia contents.

5 Conclusion and Outlook

In order to successfully integrate a wiki in an academic setting long-term an implementation strategy is required. This necessitates that all parties hitherto are included in this process. Not only the students need to be convinced of the new technology's benefits but especially educators are challenged. Also the educators have to deal with a new technology and therefore have to adapt their didactic concepts to the new circumstances in order to achieve added value. In this respect, this paper describes a concept and its realization for the integration of a wiki in an E-Learning module starting with an analysis of the target group as well as the basic conditions. Based on this, advantages for the use of a wiki were highlighted and an appropriate process framework for using a wiki was presented. This framework comprises of the instructional design phases applied to the integration of social software components exemplified by the integration of a wiki-based glossary at the University of Hildesheim as part of the CELEB project. Special attention is paid to the design and implementation phase where usability and HCI aspects need to be considered as well as didactical considerations such as the roles of the teacher and the learner and their motivation.

This paper presents a framework that aims at maintaining essential Web 2.0 attributes, such as trust, openness, voluntariness and self-organization, when applying Web 2.0 tools in institutional contexts [8]. The focus is put on usability engineering and the interface design of the wiki as well as on the necessary prerequisites concerning the implementation of the designed wiki. In order to further improve collaboration in scenarios without pre-definded groups advanced mechanisms will be investigated that can support the collaboration process. For the role of the teacher research is done for the implementation of support structures for constructing wiki scenarios and reusing them.

Acknowledgments. Our special thanks are devoted to our partners of the CELEB project as well as to the financial support from Lower-Saxony.

References

1. Downes, S.: Trends and impacts of E-learning 2.0 (last access: 2008-06-27), http://www.downes.ca/cgi-bin/page.cgi?topic=198
2. O'Hear, S.: Seconds out, round two. The Guardian (November 15, 2005) (last access: 2008-06-27), http://education.guardian.co.uk/elearning/story/0,10577,1642281,00.html
3. Kepp, S.-J., Schorr, H.: Interaktive und multimediale Lernmodule zur Gewährleistung von Authentizität. CELEB-Arbeitsbericht 1. Universität Hildesheim (2008)
4. Kepp, S.-J., Schorr, H.: Interdisziplinäre Entwicklung von E-Learning-Modulen. In: Lernen, Organisation, Gesellschaft. Das eCampus-Symposiums der Osnabrücker Hochschulen (logOS), Osnabrück, [to appear] (2008)
5. Kepp, S.-J., Schorr, H., Womser-Hacker, C., Lenz, F.: Chatten kann jede/r ;-). Integration von informellen Lern- und Kommunikationswegen und Social Software in ein Blended-Learning-Konzept für Lehramtsstudierende im Bereich Englische Kulturwissenschaft. In: Zauchner, S. et al. (eds.): Offener Bildungsraum Hochschule. Freiheiten und Notwendigkeiten. Medien in der Wissenschaft, Band 48, pp. 204–213. Waxmann Verlag, Münster (2008)
6. Farkas, M.G.: Social Software in Libraries: Building Collaboration, Communication and Community, Online. Information Today Inc., Medford (2006)
7. Jadin, T.: Social Software für kollaboratives Lernen. In: Batinic, B., Koller, A., Sikora, H. (eds.) E-Learning, digitale Medien und lebenslanges Lernen (Schriftenreihe E-Learning), Trauner, Linz, pp. 23–35 (2007)
8. Rollet, H., Lux, M., Strohmaier, M., Dösinger, G., Tochtermann, K.: The Web 2.0 way of learning with technologies. In: Uden, L., et al. (eds.): International Journal of Learning Technology, vol. 3(1), pp. 87–107 (2007)
9. Duffy, P., Bruns, A.: The Use of Blogs, Wikis and RSS in Education: A Conversation of Possibilities. In: Proceedings Online Learning and Teaching Conference, Brisbane, pp. 31–38 (2006)
10. Schwall, J.: The Wiki Phenomenon (2003) (last access: 2008-06-30), http://www.schwall.de/dl/20030828_the_wiki_way.pdf
11. Wikipedia: Special: Statistics (last access: 2008-06-27), http://en.wikipedia.org/wiki/Special:Statistics
12. Ebersbach, A., Glaser, M., Heigl, R.: Wiki. Web Collaboration. Springer, Berlin (2006)
13. Wikipedia: Wiki (last access: 2008-06-30), http://en.wikipedia.org/wiki/Wiki
14. Educause Learning Initiative: 7 things you should know about wikis (last access: 2008-06-30), http://www.educause.edu/ir/library/pdf/ELI7004.pdf
15. Wikipedia: Wikipedia: About (last access: 2008-06-30), http://en.wikipedia.org/wiki/Wikipedia:
16. James, H.: My Brilliant Failure: Wikis in Classrooms. Kaironews, May 21 (2004) (last access: 2008-06-30), http://kairosnews.org/node/3794

17. Lamb, B.: Wide Open Spaces: Wikis, Ready or Not. Educause Review Magazine 39(5), 36–48 (2004)
18. Walker, J.: Final Version of Weblog Definition. jill/txt, June 28 (2003) (last access 2007-10-11), http://jilltxt.net/archives/blog_theorising/final_version_of_weblog_definition.html
19. Guth, S.: Wikis in Education: Is Public Better? In: Proceedings of the 2007 international Symposium on Wikis, pp. 61–68. ACM Press, New York (2007)
20. Orr, D., Appleton, M., Wallin, M.: Information literacy and flexible delivery: Creating a conceptual framework and model. Journal of Academic Librarianship 27(6), 457–463 (2001)
21. Taylor, J.: Information literacy and the school library media center. Westprot CT (2006)
22. Cubric, M.: Wiki-based Process Framework for Blended Learning. In: Proceedings of the 2007 international symposium on Wikis, pp. 11–24. ACM Press, New York (2007)
23. Reinhold, S.W.: Augmenting Wiki Structure for Collaborative, Interdisciplinary Learning. In: Proceedings of the 2006 international symposium on Wikis, pp. 47–58. ACM Press, New York (2006)
24. E-Learning Center der Universität Zürich (Hrsg.): Instructional Design. ADDIE Modell (last access: 2008-06-27), http://www.elc.uzh.ch/elgrundlagen/kursentwicklung/konzeptionsraum/ instruktionsdesign.html .
25. Ebner, M., Kickmeier-Rust, M., Holzinger, A.: Utilizing Wiki-Systems in higher education classes: a chance for universal access? In: Universal Access in the Information Society. Springer, Berlin (2008), http://www.springerlink.com/content/01898k53u2041651
26. Schneider, D.K., Synteta, P., Frete, C., Girardin, S.: Conception and implementation of rich pedagogical scenarios through collaborative portal sites: clear focus and fuzzy edges. International Conference on Open and Online Learning. University of Mauritius (2003)
27. Baumgartner, P.: Web 2.0: Social Software & E-Learning. In: Computer + Personal (CoPers). Schwerpunktheft: E-Learning und Social Software 14. Jg (8), 20–22, 34 (2006) (last access: 2008-06-27), http://www.peter.baumgartner.name/article-de/social-software_copers.pdf
28. Lund, A., Smørdal, O.: Is there a space for the teacher in a wiki? In: Proceedings of the 2006 international symposium on Wikis, pp. 37–46. ACM Press, New York (2006)
29. Benson, D.E., Haney, W., Ore, T.E., Persell, C.H., Schulte, A., Steele, J., Winfield, I.: Digital Technologies and the Scholarship of Teaching and Learning in Sociology. Teaching Sociology 30(2), 140–157 (2002)
30. Bergin, J.: Teaching on the Wiki Web. In: Proceedings of the 7th annual conference on Innovation and technology in computer science education, p. 195. ACM Press, New York (2002)
31. Challborn, C., Reimann, T.: 47 Wiki Products: A comparison. International Review of Research in Open and Distance Learning 6(2), 1–5 (2005)
32. Bertram, C.: Online-Collaboration mit Wikis in Bibliotheken: Konzept zur Nutzung eines Wikis im Projekt "Lernort Bibliothek" des Goethe-Institutes Athen. Hochschule für Angewandte Wissenschaften Hamburg, Hausarbeit zur Diplomprüfung (2006)
33. Deci, E.L., Ryan, R.M.: Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology (25), 54–67 (2000)
34. Nielsen, J.: Participation Inequality: Encouraging More Users to Contribute. In: Jakob Nielsen's Alertbox: Participation Inequality in Social Design, October 9 (2006) (last access: 2008-09-14), http://www.useit.com/alertbox/participation_inequality.html

35. Dyckhoff, A.L.: Leitfaden: Einsatz von Wikis in Lehrveranstaltungen. Centrum für integrative Lejr-/Lernkonzepte. RWTH Aachen University, Aachen (2007)
36. Holzinger, A.: E-Learning Erfolgsfaktor Usability. In: Zauchner, A. (ed.) E-Austria Guide: e-Government, e-Learning, e-Health, e-Business (mit einem Vorwort von Wolfgang Schüssel), pp. 112–117. Linde Verlag, Wien (2006)
37. Ebner, M., Deutschmann, L., Zechner, J., Holzinger, A.: Qualitätssicherung durch Evaluation: Neue Ansätze zur Evaluierung selbstgesteuerter Lerntools. In: Sindler, A., Bremer, C., Dittler, U., Hennecke, P., Sengstag, C., Wedekind, J. (eds.) Qualitätssicherung im E-Learning. Medien in der Wissenschaft, Band 36, pp. 99–107. Waxmann Verlag, Münster (2006)

Web 2.0 and Social Software: Challenges and Complexity of Communication in Education

Helle Mathiasen and Lynne Schrum

University of Aarhus, Aarhus Denmark
hema@imv.au.dk
George Mason University, Fairfax VA USA
lschrum@gmu.edu

Abstract. This paper begins with an exploration of the changes that new Web tools and social software have fostered in communication in educational settings. It uses the framework of Luhmann to examine the complexity of these changes, and the potential to promote student centered learning. Through three case studies of student projects, the initial examination of an evolving educational investigation, results demonstrate the challenge for teachers to take on new roles, the reality that students' learning styles continue to drive their preferences, and the need for all teachers to understand more fully the possibilities and potential these tools offer for some students. The paper ends with a call for further research in this area.

Keywords: Web 2.0; social software; communication in education; teacher roles and functions.

1 Introduction

Technology in education has changed dramatically in the past few years as the world of the Internet has changed. It has morphed from static HTML pages where readers could find and copy information to interactive services, where visitors create and post information, and new social software expands the ways in which communication occurs between teachers and students and among students. This transition means that users work differently. No longer is it just 'find and use' information; now we have a more participatory, interactive format where everyone can create information collaboratively and share their results. Learners can now post journals, photos, movies, and more, thanks to an abundance of new collaborative tools and a multitude of Web sites. In fact, anyone can control content in a Web 2.0 world [52].

Students today are not the learners for whom our educational system was developed. Studies [25] show that students are capable, conscientious, concerned and optimistic, and as a generation, determined to succeed. Over 90% value school, plan to continue their education after secondary school, use computers and the Internet for school work and research, and spend more time using the Internet than watching television. Most pertinent to this paper, more than ¾ of them are creators of content on the Internet. Prensky [49] stated that today's students don't just use, but rather assume

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 97–112, 2008.

and expect technology, are always "on," are skilled multi-taskers, and significantly, embrace communities, collaboration and social networking.

The reality of these changes, and the changes envisioned for our learners are not based on technology alone. For example, Japan is "remaking its vaunted education system to foster greater creativity, artistry, and play. The Education Ministry has been pushing students to reflect on the meaning and mission of their lives, encouraging what it calls 'education of the heart" [48] (p. 53). Inspiring students to engage, perform, and become life-long learners is a central issue in education and thus a significant concern for educators.

In the future, "How we educate our children may prove to be more important that how much we educate them." [19] (p. 302). Thus, we begin this chapter asking what these changes mean for teaching and learning. Further, what does this mean for the educational system and for educators and their roles and functions? In what ways does planning, teaching, and interacting change in this new world? As educational leaders, we should understand changes in the Web and how they reflect changes in the world around us. Intellectually, Web 2.0 signals a transition from isolation to inter-connectedness for end users. The tools allow multiple users to participate: editing, commenting, and polishing a document rather than working alone and then publishing it.

This paper presents a theoretical framework from which to examine the implications of using new social software in the classroom. It looks at the interactional changes that may occur among the learners and between the learners and the teachers, and the ways in which every role may change. Three initial studies provide an examination of these changes, but also provide insight into the complexities of this type of research, and the tentative results that may be forever change our educational environments.

2 Lessons from Literature

Web 2.0 is an invented term, coined in 2004. It encompasses the growing collection of new and emerging Web-based tools. Many are similar in function to desktop applications, with people using their browsers for access rather than installing the software on computers. The tools are free and available to all, a change from applications that are purchased or licensed annually. Others are social in nature and promote self-expression—the social networks, blogs, wikis, photo and video sharing sites, and more [8, 52]. Anderson [1] defined educational social software as, "[...] networked tools that support and encourage individuals to learn together while retaining individual control over their time, space, presence, activity, identity and relationship" (p. 4). Fletcher, Tobias, and Wisher [17] suggest that the possibilities are endless, and state,

> …classroom teachers can locate and assemble instructional objects from the Internet or the Web for students to use individually, collaboratively, or under instructional guidance. Parents can access the same materials to see for themselves what students are learning in school or to pursue their own learning. (p. 97)

Social software tools include, but are not limited to, discussion forums, file sharing, chat, e-mail, weblogs, wikis, social bookmarking and RSS feeds. The term has not grown out of an educational discussion, and social software has not been developed specifically for learning. Anderson [1] introduces the concept of educational social software to initiate a discussion of social software in relation to learning. Social software represents a new approach to interactivity and communication. The focal characteristics of social software are flexibility and individualized use. Our approach, however, is different than that found within open and distance learning and lifelong learning. In our case, flexibility and individual use is not an institutional demand, bur rather a pedagogical principle, for which we will argue theoretically through the concept of self-organized learning environments [11] (38).

Other aspects of using these new technologies are worth consideration. Ebner and Holzinger [16], in an investigation of user-centered game based learning in a post-secondary environment, demonstrated that student engagement resulted in learning in a civil engineering course. They found that "joy" in the learning environment may encourage learning. Further, one study demonstrated support for the techno-social perspective, and found "Technology characteristics affect the ability of weblogs to create and deliver content, as well as to nurture social circles around them" ([14] (p. 796).

Salomon, Perkins, and Globerson [50] identified the key difference in learning with computers as entering into an intellectual partnership. The computer does not control the learning process. Rather, the learner enhances the computing capability of the technology and the computer augments the thinking and learning of the user. In this mode, students learn with computers rather than from them. Computers become "mindtools" that enhance abilities to construct knowledge by exploring, doing, conversing, and reflecting [21]. "Mindtools actively engage learners in creation of knowledge that reflects their comprehension and conception of the information rather than replicating the teacher's presentation" (p. 10). And although much of the focus on using ICT with students happens in groups, Jonassen [21] suggests that while learners may be part of a group, learning from each other and interacting in a social context, ultimately it is the individual who gains knowledge and skill. Lave and Wenger [24] add to this understanding by claiming that learning is a process of identity building. All such interactions provide "challenging experiences that require learners to rethink their understandings based on evidence from experience" [12] (p. 64).

Ausubel [3] introduced the ideas of "meaningful learning" in education. He stated that the most important thing learners bring to an instructional environment is what they already know. Meaningful learning occurs when an individual consciously and explicitly ties new knowledge to relevant concepts within his/her constructed knowledge. When this occurs, it produces a series of changes within our entire cognitive structure. Existing concepts are modified and new linkages between concepts are formed. Meaningful learning occurs when new, incoming information is connected to existing knowledge.

Ultimately, the creation of learning experiences that take advantage of the unique affordances of new technologies require educators, as key to any meaningful changes [4], to reconceptualize their new role. Berge [7] identified four primary types of communication that educators must incorporate when they are using technology. These include pedagogical, social, managerial, and technical. Assessment of learning should also take into account that "The purpose of interactive activities is to involve students in the kinds of analysis and synthesis processes essential for deep understanding and application" (Olgren, 1998, p. 88). The challenge is that much of our educational research has resulted in identifying barriers to implementation (Hernandez-Ramos, 2005; Norris, Sullivan, Poirot & Soloway, 2003; Sandholtz & Reilly, 2004). Doering and Veletsianos (2008-09) suggest that educational researchers must change their questions to better answer questions about teaching and learning:

> Rather than asking how many teachers know how to use technology, researchers need to ask how technology is used in the curriculum, classroom, and schools. Such investigations may yield fruitful knowledge as to how teachers decide to use the technology available to them, which in turn may help researchers understand the factors driving technology use, or lack thereof. (p. 102)

Authentic assessment has a high degree of similarity to real world tasks, which is exactly the type of situation available in the new world of Web 2.0 tools. Educators are more able to evaluate learners in critical thinking and problem solving based on thoughtful consideration of learners' outcomes [43, 44, 46]. We are finding exemplars in which students are creating complex podcasts (e. g., plate tectonics, glaciology, edaphology, limnology) for their colleagues as study guides; writing and sharing rap songs about elements of the Periodic Table; schools around the world now use these new tools to create online newspapers that now include stories, blogs, video/audio casts and more; and students authentically design solar passive cabins and study astronomy by examining the sky in real time. However, it is important to ask about the real changes in communication, roles of teachers and learners, and a larger educational perspective in this new world of social software.

3 Theoretical Framework

3.1 Complexity

In this theoretical framework, we are dealing with complexity rather than cause-effect relations. "Complexity means being forced to select; being forced to select means contingency; and contingency means risk." [29] (p. 25). Complexity is therefore considered a surplus of possible options:

> We will call an interconnected collection of elements 'complex' when, because of immanent constraints in the elements' connective capacity, it is no longer possible at any moment to connect every element with every other element. [29] (p. 24*)*

He explains further:

> …for each system the environment is more complex than the system itself. … There is, in other words, no point-for-point correspondence between system and environment… [29] (p. 25)

This paper concentrates on two types of systems: psychic systems and social systems. Both types of systems are based on meaning, implying that they choose to actualize something and leave other things alone. The individual system's unique selection decides what the system chooses to actualize and therefore communication and knowledge construction is “ruled” by contingency. Accordingly, the concepts and theoretical framework presented have consequences for the way the concepts of learning and teaching are defined. The specific characteristics concerning psychic systems and social systems are unfolded below.

According to Luhmann [26, 28, 29], systems are by definition closed, autopoietic (self-reproducing), self-referential and autonomous – and hence unique. In this sense the individual person, for instance, student as well as teacher, is unique. Each system, each person (seen as a system) constructs his/her own way of observing the world. According to the definition of systems, a system’s environment is specific to the system, which means that the environment is system-related. Each student and each teacher observes and constructs his/her own environment. In principle each system, hence each person, constructs their specific environment. In other words, insight lies in the eyes of the beholder. This insight has consequences for the approach to teaching and learning and to the educational system.

In this theoretical framework, knowledge is seen as a result of learning processes in which communication has a pivotal role. Communication constitutes the environment for conscious activities and for mental constructions. In other words, communication is seen as an essential part of an environment for learning processes. We introduce here social systems (maintained by communication activities) and psychic systems (maintained by conscious activities, including intuitions, emotions etc).

Communication promotes understanding, and understanding promotes construction of knowledge. This means that communication is the nutrition for learning processes and knowledge construction. Communication can ‘perturb’ consciousness and activate/maintain learning processes. The theoretical framework therefore offers an important point about the relationship between communication and processes of learning and construction of knowledge. Conscious activity and communication are mutually dependent. Social systems maintain themselves through communication. Thus, communication is the minor element in social systems, and this minor element is defined as the synthesis of three selections: the selection of information, the selection of utterance, and the selection of understanding [27].

All observation involves operations internal to the system, that is, the system characteristics. Thus, Luhmann calls his form of constructivism “operative constructivism” [26]. The communicator or the utterer – or the “alter,” as Luhmann prefers to call it – selects the information and utterance, and the addressee – the “ego” – selects the way in which the communication is understood [29]. Regarding the concept of observing as related to systems and its (self-constructed) environment, the utterer could be named the observed system and the adresseee could be named the observing system. Knowledge is constructed through operations of consciousness and related to the observing systems. Thus, knowledge constructions demand observation and selections. Luhmann writes that “Communicative success is the successful coupling of selections” [29] (p. 159).

A communication unit (see Figure 1) is for example, when a student pays attention to a lecture on video, reads a book or attends a lecture. One communication unit is actualized in this context when, for example, the students select understanding from the teacher’s uttered information. If the social system is going to maintain itself (e.g., teaching) the student and teachers continuously shift between utterance and selected understanding of the utterance and the action of a new utterance inviting to selection of further utterance. Here the point is that it is an iterative process, relating to both the utterer (observed system) and the addressee (the observing system). In other words the observing system and the observed system are switching roles in a continuous and iterative process.

This demands that the students, as observing systems and observed systems focus their attention on the communication, which might be the video, the book, the lecture, or a discussion. Communication, as well as conscious activities, is based on meaning; that is, social systems and psychic systems are meaning-based systems. "Meaning is a general form of self-referential adaptation to complexity […]," writes Luhmann [29] (p. 71).

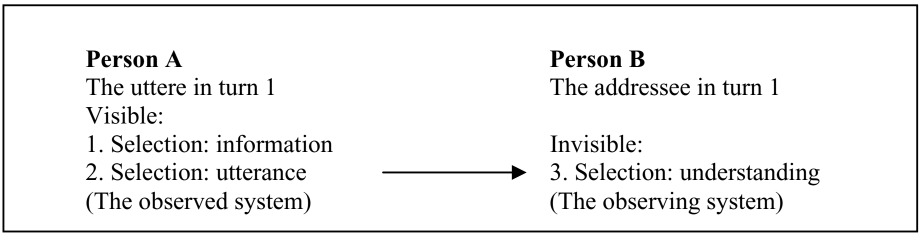


Fig. 1. One communication unit

Social systems as well as psychic systems are based on meaning, implying that they choose to *actualize* something and leave other things alone. Based on the system characteristics mentioned above, the result is that in principle the individual system's *unique* selection decides what the system chooses to actualize. In other words, all observations and selections are systems related.

> In every meaning, formulated positively or negatively, the three meaning dimensions are available as forms of further reference and as reflection tools. In general, the primary decomposition of meaning lies then in these three dimensions, and every thing else is a question of their recombination. [29](p.76)

The three dimensions are the "fact dimension," the "temporal dimension," and the "social dimension." Every operation relate to these dimensions. For example, a dialogue will always include a theme, a 'past-now-future,' and a social dimension. The communication will convey a theme over time and demand at least two persons, an utterer and an addressee (cf. Figure 1). In class the typical picture is a lecture and a group of students (e.g., groups between 15- 100 students).

Because learning depends on selections made by the psychic system (conscious activities), seen as autopoietic (self-reproducing), self-referential and autonomous which operate in and maintain themselves via conscious activities (e.g., thoughts, emotions or intuitions), learning cannot be organized. Learning develops from the selections of the psychic system. The question is how to develop a learning environment which supports and facilitates students' learning processes and knowledge constructions. That is, a class with 40 students in principle will result in 40 different ways of observing and therefore 40 ways of knowledge construction. Given that there is no overall solution, it becomes a question of how to meet this challenge in a variety of ways.

Social software in its broader sense gives some new possibilities and satisfies some needs if the possibilities are known. We term such a learning environment a self-organized learning environment. A self-organized learning environment is an environment in which students are encouraged to make their own selections and govern their (own) learning activities. If the point of departure is that these activities can be fruitful for knowledge constructions, the challenge is to empower students to organize their own learning activities as well to empower teachers regarding their didactic and pedagogical approach to the communication forums offered by Web 2.0 and social software products. Consequently, students' self-organized so called learning environment could be named a teaching environment, in the sense that the educational system in 2008 empowers this kind of organization of learning activities to the students. We must be mindful that all the communication activities are in one way or another linked to the educational system. Accordingly, we are dealing with students, teachers and specific educational aims, but in a reconceptualized framework that includes social software.

4 Three Cases

The three cases presented in this paper are first steps at examining the framework guiding the understanding and implementation of it in authentic classroom situations. Gaining insight into teachers' and learners' perspectives is essential if further research is to be designed; the use of technology in classrooms without such information will only result in future haphazard implementation efforts.

4.1 Institutional and Educational Framework

Case I (2004-2005) centers on three projects in which students mostly had to organise their project work themselves during one semester. The teachers referred to themselves as consultants due to the nature of their activities and functions. They considered the consultant role different from the traditional teacher role, associating it with functions like giving guidance and coaching. The students chose to use different types of web-based communication forums, such as their own group conference in the conference system, messenger, mail, and chat [30, 31, 32, 35].

Case II (2005-2006) centers on one three-month project in which the students had to organise their project work themselves during the last part of a semester. The teachers referred to themselves as supervisors and guidance counsellors due to the nature of their activities and functions [36, 40]. The students in Case II chose to use the same kinds of web-based communication forum as those mentioned in Case I.

Case III (2008) centers on a six month semester, where the students' learning processes as well as their group work was planned to be supported by different categories of podcasts (categories including five minute instructions/demonstrations, seven minute demonstrations including video streaming and power point presentations, and 45 minute lectures (talking head)) among other mentioned net-mediated communication forums (Mathiasen, in press).

In Case I the students ranged from 17 to 20 years of age (the age of upper-secondary school students in Denmark), in Case II the students were in their third year at university (in Denmark), and in Case III the students were in their fourth year at university.

4.2 Research Design

The case studies were based on a consecutive research design inspired by systems theory [33, 34]. The research projects made use of the following concepts and their possible connections: student and teacher roles, learning and teaching, and used a variety of data collection methods, including observations, individual/group interviews, print outs of the conference communication, questions asked in plenum sessions, and final written student evaluations [11, 47].

The research projects focus on student and teacher roles in various communication forums such as lectures (ordinary classroom interactions), problem-based self-organised group work, and web-based discussion forums and individual guidance. The Case III focus was specific to the students' use of and reflection on different podcast categories offered by the teacher. One category was named "talking head," which refers to the teachers' or a guest lectures' giving a lecture (typically 45 minutes) on video, a "one-way" and asynchronous category. In this, the students did not have the possibility to ask questions (written or oral) during the presentation of the video. Another category was named "demonstration/instruction" (typically five to seven minutes).This category was also a "one-way" and asynchronous category. The category was split up into a category A, where the teacher's voice supplied the video/audio stream (e.g., demonstrations of a practice context using a specific tool), and a category B where the teacher presented a PowerPoint and throughout it supplied comments as she/he went through the presentation, including giving references to literature, that would be helpful concerning the examination. Both the "talking head" and the "demonstration/instruction" category are meant as a learning resource and as a supplement to the traditional face- to- face teaching environment. Case III includes students' use of weblogs, wikis and the conference system offered by the university.

In all three cases the observed courses were organised as problem-based project groups combined with lectures, so in all three cases the teaching was organised as a mixture of ordinary classroom interactions and web-based communication. Data collected from these studies were examined through the theoretical lens, and using the constant comparative method [10], a systematic technique that employs various levels of coding to develop a 'grounded' theory of the phenomenon being studied.

4.3 Findings

The study demonstrates that the students involved in Case I were engaged in decoding the new roles and functions in respect to their status as upper-secondary school students, with everything this involves regarding what they should communicate about, with whom, and in which communication forums. In particular, the students focussed on the teacher's ability to gain insight into their academic abilities when the possibilities for communication are expanded with other forums besides the traditional class-based communication forum. Further, the study demonstrates that the students did not regard this as an important means of supporting their learning processes, but as a possibility for the teachers to have a broader basis on which to assess them (net-based written texts can be reintroduced in new/other connections). They thought it was easier to "hide" their academic skills and knowledge in traditional class-based teaching. For example, they could immediately take action by correcting a misunderstanding or incorrect answer, or just choose to be absent from class. Thus, the introduced concept of communication, the conditions for communications is decisive when the teaching environment (seen as a communication environment) is in a face-to-face context or a written net-mediated context. That includes which roles and function the teacher and the students are expecting from each other.

When the teachers were made aware of the students' scepticism about the teachers assuming the role of, for instance, consultant, and about the requirement of using conference forums, at first they were unable to understand the complexity of the problems. Later, these turned into eye-opening problems for the teachers. The performance of a "teacher" role and its related functions implies that the system makes demands on the students, which also applies, for instance, to the consultant role. For every type of role that the teachers choose to assume, the students themselves should, in principle, assume a corresponding role. This may be considered an obligation to actualize an increase in complexity, in that there are several possible student roles and functions that can be linked to several types of teacher roles and functions, such as conflict resolver, organiser and decision maker. In other words, combinations of the plurality of possible roles, functions, and communication forums are an indication of complexity that the system continuously attempts to reduce. The system actualizes the horizon of expectations more or less explicitly and hence also the framework of the communication in the broadest sense. The students tell about a problem, when the educational system expands the teaching environment (with social software) but they cannot simultaneously reflect and explicate new actualized expectations to the participants in the specific social system. In other words we are dealing with context dependent communication, which demands an explicit description of the different participants' roles and functions.

In Case II, the teachers observed that about 50% of the students experienced the project process as worthwhile. The students used the net-mediated communication forums and were self-organised during all three months. Most of the groups asked for continuous guidance meetings with one of the teachers. The other half of the students were observed to have the same problems as those described in Case I. They had difficulties finding their different roles and functions during the group project process.

The students' reactions can be categorized into two groups. One group observed they had developed and improved their academic and social qualifications through their use of a variety of communication forums and through the demands of the self-organised activities required by the project. In the future they would prefer to participate in courses organised as a variety of communication forums, each of which has its own advantages. The other group preferred traditional face-to-face class interaction with its well-known roles. This group of students preferred having a semester timetable and being informed by the teacher of the next homework schedules and assignments/marks. An overall question is here actualized: What does the educational system prefer as the outcome of an education? Do they wish to encourage students who have the ability to learn/construct knowledge in different and contemporary communication forums or students who prefer to be learners in safe environments, with well known roles and demands from the educational system?

In Case III where the focus was on the students' use of and reflections on podcasts as one of a broad spectrum of communication forums the students again divided into two groups. One sub-group prefers the face-to face communication in class, including the well-known roles as student and teacher. They want the possibility to interrupt "in real time" and ask questions which emerged during the teachers' lecture/demonstration/instruction. The other group (with the fewest number of students) thought it was a flexible way of "being a student." The time and place dimension was in focus in their argumentations. Some of the students in this group stated that they had a more or less indolent approach to their studies, and the available podcasts became a very comfortable way of being updated and for preparing for the exam. Other students in this group found that the podcasts were a comfortable way of preparing for the specific lesson.

Both groups found that the possibilities of replaying the podcast was an advantage and saved them from asking the teacher some of the questions they had. Nearly all students thought the podcast category "talking head" was a waste of time during the semester; however, nearly all the students recognized this specific podcast category as an option when they prepared for the exam. The category "demonstration/instruction" was not present when they prepared for taking their exam.

They assumed that if they could reproduce the "talking head," they would be more successful on the exam. Roles and functions have not in any significant degree been questioned when the focus is on the use of podcasts. Thus students and teacher roles seem not to create an increase in complexity. The system actualizes a well-known horizon of expectations, more or less explicitly.

4.4 Cross-Case Findings

It can be seen from these cases, and from other research [22, 54] that just as teachers differ in their skills and desire to employ these new forms of communication and learning, students also have individual reactions to learning in new ways and in taking advantage of the self-organized learning environments. It does require a more active role in learning, and many students will require knowledge about the new roles and functions, enhanced expectations, and personal learning strengths.

A larger concern can also be identified. The educational system needs to rethink the traditional approach to teaching and learning. With the reality of students' lives outside of the classroom, we are truly no longer dealing with the one-fits-all. Now with the development of different categories of social software, with the potential of serving each learning system, the educational system has an opportunity to provide for each learner's needs and expectations. It is possible to make a start toward this goal by identifying successful ways to implement net-mediated tools available for free. This change also requires that teachers have the professional development they need to understand and select the tools that support their instructional goals while recognizing the opportunity and challenges.

5 Social Software and Its Potential

We argue that other social software tools can help better support self-organized learning environments, and, thus, can supplement the use of asynchronous discussion forums and file sharing. Different social software tools can facilitate students' development of and engagement in networks that are not organized by the educational system. Further, social software supports students' use of resources not provided by the educational system with the intention of reaching specific goals of the educational system. Social software can provide students with opportunities for communication and social relations, which can empower them to develop self-organized learning environments.

In that respect, social software can be considered a supplement to the organized institutional setting. Use of social software is not determined by the educational system, and the communication supported by social software does not take place within the institution. Anderson [1] suggests:

> I think that Educational Social Software will find its real niche in self-paced, continuous enrollment, informal, lifelong learning and other more self directed forms of learning. Cohort based models of distance and blended learning boast LMS tool sets that were specifically designed for this institutionally driven context and the added value of social software is considerably smaller in less constrictive forms of learning. (n. p.)

The World Wide Web provides massive resources; their potential in relation to education is enormous, but the complexity provides a huge barrier. Search engines like Google do an impressive job, but it is still difficult for an individual to navigate the resources on the Web, and to value their relevance. If project work – which allows students to structure their own work process – is combined with the use of social software, it is perhaps possible to empower students to navigate the Web by actively using it as a major resource for their self-governed work.

Social software can provide tools for a personal and individual use of the Web which is based on social networks and communication. Social software enables a personalization and individualization of the Web. It can be used to develop relationships with people. Access to all the social software tools (weblogs, wikis, etc.) also means access to resources such as links to Web sites, papers, references, and so on. Using social software represents an alternative way to navigate the Web instead of only using search engines. Engaging in networks through public discussion forums and weblogs will enable students to find resources through people in their network and to engage in discussions not controlled by the educational institution.

Social software tools can be used to form networks or communities which can initiate self-organized project related discussions. Social bookmarking can enrich this network by providing students with a network of references from other people. These tools support development of social networks and can therefore facilitate communication and discussions. A wiki – or similar tools – can support the process of collaboration. At the same time, a wiki can be used to make the project available for other people. Such tools can support students in their self-governed work; for instance, they can easily collaborate on writing an assignment.

Weblogs in particular provide an example of the potential of social software. Weblogs differ from discussion forums and conferences in an important way. As opposed to discussion forums and conferences, weblogs are owned by the individual student. A study by Andreasen [2] shows differences in using a discussion forum or weblogs within the same course. He concludes:

> The learning potential that can be said to exist in the use of weblogs in relation to a course conducted over the internet relates partly to the increase in

> the students' opportunities for making their own voice heard, and partly to the active exchange with and reflection on other students' weblogs. (p. 86)

Andreasen [2] argues that weblogs support development of "individual voices." These individual voices are important to the development of students' independent use of the Web. Students can form their own networks which is different from participating in discussion forums within an educational setting. The result is what could be termed 'self-organized networks,' networks developed by students themselves. A similar, but more formalized approach is suggested by Koper [23] who uses the concept of 'learning network.' He stated, "Self-organised learning networks provide a base for the establishment of a form of education that goes beyond course and curriculum centric models, and envisions a learner-centred and learner controlled model of lifelong learning" (p. 1).

The use of social software suggested in this paper takes Koper's approach a step further. Self-organized networks are completely organized by the students without any influence from the educational institution.

6 Conclusion

This paper has provided an overview of the new Web 2.0 and social software as they relate to educational communications, roles and functions of educators and learners, and the possibilities for expanding the ways in which students learn and teachers establish educational environments. Looking through the theoretical lens of Luhmann [26, 27, 28, 29], it sought to provide an explanation of the complexity inherent in all communication, and in particular the communication of teaching and learning. The addition of new forms of social software provides an even more compelling need to understand the nature of the changes that may occur.

This paper did not address another aspect to the use of social software in educational environments and we are not ignoring the significance of it. We refer to the need for teachers to not only be willing to learn new roles and functions, but also the necessity for them to be familiar and comfortable with designing curricular experiences in this new milieu. Much research has identified the barriers, intrinsic and extrinsic to reaching this goal. These barriers include lack of time, lack of preparation, lack of technical, administrative, and pedagogical support, and individual resistance [5, 9, 41, 53, 55, 57]. Any changes in teaching and learning, or in roles and functions will require that we support and develop educators who are ready for using the technology effectively.

Through the three cases presented, we can see that introduction of social software and modified assignments works very well for some students at this time. However, the findings also point out the reality that the educational community can not assume that all students create, respond, and assimilate in the same ways. Differentiation must become a significant part of any educational environment in significant ways.

Clearly more research is required to begin to understand for whom this is the best way to learn, and for whom other, perhaps more traditional, methods are appropriate. According to Means and Haertel [42], all stakeholders in the educational system want to know more than if a piece of software or a web resource teaches something; instead, the system needs to know if the exposure to technology will ultimately have lasting value and impact on learners' lives. Most importantly, this paper has identified the need for significant research concerning the new net-mediated communication forms, and the related expectations for the participants in these activities. Frequently, researchers believe that they must do experimentally designed studies that use generic technologies or activities across the curriculum, and perhaps even across grade levels, instead of focusing on more deeply substantive questions about cognitive outcomes (see, e.g., [15]). Researchers and educators alike have pointed out that the effectiveness of the technology is tied directly to the efficacy of the instructional design, content, and teaching strategies employed by the teacher [19, 56]. Unfortunately, today this situation still exists. We believe that researchers must explore the characteristics of learners that promote success in these interactive settings, and return to helping learners understand more fully their individual strengths and challenges. It is also important to recognize the challenges inherent in conducting research in school settings and also teasing out the effects of a technology's use from all the other variables found in educational environments [6]. We must investigate teachers' roles and functions in this new environment and assist them in making the transitions necessary. As a community, we still do not recognize the affordances and constraints of these powerful tools, and when it is appropriate to use each type of social software. This paper has begun the dialogue and we challenge researchers to collaborate on creating a powerful research agenda for the future.

References

1. Anderson, T.: Teaching a distance education course using educational social software [Online] (last accessed 2008-07-15), http://terrya.edublogs.org/2006/01/02/teaching-a-distance-education-course-using-educational-social-software/
2. Andreasen, L.B.: Weblogs as forums for discussion – an alternative to the computer conference as a standard in online learning? In: Buhl, M., Holm Sørensen, B., Meyer, B. (eds.) Media and ICT – learning potentials, [Online], pp. 71–88. Danish University of Education Press (2006), http://www.dpu.dk/site.asp?p=8803
3. Ausubel, D.P.: The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology 51, 267–272 (1960)
4. Bai, H., Ertmer, P.: Teacher educators' beliefs and technology uses as predictors of preservice teachers' beliefs and technology attitudes. Journal of Technology and Teacher Education 16(1), 93–112 (2008)
5. Bauer, J.: Toward technology integration in schools: Why it isn't happening. Journal of Technology and Teacher Education 13(4), 519–546 (2005)
6. Bell, L., Schrum, L., Thompson, A.: Introduction and Overview. In: Bell, L., Schrum, L., Thompson, A. (eds.) Framing Research on Technology and Student Learning in the Content Areas: Implications for Educators. Information Age Publishing, Inc., Charlotte (in press)
7. Berge, Z.L.: Facilitating computer conferencing: Recommendations from the field. Educational Technology 15(1), 22–30 (1995)
8. Boster, F.J., Meyer, G.S., Roberto, A.J., Inge, C., Strom, R.: Some effects of video streaming on educational achievement. Communication Education 55, 46–62 (2006)

9. Butler, D.L., Sellbom, M.: Barriers to adopting technology for teaching and learning. Educause Quarterly 25(2), 22–28 (2002)
10. Corbin, J., Strauss, A.: Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative Sociology 13, 3–21 (1990)
11. Coulon, A.: Ethnomethodology. SAGE Publications, Inc., Thousand Oaks (1995)
12. Dhindsa, H.S., Anderson, O.R.: Using a conceptual-change approach to help preservice science teachers reorganize their knowledge structures for constructivist teaching. Journal of Science Teacher Education 15(1), 63–85 (2004)
13. Doering, A., Veletsianos, G.: Hybrid online education: Identifying integration models using adventure learning. Journal of Research on Technology in Education 41(2), 101–119 (2008-09)
14. Du, H.S., Wagner, C.: Weblog success: Exploring the role of technology. International Journal of Human-Computer Studies 64(9), 789–798 (2006)
15. Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al.: Effectiveness of reading and mathematics software products: Findings from the first student cohort (Publication No. 2007-4005) (last accessed 2007-04-27), `http://ies.ed.gov/ncee/pdf/20074005.pdf`
16. Ebner, M., Holzinger, A.: Successful implementation of user-centered game based learning in higher education: An example from civil engineering. Computers & Education 49(3), 873–890 (2007)
17. Fletcher, J.D., Tobias, S., Wisher, R.L.: Learning anytime, anywhere: Advanced distributed learning and the changing face of education. Educational Researcher 36(2), 96–102 (2007)
18. Friedman, T.: The World is Flat: A Brief History of the Twenty-First Century, Farrar, Straus, & Girous, New York (2005)
19. Glennan, T.A., Melmed, A.: Fostering the Use of Educational Technology: Elements of a National Strategy. RAND Corporation, Washington (1996)
20. Hernandez-Ramos, P.: If not here, where? Understanding teachers' use of technology in Silicon Valley schools. Journal of Research on Technology in Education 38(1), 35–46 (2005)
21. Jonassen, D.H.: Computers as mindtools for schools: Engaging critical thinking, 2nd edn., Merrill, Upper Saddle River, N.J (2000)
22. Kjær, A., Mathiasen, H.: IT: A challenge for the educational system. In: Dirckinck-Holmfeld, Fibiger, B. (eds.) Learning in Virtual Environments Samfundslitteratur, Copenhagen, pp. 142–165 (2002)
23. Koper, R.: Use of the semantic web to solve some basic problems in education: Increase flexible, distributed lifelong learning, decrease teacher's workload. Journal of Interactive Media in Education 6 (2004), `http://www-jime.open.ac.uk/2004/6/`
24. Lave, J., Wenger, E.: Situated Learning. Legitimate peripheral participation. University of Cambridge Press, Cambridge (1991)
25. Lenhart, A., Madden, M., Macgill, A.R., Smith, A.: Teens and Social Media: The use of social media gains a greater foothold in teen life as they embrace the conversational nature of interactive online media, Pew Internet & American Life Project, Washington, D.C (2007)
26. Luhmann, N.: Erkenntnis als konstruktion. Benteli Verlag, Berlin (1988)
27. Luhmann, N.: What is communication? Forum. Communication Theory 2(3), 251–258 (1992)

28. Luhmann, N.: How can the mind participate in communication? In: Grumbrecht, H.U., Pfeiffer, K.L. (eds.) Materialities of Communication. Stanford University Press, Stanford (1994)
29. Luhmann, N.: Social Systems (Trans. J. Bednartz Jr. & D. Baecker). Stanford University Press, Stanford (1995)
30. Mathiasen, H. (ed.): Udviklingsprojektet ”Det virtuelle gymnasium.” Følgeforskningsrapport 1, skoleåret 2002/2003, Bilagsmateriale, Copenhagen, vol. II. Undervisningsministeriets forlag (2003)
31. Mathiasen, H. (ed.): Udviklingsprojektet ”Det virtuelle gymnasium” Følgeforskningsrapport 2, Undervisningsministeriets forlag, Copenhagen (2004)
32. Mathiasen, H.: High school and learning environments. In: The Annual Meeting of the American Educational Research Association, San Diego, CA (2004a)
33. Mathiasen, H.: Expectations of technology: When the intensive application of IT in teaching becomes a possibility. Journal of Research on Technology in Education 36(3), 273–295 (2004b)
34. Mathiasen, H.: A Consecutive Research Design Inspired by System Theory. In: The Annual Meeting of the American Educational Research Conference, San Francisco, USA (April 2006a)
35. Mathiasen, H.: Project and web-based teaching. In: Buhl, M., Holm, B., Sørensen, B., Meyer, B. (eds.) Media and ICT. Danish University of Education Press, Copenhagen (2006b)
36. Mathiasen, H.: Undervisning og kompetenceudvikling. In: Mathiasen, H. (ed.) Læring og it, kompetenceudvikling på de videregående uddannelser. Aarhus University Press, Århus (2006c)
37. Mathiasen, H. (ed.): Læring og it, kompetenceudvikling på de videregående uddannelser. Aarhus University Press, Århus (2006)
38. Mathiasen, H.: Is there a nexus between learning and teaching? In: Holtham, C., Nygaard, C. (eds.) Understanding Learning-Centred Higher Education. Copenhagen Business School Press, Denmark (2008)
39. Mathiasen, H.: Podcast et fleksibelt medie? – casestudie af studerendes brug af og refleksioner over brug af ”on demand” midler. In: Bang, J., Dalsgaard, C. (eds.) Vidensmedier i videnssamfundet, Aarhus University Press, Århus (in press)
40. Mathiasen, H., Dalsgaard, C.: Self-organized learning environments. In: EISTA 2007 Proceedings, pp. 290–294. International Institute of Informatics and Systemics Press, Orlando (2007)
41. McKenney, S.: Technology for curriculum and teacher development: Software to help educators learn while designing teacher guides. Journal of Research on Technology in Education 38(2), 167–190 (2005)
42. Means, B., Haertel, G.D. (eds.): Using technology evaluation to enhance student learning. Teachers College Press, New York (2004)
43. Mehrotra, C.M., Hollister, C.D., McGahey, L.: Distance learning: Principles for effective design. Sage Publications, London (2001)
44. Naidu, S., Barret, J., Olson, P.: Improving instructional effectiveness with computer-mediated communication. In: Squires, D., Conole, G., Jacobs, G. (eds.) The changing face of learning technology, pp. 112–125. University of Wales Press, Cardiff (2000)
45. Norris, C., Sullivan, T., Poirot, J., Solloway, E.: No access, no use, no impact: Snapshot surveys of educational technology in K-12. Journal of Research on Technology in Education 36(1), 15–28 (2003)

46. Olgren, C.H.: Improving learning outcomes: The effects of learning strategies and motivation. In: Gibson, C.C. (ed.) Distance learners in higher education, pp. 77–95. Atwood Publishing, Madison (1998)
47. Patton, Q.M.: Qualitative Evaluation and Research Methods. Sage Publications, Inc., London (1990)
48. Pink, D.H.: A Whole New Mind: Why Right-brainers will Rule the Future. Riverhead Books, New York (2006)
49. Prensky, M.: Digital natives, digital immigrants. On the Horizon 9(5), 1–6 (2001)
50. Salomon, G., Perkins, D., Globerson, T.: Partners in cognition: Extending human intelligence with intelligent technologies. Educational Researcher 20(4), 2–9 (1991)
51. Sandholtz, J.H., Reilly, B.: Teachers, not technicians: Rethinking technical expectations for teachers. Teachers College Record 106(3), 487–512 (2004)
52. Solomon, G., Schrum, L.: New Tools, New Schools: Getting started with Web 2.0. International Society for Technology in Education, Eugene, OR, USA (2007)
53. Staples, A., Pugach, M.C., Himes, D.J.: Rethinking the technology integration challenge: Cases from three urban elementary schools. Journal of Research on Technology in Education 37(3), 285–311 (2005)
54. Stenalt, M.H., Jørgensen, D.S.: Using E-tivities, video dialogues and informal communication to motivate university teachers online. In: Proceeding of the 13th International Conference of European University Information Systems, EUNIS 2007. Grenoble Universities, France (2007)
55. Williams, H., Kingham, M.: Infusion of technology into the curriculum. Journal of Instructional Psychology 30, 178–185 (2003)
56. Winn, W.D.: Current trends in educational technology research: The study of learning environments. Educational Psychology Review 14(3), 331–351 (2002)
57. Zhao, Y., Frank, K.A.: Factors affecting technology users in schools: An ecological perspective. American Educational Research Journal 40, 807–840 (2003)

Using ePortfolios Enhancing for Learning through Computer-Mediated Interaction in a Course on HCI

Sonja Kabicher[1], Simone Kriglstein[2], Kathrin Figl[1], and Renate Motschnig-Pitrik[1]

University of Vienna, 1010 Vienna, Austria
[1] Research Lab for Educational Technologies
{sonja.kabicher,kathrin.figl,renate.motschnig}@univie.ac.at
[2] Department of Knowledge and Business Engineering
simone.kriglstein@univie.ac.at

Abstract. Web-services have established their role as facilitating tools for interactive and collaborative learning that supports social knowledge construction. In particular, ePortfolios are gaining importance in supporting the learning process and students' reflection. This motivated us to introduce and investigate an interactive mode in an academic course on Human–Computer Interaction (HCI) that adopted ePortfolio or diary-services, interactive spaces for team projects, and personal as well as interpersonal reflection. The qualitative content analysis of the ePortfolio entries illuminates those aspects in the learning process that were unique due to the computer-mediated interaction offerings in the course.

Keywords: Teaching Human–Computer Interaction and Psychology, ePortfolios, online reflection, online diaries, qualitative content analysis, assessment.

1 Introduction

The ePortfolio (electronic portfolio) is currently revolutionizing the world of learning [1] [2]. Particularly in the education sector, there are many initiatives in order to achieve the European goal "ePortfolio for all" by 2010. ePortfolios shall play the role of the lifelong learning companion that facilitates meaningful, self-directed and autonomous learning, capturing learning experiences, and planning personalized learning paths and careers.

In this study, we investigate the use of ePortfolios in an academic course on Human-Computer Interaction and Psychology (HCI and Psychology). The ePortfolio was part of the so-called "interactive mode" of the lecture which aimed to increase (compared to a conventional lecture) freedom, responsibility, awareness of students own learning, and decreases student's dependence on formal exams. This follows our strategy of applying/living humanistic and in particular person-centered principles within Technology enhanced Learning (TEL) [3] [4] [5] [6] [7]. Supplementary to a team-project, students were asked to reflect their learning experiences which they could collect individually and in teams by using the electronic diary functionality offered via the course's learning platform. In the context of this study, we analysed ePortfolios of 31 students by qualitative content analysis in order to find out whether the use of e-portfolios could enhance students' reflection on the input of the lectures

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 113–124, 2008.

and deepen students' learning experience. Furthermore, students' assessment of the e-portfolios and their way of dealing with the e-portfolios were analysed.

The paper is structured as follows. In the next section a brief insight into ePortfolio theory is given. Section three describes the context of our study, in particular the course on HCI and Psychology is illustrated. Section four includes the empirical study. The final part of the article gives a conclusion.

2 ePortfolios

Carmean and Christie [8, p. 34] define effective ePortfolios as "*a purposeful collection of student work that exhibits a learner's efforts, progress, and achievements in one or more areas.*" This ePortfolio definition is derived from Barrett [9], and comprehends an understanding of ePortfolio as a means of communicating growth made by the learner. ePortfolios are often used as a tool for assessment of learner performance, documentation of learner progress, and representation of outstanding learner achievements and records. Nevertheless, they can play a variety of other roles in educational environments (e.g. ePortfolios can be used for artefact creation and sharing, for setting personal learning goals, for practice using different types of technology to successfully solve required tasks, and for communication in order to provide feedback, for example from the teacher, peers, or outside experts, etc.). A very effective strategy in education is using ePortfolios for reflection [10]. In this use case, the ePortfolio facilitates learners' reflection on their learning process. The web-based services guiding and supporting learners in their learning processes manifest human-computer interaction in several ways: First, students interact with the services in being loosely accompanied by them in their reflection. Second, educators have easy access to students' elaborations and documents and, reciprocally, can send comments to students. Third, peers can mutually read each other's reflections and thus gain additional perspectives and inspirations. In brief, the computer significantly increases the interaction among all parties and thus facilitates social knowledge construction. On a course level, ePortfolios (or reflective diaries) facilitate students to document and observe own learning development and to control the fulfilment of their learning goals. Students can be supported in writing their diary by providing various questions that inspire reflection [11].

3 Description of the Course Human-Computer-Interaction and Psychology

The course on human-computer-interaction and psychology was held at the University of Vienna as part of the Computer Sciences bachelor curriculum [12]. The course is placed in the fourth semester of the curriculum and its first iteration took place in summer term 2008, which is, as well, the time period of our study. As the course is one of the required courses of the Computer Science core curriculum, 116 students participated.

The course was organized as a lecture with regular face-to-face units and two modes of assessment. Students could choose among the conventional written examination at the end of the lecture and the so-called "interactive mode" supported by a tutor. The goal of the interactive mode of the lecture was to offer students the possibility to learn and deepen own interests and knowledge in the area HCI and Psychology in an active and self-directed way during a lecture where usually regular frontal teaching takes place. It was an additional mode, which ran parallel and in a linked way to the weekly lectures. The motivation to offer the interactive mode in the lecture was to facilitate individual learning with

- learning experiences and self-initiated processes,
- use of HCI for practical tasks,
- reflection of subject-specific and personal learning,
- more social interaction,
- connection of education and research for new expertise (optionally), and
- less exam stress.

The 31 students, who participated the interactive mode of the course, had to elaborate a team-project and write entries into an accompanying ePortfolio, which we called "diary".

First of all, students had to sign up for the interactive mode at the very beginning of the lecture. Then, students were asked to build small teams (2-3 members per team) and elaborate project proposals. These project proposals were reviewed by the tutor. Project goals, output and working plan were discussed in a personal meeting with the team. If the project proposals met the requirements, expectations, desired learning goals and outcomes of both, the educator and the team, then they acted as online learning contracts between the educator and the students [13] [14]. Using learning contracts helped us to facilitate freedom and self-initiated learning in the lecture. Students could define and follow their own learning plans and targets with a substantial degree of responsibility for their own learning [15]. They could explore and elaborate topics of interest in an active, self-directed way and thus deepen their knowledge in a special application of the course's subject matter. These activities were accompanied by the educator by giving feedback to students in online and face-to-face meetings to approve the self chosen themes. After reviewing and approving the project proposals, the teams elaborated contributions and deliverables (e.g. a project specification, project documentations and project output deliverables like videos, learning objects, slides). During this work phase, three work-in-progress inspections were done, in which particular submissions of the teams were reviewed and evaluated. In the last unit of the course, students presented their team-projects. This allowed all students to gain some insight into the interesting topics and results elaborated by the participants of the interactive mode of the course. In addition, presenters learned through experience how to mediate HCI aspects in a brief slide presentation.

Simultaneously, participants of the interactive mode reflected their individual and team learning in an ePortfolio (electronic diary). The diary was technically realized as a Blog which was offered as extended functionality on the learning platform CEWebS

(Cooperative Environment Web Services) [16] CEWebS provides a web-service-based architecture for cooperative environments. Web services are the central component of CEWebS in order to put blended learning scenarios into practice. (To read more about the CEWEbS architecture see [17, 18] [19]). Students could submit their diary entries with read permission for either all participants of the course, for only the facilitator, or for nobody in order to write diary entries for their personal use. Writing the diary was organized in four phases (derived from the “Writing-Survey-Writing Cycle” [20] [21]) during the course:

- **Phase 1: Personal state of knowledge and expectations.** Phase one started right at the beginning of the course and students were asked to write about what they thought they already know about HCI and Psychology, what experiences they made in this area and what their attitudes are concerning HCI and Psychology. Furthermore, we asked students to formulate their expectations on the HCI and Psychology course.
- **Phase 2: Continuous reflection.** Phase two immediately followed phase one and included questions concerning experiences and impressions of our students in the HCI and Psychology course, team work and team learning experiences, activities and experiences in the area of HCI and Psychology that went beyond course requirements (e.g. observations in the internet, working environment; literature research; etc.)
- **Phase 3: Experiences collected in the team-project and in HCI and Psychology lectures.** The question pool of phase three concentrated on experiences collected in the team-project and in the HCI and Psychology module. Some exemplary questions were: What worked successfully in the project? What would you change in the project in order to optimize HCI and Psychology aspects? What did you take home with you from the module/course HCI and Psychology, and what not? For what can you use your collected experiences from the HCI and Psychology module/course?
- **Phase 4: Final reflection.** Phase four was the last phase of the diary. We used this phase to ask students about their impressions to the diary, experiences with writing such a diary, the offered tool function, the recommended amount of words per diary entry. Finally, we asked students to write some recommendations for students participating the course/module HCI and Psychology in the coming term.

We suggested students to write around 100 to maximal 300 words, and not to exceed a time period of one hour per entry per week.

Before the final written exam took place, the instructor and tutor evaluated students’ achievements in the interactive mode. In the interactive mode, students could in the best case obtain the grade C (satisfactory). To improve their grade (to B – good, or A – very good), students had to pass the final written exam. The points obtained from the exam were added to the points collected throughout the interactive mode. The HCI and Psychology course scenario is illustrated in Fig. 1 and Fig. 2 shows Person-centered e-Learning patterns which we used.

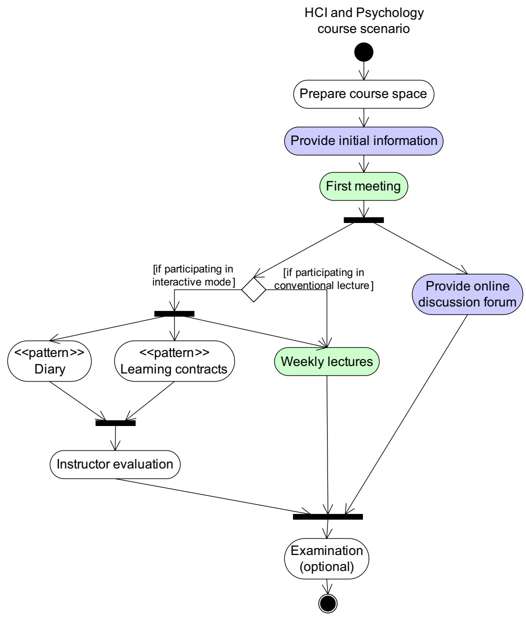


Fig. 1. Illustration of the HCI and Psychology course scenario as an activity diagram using coUML notation and Person-centered e-Learning patterns proposed by Derntl [22] [23]

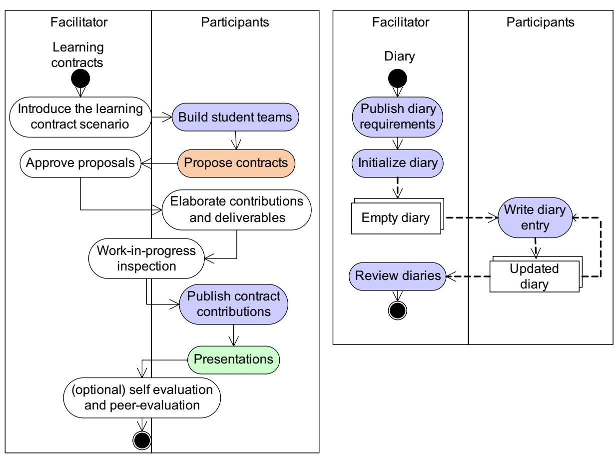


Fig. 2. Illustration of the Person-centered e-Learning patterns "learning contracts" [14] and "diary" in CoUML notation [23]. Colours of the activities represent different modes of presence: a green activity means that it takes place in face-to-face (or present) mode; a blue activity takes place on the web; a red activity illustrates that the activity takes place in a blended (mixture of face-to-face and web) mode; a white activity communicates that the mode of presence is not specified.

4 Empirical Study

Text and qualitative content analysis were applied to evaluate these reflections in the style of Mayring [24].

The primary goal of using a qualitative content analysis was to investigate students' reflection about their learning development and how the teaching method can support the understanding of HCI. Therefore, the following research questions were investigated in the qualitative content analysis:

- How did students experience the interactive mode (e.g. positive and negative feedback)?
- Could they recognize connections between the learning input of HCI and their life?
- Did the use of ePortfolios enhance students' reflection on the theoretical input of the HCI lectures?
- How did students assess the use of ePortfolios?
- What did students share in the ePortfolios?

4.1 Sample

The interactive mode was proposed for the first time in the bachelor studies of Computer Science. From 116 students who took the course 31 (~ 27%) participated in the interactive mode, divided into 13 teams (1-3 students per team). About 64% of the participants were male (20) and 36% female (11). About 74% of the participants successfully finished the interactive mode and the dropout rate was 26%. Students who dropped out the interactive mode claimed about too little time to meet all requirements and too much workload of the mode. All of the 31 participants of the interactive mode wrote online "diaries". Consequently, 31 online diaries with a total word count of 48763 could be analyzed. Students' online diaries were to be provided on the course's e-learning platform.

4.2 Results

First we analysed the situation in which the text is originated and its characterisations. In order to provide inter-subjectivity, categorisation was undertaken by two researchers. Finally, we verified, merged and interpreted the essential categorisations. Table 1 shows the final categorisations.

In the ePortfolios - or "online diaries", as most students named them, students wrote about a variety of topics, often closely related to the guiding questions for the reflections. They wrote about the interactive mode, their team project, the teamwork, the lectures, the lab courses and about connections between the theoretical input of the lectures and experiences in their everyday life. Often students wrote about the lectures (19 statements), they narrated the theoretical inputs (23 statements), reflected on them and gave feedback on lectures (18 statements) and specific lecturers (6 statements). In general, for most topics, statements addressing these three different levels could be found:

- **Narration:** students memorized what they could remember e.g. from a guest lecture; example for narration of the lecture: "Today we discussed the short term memory (how to adjust to the context of the user) and the long term memory."
- **Reflection:** students made connections between things they had learnt in the lectures and things that they encountered in everyday life or their previous knowledge; additionally they critically questioned input of the lectures); example of reflection on the lecture: "Another interesting point brought up in the lecture was the use of metaphors; thereby consistent symbols are used for the representation of functionalities. Some symbols result from the beginning of visual computing history, as for example the symbol of floppy disks, which still is used almost in all programs as a symbol for saving, although floppy disks are - if used at all - used very seldom. A change of the symbol probably cannot be recommended right now, since most users are familiar with this metaphor and a change would lead to problems. Nevertheless, I am curious, how long it will take, until the saving-symbol will be changed or whether we will still use the same symbol in 50 years."
- **Feedback:** students assessed instructional methods used in the course or they proposed possibilities for improving the interactive mode; example of feedback regarding the lecture: "The last lecture was again, very interesting; it was nice to see how our eyes can fool us and that perception is not as perfect and reliable. According to my opinion it was nice to see, that the lecture is always presented with pictures and videos, what deserves praise, because it contributes to a better understanding of the subject matter."

Table 1. The merged category system of the qualitative content analysis

Category	Subcategory	Statements
Impression of the Lecture	General	19
	Narration	23
	Reflection & Feedback	24
General positive attitude to HCI	Positive	44
Feedback about the interactive mode	Time-consuming	8
	Effective learning	12
	Grading	6
HCI in everyday life		24
Lab course		17
Language		6
Diary	Personalization	43
	Feedback about the idea	21
	Feedback about the tool	11
	Experience with writing the diaries	11
	Number of words	12

We found out that the students wrote a lot about their **positivities according to HCI** (44 statements). They noted that HCI is a very important aspect in Computer Science and an essential factor for the success of a product. Examples were "Computer Science should not only be engaged with functional aspects of products. It is also necessary to know how humans interact with these products.", "I think it is also important to learn how to analyse Usability, because we want to develop homepages or applications, which are more user-friendly. This topic will be also important for our future jobs", and "The gain in experience from the HCI lecture, exercise and project will be reflected in future projects (homepages as well as software) according to an optimal adjustment to user groups".

Furthermore, students gave us **feedback about the interactive mode** (26 statements). Three different kinds of statements could be found:

- **Effective Learning:** students found that the interactive mode gave them the possibility to learn HCI more effectively (12 statements); example for effective learning of the interactive mode: "It was a very interesting module, because we learned a lot about interaction. It was never boring and stimulated our creativity."
- **Time-Consuming:** students noted that interactive mode is more work than a "traditional" lecture with a final examination (8 statements); example for time-consuming of the interactive mode: "Perhaps it is not a bad idea, but it is very time-consuming, particularly the diaries, because of the regularly required entries."
- **Grading:** students wrote that the grading of the interactive mode was not clear or fair enough (6 statements); example for grading of the interactive mode: "The specifications and the assessment criteria were not clear enough. Additionally, I think it is better to get more points, because it should be an alternative to the lecture. But if you want a better grade you need the exam in addition, because with the points of the interactive mode alone you can only get an average grade."

Additionally the diaries showed that the interactive mode made students more aware of **HCI in everyday life** (24 statements). For example 2 students could see conclusions between an exhibition and HCI: "I visited the exhibition 'Dinner in the Dark'. [...] Since this visit, I think about accessibility of websites, computers and operation systems for blind people and I'm appalled that there is so little done for these people."

In addition, the **lab course** of the module HCI played a significant role in the diaries (17 statements). Students wrote that they could implement what the learned in the interactive mode as well as in the HCI lecture and in the HCI lab course. They liked the fact that the lab course was matched with the lecture (e.g. "For us it is easier to learn the theory from the lecture with the help of the practical exercises. Therefore, we get a good overview about good web designs.").

Another interesting aspect was the **personification** of the online diaries. In at least 8 student's diaries a personification of the diary could be found, although this was never explicitly asked by the instructor or the tutor of the interactive mode. Examples for personalization of the diary were for example: "Dear diary. I am sorry that I neglected you the last days, but I had much to do and I am really sorry ☺" , So, dear diary, that was it for a start, if anything else happens, you will be the first to know it, bye" or "What a wonderful day it is today! Please excuse that I didn't write into you for a long time. No, there isn't anybody else. No, you are the only one for me. Let me tell you what happened in the meantime while we haven't seen each other."

Another aspect mentioned by students was the **language**. The course language was German and 6 students noted that the language was an additional challenge for them (e.g. “I’m worried about the foreign language.”)

In the last part of the diary, students were asked how they found the idea of the online diaries (21 statements), whether they felt well supported by the tool (11 statements), how they experienced writing the diaries (11 statements) and fulfilling the proposed word count (12 statements).

Interestingly four statements described students’ scepticism towards writing the diaries in the beginning, but then the students’ appraisal got better and they valued the diaries (e.g. “I liked the learning diary, although I have to admit that it was a burden in the beginning, because I didn’t know what to do with it and especially, what to write into it.”). Despite of one student who found writing the diary annoying, all other statements by students were positive. Online diaries were experienced especially valuable for memorizing inputs of the lectures (4 nominations, e.g. “Very good [idea], because I could do a repetition of the lecture and because I additionally wrote it down, I could memorize much more, than I would have thought.” Two students valued the feedback aspect of the diaries (e.g. “I found it good, because it is important for the lecturers. Due to our [written] opinions, the lecturers could get a feedback and now can optimize the courses and transfer the course mode to other courses as well.” and further two appraised the possibility to document the project of the interactive mode (e.g. “The learning diary was a good possibility for us to track our project. For me it was an opportunity to plan the project in the right way from the beginning on.”)

4.3 Implications for Human Learning

Many students described the interactive mode of the course on human-computer-interaction and psychology as a possibility to learn HCI effectively. Learning effectively in this case means, for example, to learn continuously by regularly reflecting themes heard in the lecture and working on self-initiated HCI projects. We observed that students were enthusiastic and highly motivated to work on their project as they could elaborate their own project-themes and organize their project-work according to their own ideas. During the project work students could pursue their interests in HCI and develop their creativity. Entries of the diaries showed, that students reflected on various aspects of HCI during the whole summer term, which were discussed in the lecture as well as observations beyond course borders by students. Thus, students learned actively and without or just with a low level of stress during the whole semester term, in contrast to cumulatively learning several days before the final exams. As the diary provides space for students to reflect on lectures, reflected topics could be memorized easier. One student argued: “The idea of the diary was great, because I could repeat what I heard in the lecture and as I wrote the reflection in the diary, it was easy for me to remember things.“ Another positive implication of the diary was that students could use it as a kind of learning protocol and document what they have done and learned during the semester. This learning progress documentation gives students transparency concerning their learning status. For example, one student wrote: “Furthermore the diary gave insight into what things were done or learned and who made what in the team project.”

Students' performance of the interactive mode was high and manifested itself in creative projects. For example, one student project was the elaboration of a cyber cell phone 3D-model focusing on design and user interface aspects. Another example was the elaboration of a video illustrating the functionality of face-tracking in cars, elaboration of possibilities of internet accessibility for blind and visually handicapped persons, etc.). Also, detailed diaries (most of the students wrote among 8 and 15 diary entries) confirmed students' interest in active learning.

4.4 Suggestions for Improvements

The results of the qualitative content analysis offer the possibility to recognize improvements for the future:

- **Language:** For students, whose mother tongue is not German, it would be helpful to make it explicit that projects and entities of diaries can also be written in English.
- **General process of the interactive mode:** Based on the fact that the interactive mode was proposed for the first time, it was not exactly clear how time-consuming the projects are. It would make sense to think about the point-system in connection with expenditure of time (e.g. to give bonus points for extensive projects). All requirements regarding main aspects of the interactive mode need to be presented more transparently in the beginning of the course.
- **Exam:** It seems reasonable to decouple the interactive mode from the traditional exam.

Issues noted positively and hence to be kept were:

- **Project:** Students could realize their own project idea and could choose their group members by themselves.
- **Lab Course:** Students could put the learned input of the HCI lecture in the interactive mode as well as in the HCI lab course into practice. This provided a deeper understanding of the field.
- **Lecture:** Students found the integration of additional practical examples (e.g. showing videos or comparing homepages) in the lectures units as positive.

5 Conclusions and Further Work

Our experience in the interactive mode of the "lecture" in HCI and Psychology confirmed our initial hypothesis regarding the enhancement of learning due to increased reflection and various forms of interaction. Although the new and innovative interactive mode was met with scepticism in some students initially, and a few students dropped out in the very beginning, all students who completed it perceived it as constructive and effective. Retrospectively, the manageable increased effort on the side of instructors contributed to meeting learning goals on the level of professional as well as generic competencies, in particular communication, giving and receiving feedback, team competence, presentation skills and self-directed learning.

The course example described in this paper showed that usable and simple web-service technology has the potential to promote sustainable, deep and social learning.

This is particularly the case if it is applied by educators who give feedback and act as facilitators rather than acting as pure experts and examiners. Thus, diary and interactive team-space web-services proved invaluable in facilitating improved interaction and, in general, raised awareness of the importance of computer-mediated interaction in learning. Web services play the role of enabling technology with simple, usable user interface.

Further work will redesign the course on HCI and Psychology along the "lessons learned" from our initial experience. The interactive mode will be refined and complemented, and offered as a self-contained alternative to the traditional exam. The course will be tracked by empirical studies in order to provide still more insight into the dynamics of learning. In particular, we are interested in students' unfolding in the area of generic competences which appear to be hard to measure objectively.

References

1. Jafari, A., Kaufman, C.: Handbook of Research on ePortfolios. Idea Group Reference, London (2006)
2. Hartnell-Young, E., Morriss, M.: Digital Portfolios, 2nd edn. Corwin Press, California (2007)
3. Motschnig-Pitrik, R.: Person-Centered e-Learning in Action: Can Technology help to manifest Person-centered Values in Academic Environments? Journal of Humanistic Psychology 45(4), 503–530 (2005)
4. Debevc, M., Kosec, P., Holzinger, A.: Examples of Using Technology in Teaching: Human-Computer Interaction According to the Bologna Process. In: Proceedings of the IEEE ITI 2008, 30th International Conference on Information Technology Interfaces (2008)
5. Motschnig-Pitrik, R., Holzinger, A.: Student-Centered Teaching Meets New Media: Concept and Case Study. Journal of Educational Technology & Society 5(4), 160–172 (2002)
6. Holzinger, A., Motschnig-Pitrik, R.: Considering the Human in Multimedia: Learner-Centered Design (LCD) & Person-Centered e-Learning (PCeL). In: Micheuz, P., Antonitsch, P.K., Mittermeir, R. (eds.) Innovative Concepts for Teaching Informatics Carl Ueberreuter, Vienna, pp. 102–112 (2005)
7. Ebner, M., Holzinger, A., Maurer, H.: Future Interfaces for Technology Enhanced Learning. In: Proceedings of the Universal Access to Applicatoins and Services. LNCS, vol. 4556. Springer, Heidelberg (2007)
8. Carmean, C., Christie, A.: ePortfolios: Constructing Meaning Across Time, Space, and Curriculum. In: Kaufman, C., Jafari, A. (eds.) Handbook of Research on ePortfolios. Idea Group Reference, London (2006)
9. Barrett, H.C.: Strategic questions to consider when planning for electronic portfolios. Learning & Leading with Technology 26(2), 6–13 (1998)
10. Sherman, G.: Instructional Roles of Electronic Portfolios. In: Kaufman, C., Jafari, A. (eds.) Handbook of Research on ePortfolios Idea Group Reference, London (2006)
11. Stangl, W.: Lerntagebücher als Werkzeug für selbstorganisiertes Lernen (last access: 14.08.2008), `http://www.stangl-taller.at/ARBEITSBLAETTER/LERNTECHNIK/Lerntagebuch.shtm`
12. Kabicher, S., Derntl, M., Motschnig-Pitrik, R.: Towards an Active Curriculum for Computer Science. In: Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008, Vienna, Austria (2008)

13. Motschnig, R., Derntl, M., Mangler, J.: Web Support for Learning Contracts: Concept and Experiences. In: Proceedings of the 2nd International Conference on Multimedia and Information & Communication Technologies in Education, Badajoz, Spain (2003)
14. Motschnig, R., Derntl, M., Figl, K., Kabicher, S.: Towards Learner-Centered Learning Goals based on the Person-Centered Approach. In: 38th ASEE/IEEE Frontiers in Education Conference, Springs, NY, Saratoga (2008)
15. Rogers, C.R.: Freedom to Learn for the 80's. Charles E. Merrill Publishing Company, Columbus (1983)
16. Mangler, J.: CEWebS - Cooperative Environment Web Services. Faculty of Computer Science, Univ. of Vienna, Master thesis (2005)
17. Derntl, M., Mangler, J.: Web Services for Blended Learning Patterns. In: Proceedings of the IEEE International Conference on Advanced Learning Technologies, Joensuu, Finland (2004)
18. Mangler, J., Derntl, M.: CEWebS - Cooperative Environment Web Services. In: Proceedings of the 4th International Conference on Knowledge Management (I-KNOW 2004), Graz, Austria (2004)
19. Mangler, J., Motschnig-Pitrik, R., Derntl, M.: Developing Open Source Web Services for Technology-Enhanced Learning. In: Proceedings of the Sharable Content Objects (SCO 2005), Brno, Czech Republic (2005)
20. Peterson, B., Roemmer-Nossek, B., Logar, S.: Using creative writing techniques for ePortfolio work: Writing to enhance your competencies and evelop your own voice. In: 2nd Thematic Conference Lifelong Learning - ePortfolio and Open Content, Katowice (2007)
21. Roemmer-Nossek, B., Peterson, B., Logar, S., Zwiauer, C.: ePortfolio Pilot Project at the University of Vienna - Towards a Framework. In: ICL 2007, Villach, Austria (2007)
22. Derntl, M., Motschnig-Pitrik, R.: Towards a Pattern Language for Person-Centered e-Learning. In: Proceedings of the Society for Information Technology & Teacher Education International Conference (SITE 2003), Albuquerque, New Mexico (2003)
23. Derntl, M.: The Person-Centered e-Learning Pattern Repository: Design for Reuse and Extensibility. In: Proceedings of the ED-MEDIA 2004: World Conference on Educational Multimedia, Hypermedia & Telecommunications, Lugano, Switzerland (2004)
24. Mayring, P.: Qualitative Inhaltsanalyse (Qualitative Content Analysis), 8th edn., Beltz, Weinheim (2003)

Knowledge-Based Patterns of Remembering: Eye Movement Scanpaths Reflect Domain Experience

Geoffrey Underwood*, Katherine Humphrey, and Tom Foulsham

School of Psychology, University of Nottingham, UK
geoff.underwood@nottingham.ac.uk, lpxkah@nottingham.ac.uk, lpxtf@nottingham.ac.uk

Abstract. How does knowledge of a domain influence the way in which we inspect artefacts from within that domain? Eye fixation scanpaths were recorded as trained individuals looked at images from within their own domain or from another domain. Sequences of fixations indicated differences in the inspection patterns of the two groups, with knowledge reflected in lower reliance of low-level visual features. Scanpaths were observed during first and second viewings of pictures and found to be reliably similar, and this relationship held in a second experiment when the second viewing was performed one week later. Eye fixation scanpaths indicate the viewer's knowledge of the domain of study.

Keywords: Eye movements; scanpaths; knowledge-based processing; saliency.

1 Introduction

When we are introduced to a domain of knowledge and educated about its aspects, our perceptions of the domain are changed, and our enquiries of new artefacts may reflect these altered perceptions. One tool that can be used to investigate the perceptions of knowledgeable individuals who are inspecting pictures of domain-specific artefacts is the recording of the viewer's eye movements. It is known that the eye fixations of trained drivers, for example, are different from those of novices, even when watching roadway scenes rather than engaging in the active task of driving itself [1, 2]. In normal vision, scanning behaviour consists of saccades: fast, ballistic shifts of gaze that bring regions of interest onto the area of the eye with the highest resolution (the fovea). The measurement of these movements is now common in many diverse areas of cognitive psychology including research into reading [3] the perception of pictures and scenes [4], problem solving [5] and complex behaviours such as driving [1]. A variety of measures are taken in such experiments, and these measures reflect assumptions about the functioning of the visual-cognitive system: firstly, that the accuracy with which saccades are targeted is an indication of early attentive processing based on peripheral vision (for example, in the analysis of saccade landing positions in reading, or the measurement of saccade lengths as an index

* We are grateful to the U.K. Engineering and Physical Sciences Research Council for support (EPSRC award EP/E006329/1), to Laurent Itti for the use of his saliency software, and to two anonymous reviewers for their comments.

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 125–144, 2008.

of performance in visual search); secondly that the position and duration of a fixation are reliable diagnostics of what is being processed and the difficulty of this processing. Thus fixation durations and gaze durations (the sum duration of consecutive fixations within a region) are commonly explored to indicate cognitive processing of different stimuli. What these measurements have in common is that they generally consider each eye movement event independently. However, the pattern of inspection can only be revealed by considering a sequence of successive fixations. These patterns are sometimes referred to as *scanpaths* or *scan patterns* [4].

What determines where the eyes will move to next? Early researchers considering this question recorded a large variation in the scanpaths made when observers viewed complex stimuli such as pictures, but there were patterns [6, 7]. A glimpse at some scanpaths (see Figure 2) shows that the places where people fixate, and the movements between them, are not random, and neither are they regularly distributed across space, as we might suppose if the visual system were trying to sample the whole scene uniformly. If looking at a picture containing people for example, fixations tend to be focused on the faces. In fact, fixations across many stimuli tend to cluster on regions rated informative [8] and some researchers have analysed the low-level statistical properties of these image regions [9], in an attempt to identify the bottom-up determinants of attention.

In addition to being tied to the visual features present in a stimulus, the pattern of eye movements made by an observer is known to vary according to the task being undertaken. In his often-cited early work on eye movements, Yarbus [7] highlighted the fact that scanpaths exhibited when viewing the same stimulus would be quite different if the viewer was given a different task. Two commonly studied experimental tasks are looking at a scene in order to remember it for later and searching a scene for a specific target. The between-subjects variation in scanpaths has led some to label scanpaths as distinctively idiosyncratic, presumably reflecting personal knowledge, experience or viewing strategy. To study these top-down aspects of overt attention it is useful to be able to compare scanpaths across viewings, stimuli and individuals. Before looking at this technique in more detail, we will consider a specific theory for which scanpath comparison is particularly important. It is necessary here to distinguish between scanpath theory and the measurement of observed scanpaths.

Scanpath theory is an ambitious set of ideas that were originally proposed in two papers by Noton and Stark [10, 11]. The theory describes scanpaths as controlled by internal, cognitive models representing the viewer's expectations of the scene. These models might represent the saccades involved in viewing a picture or scene as a kind of structure or syntax that binds together the features processed at fixation. When viewing the same scene again, as in the test phase of a recognition experiment, a scanpath might be re-invoked or checked against the external stimulus. The main evidence for scanpath theory came from experiments showing that scanpaths recurred when stimuli were reviewed in a recognition task. In Noton and Stark [10] this conclusion was reached based on subjective observation of the patterns shown by each subject and there was no quantification of the similarity between the scanpaths. Other researchers have examined the presence of repetitive scanpaths when imagining a previously seen image [12, 13].

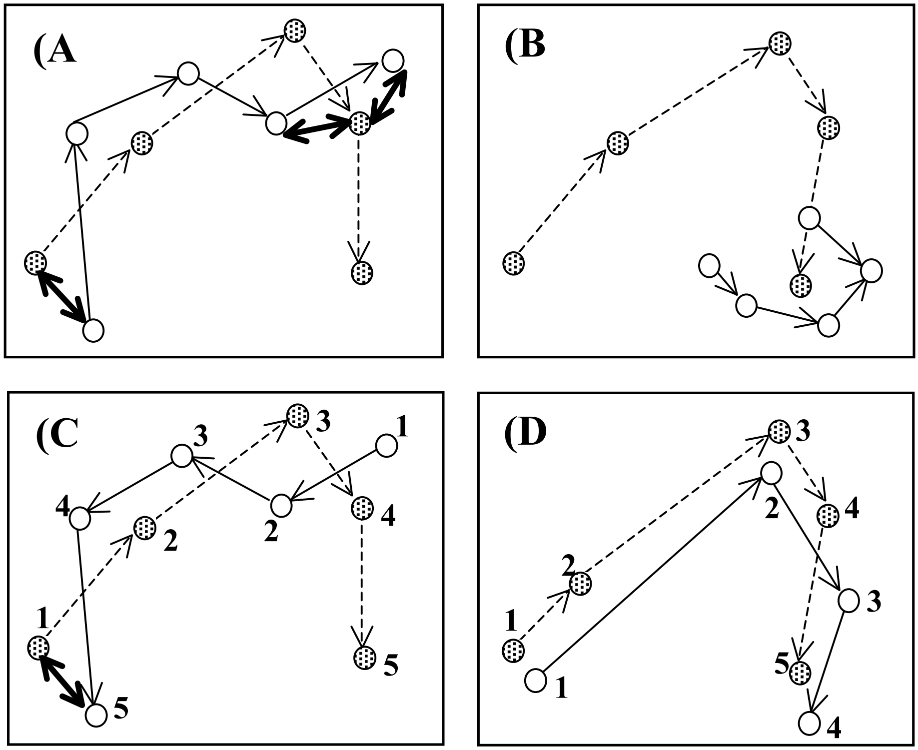


Fig. 1. Calculating the linear distance between two hypothetical scanpaths. Circles show fixation locations. (A) Each fixation is compared to its closest neighbour in the other scanpath. This distance is illustrated for three fixations (bold arrows). (B) This metric is confounded by differences in the spatial variability of the two scanpaths. All of the fixations in one scanpath (open circles) will be compared to just one in the other set, leading to a low mean distance despite very different patterns. (C) The metric ignores sequence information. Ordinal position is emphasised with numbers. Note that the first fixation is compared to the fifth fixation in the other set. (D) If each fixation is compared with that in the same serial position, small differences skew later comparisons. In the case illustrated, despite broadly similar scanpaths, the distance between second and subsequent fixations is large.

Little support has been found for scanpath theory and it has difficulty explaining some phenomena. For example, it is not necessary to move ones eyes to encode or recognise a picture, and the apparently large amount of variability within the patterns shown by a single person viewing the same stimulus also make a strong version of scanpath theory untenable [9]. As a result, some researchers prefer to use the term *scan patterns* rather than *scanpaths* [4], in order to dissociate the fixation recordings from the theory. We do not rely upon any of the assumptions of scanpath theory here. Instead, we will restrict ourselves to a discussion of how the relationships between scanpaths might be quantified, for purposes of comparing the fixation patterns recorded during the viewing of the same picture on two separate occasions.

1.1 Methods for Comparing Scanpaths

In this section, several different basic methods for comparing scanpaths will be reviewed. Figure 1 shows some hypothetical scanpaths and these will be used as examples here. Alongside the practicality of using a method on large datasets, the main criteria for evaluating these methods will be whether it appropriately captures the degree of similarity between different scanpaths and the ease of testing this statistically.

1.1.1 Distance-Based Methods

Scanpaths are inherently spatial. As a result it would seem most appropriate to measure the distance between two scanpaths superimposed on the same visual area. A metric developed by Mannan and colleagues [9, 14, 15] computes the similarity between scanpaths by measuring the distance between each fixation in one set and its nearest neighbour in the other. Scanpaths that are more similar, in the sense that they dwell on locations close to each other, will show a smaller average linear distance. Figure 1 depicts this measurement for several different comparisons. The average linear-distance is defined as D, where

$$D^2 = \frac{n_1 \sum_{j=1}^{n_2} d_{2j}^2 + n_2 \sum_{i=1}^{n_1} d_{1i}^2}{2 n_1 n_2 (a^2 + b^2)} \tag{1}$$

and where n_1 and n_2 are the number of fixations in each scanpath and a and b are the dimensions of the image. d_{1i} is the distance between the *ith* fixation in the first set and its nearest fixation in the second set, and d_{2j} is the same distance for the *jth* fixation in the second set.

Computation of this measure is straightforward from the fixation coordinates that normally make up raw fixation data. In addition, it is robust to scanpaths with different numbers of fixations and is scaled relative to the size of stimulus being viewed (due to the term $a_2 + b_2$). In order to produce an estimate of the absolute degree of similarity, Mannan et al. [9] compute the similarity index, I_s, by comparing the average linear distance between two scanpaths with that between randomly generated scanpaths of the same size (D_r):

$$I_s = \left(1 - \frac{D}{D_r}\right)100 \tag{2}$$

This gives a value between 0 (chance similarity) and 100 (identity), with negative values indicating scanpaths that are more different than expected by chance. The distance between randomly generated scanpaths (D_r) produces the normally distributed similarity that would be expected from chance or uniform scanning. This distribution was examined by Mannan et al. [9]. For a constant display size, the average random distance gets smaller as more fixations are added to the scanpath (as n_1 and n_2, which do not have to be equal, increase).

The second problem occurs when the spatial distribution in one set of locations is very different from that in the other (see Figure 1b). This leads to multiple fixations being compared to a single location in the other scanpath, potentially producing high similarity from two scanpaths that appear very different. Similarly, one outlier will skew two otherwise very similar scanpaths. Tatler, Baddeley, & Gilchrist [16] identified these problems, pointing out that the linear distance method is fundamentally confounded by differing amounts of variability in the two scanpaths. Henderson, Brockmole, Castelhano, & Mack [17] proposed a "unique assignment" variant of the linear distance metric whereby each fixation is paired with just one other. All possible pairings are computed, and that chosen which minimises the average distance. A disadvantage of this approach, and of the serial position version, is that they require equal numbers of fixations in each set.

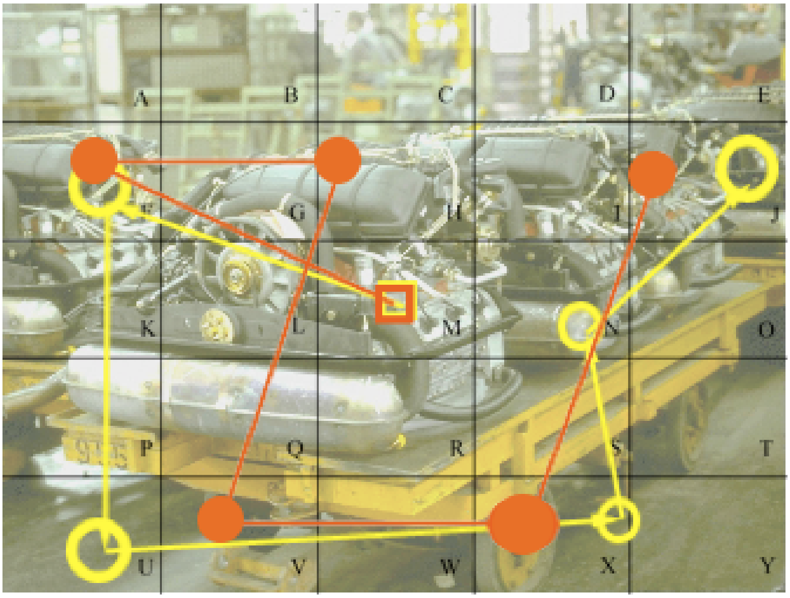
Figure 1 illustrates two specific problems with the linear distance method. Firstly, the measurement does not take into account the temporal sequence of the scanpath. Fixation locations are compared to whichever fixation is closest, regardless of when it occurred. As it ignores the order information, this metric would give extremely high similarity to the example in Figure 1b, despite the fact that in one scanpath the observer starts at the bottom left and works upwards whilst in the other they do the opposite. One way to avoid this problem might be to compute a "serial position" version, where the distance is computed between each fixation and that fixation which occurred in the same serial position in the other scanpath. However, this would be skewed by any small deviations, as illustrated in Figure 1d.

1.1.2 String Edit-Distance

In order to capture the temporal order of scanpaths, several researchers have utilised a method designed specifically for sequence analysis: the Levenshtein distance [18], or simply string edit-distance [12, 19]. This algorithm is an extension of the Hamming distance that gives the difference between two strings of symbols in terms of how many positions are identical. The edit-distance is defined as the number of editing operations (deletions, insertions and replacements) required to turn one string into another, and this distance will decrease as strings become more similar. A method (based on discrete dynamic programming) is available which computes the minimum number of operations required, and this distance has been used for comparing a range of different strings of items, from DNA sequences to birdsong [20]. Figure 2 illustrates how this method can be applied to eye movement sequences. The visual stimulus is divided into regions, each of which is allocated a letter. Each 2-dimensional scanpath can now be transformed into a character string, and the edit distance between the two can be computed. It is often desirable to compare similarity across scanpaths of different lengths, so the distance can be normalised by the number of fixations, and an index of similarity, which here we call s calculated from its reciprocal distance between two scanpaths is calculated as the minimum number of steps required to transform one string into another:

$$s = 1 - \frac{d}{n} \tag{3}$$

where d is the edit-distance between two scanpaths and n is the number of fixations within the longest scanpath. This metric is equivalent to Ss, the first of three string-based similarity measures identified by Privitera [21]. Ss is the sequential similarity, whilst Sp (locus similarity) represents the number of characters shared by both



String 1: *FUXNJ* (open circles)   String 2: *FHVXJ* (filled circles)

Editing cost in transforming String 1 into String 2:

String item 2:	*U* to *H*	= one replacement
String item 3:	*V* inserted	= one insertion
String item 4:	*N* deleted	= one deletion

Total editing "cost" = 3 changes in a string of 5 items
Normalised difference = 3/5 = 0.6
String similarity = 1 – 0.6 = 0.4

Fig. 2. Each circle indicates an eye fixation, and here two sequences of fixations are represented. Fixation durations are suggested by variations in the sizes of the circles. A string-editing procedure is used to evaluate scanpath similarity by calculating the "editing cost". The distance between two scanpaths is calculated as the minimum number of steps required to transform one string into another. This edit cost can be normalised and converted into a standardised similarity score, where a score of 1 represents two identical strings.

strings, giving an index of the positions both scanpaths dwell on, regardless of order. The final metric mentioned is *St* (transition similarity), which consists of Markov matrices of region transitions.

There are several important decisions to be made if using the edit distance method. Firstly, how is the region schema produced? In some cases there are clear areas of interest that can be predefined by the researcher. These might correspond to areas of a display or particular objects in a scene. However, in other cases it might be desirable to look at scanpaths over the whole image and to use regions of a constant size. In this situation the image can be divided into a grid, although this raises the question of how large these regions should be. A third possibility is to use the fixation data themselves to produce the regions, using statistical clustering techniques, for example Privitera & Stark [22]. Thus the region schema could be found which divided the image into the desired number of regions. It might be useful to perform the analysis with several different sets of different regions, perhaps of varying sizes. Similarity that is present at several scales and robust to changes in the region organisation is likely to be the most reliable. Choi et al. [19] argue that the estimates of similarity that they give are robust, whether using 10 or 15 regions. Figure 3 illustrates the consequences of varying the grid size, and can be used to suggest an optimal size in which increasing the number of grids has no further discernible effect on the Levenshtein distance value.

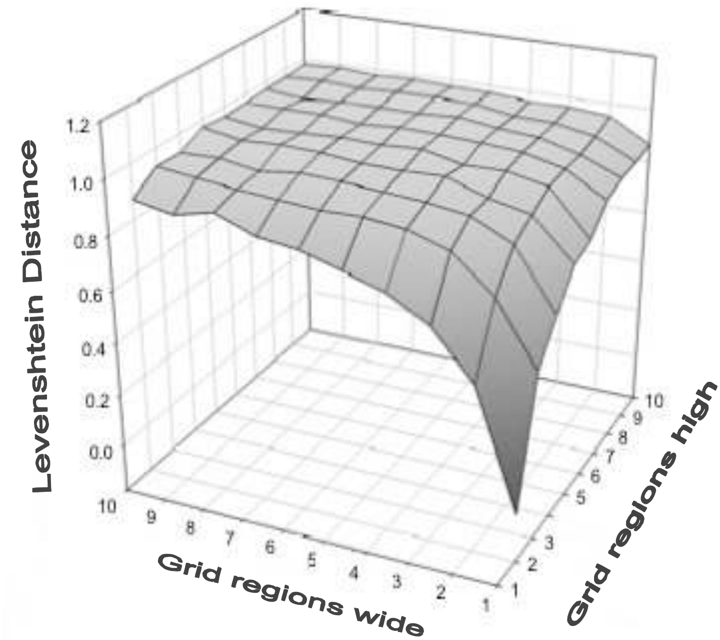


Fig. 3. The normalised Levenshtein string-editing distance between the fixations in two randomly generated scanpaths varies with the size and number of regions. Data are based on scanpaths over an area divided into a grid of regions with dimensions varying from 1 x 1 (only one region where all fixations are evaluated as equal) to 10 x 10 (100 regions). Note that the distance expected by chance increases as a finer grid is used.

A second decision that is commonly made is to condense consecutive fixations on the same region, which would result in repeated symbols in a string, into a single character. Groner, Walder, & Groner [23] make a distinction between local and global scanning, with the former consisting of small readjustment saccades, of less theoretical interest. Thus the coarser scale movements between regions may be more useful. Of course, in combination with decisions regarding the size and shape of regions this will have a profound effect on estimates of scanpath similarity.

How can we calculate the significance of *s*? As with distance-based methods this problem amounts to comparing experimentally derived similarity with a random or chance estimate. The chance similarity can be easily calculated as the probability that any two characters will be the same. Thus for a 5 by 5 grid there is a 1/25 chance that any two fixations will be in the same place. The similarity of randomly generated scanpaths could be used as the denominator in an estimate of absolute similarity analogous to equation (2). Randomly generated similarity varies with the size and number of regions (as demonstrated in Figure 3), so the same grid would need to be used as that used with experimentally derived data. The random model might also need to be adjusted to take into account other biases present in the experimental sample. For example scanpaths might be constrained to start or finish in a particular place, and this would affect similarity estimates.

Although the editing-distance method successfully captures the temporal sequence of scanpath data, it reduces all spatial information to a binary choice where fixations are either in the same region or they are not. This seems intuitively unsatisfactory and leads to some comparisons being equivalent despite large differences (see Figure 1b). This problem makes the regions used, often a fairly arbitrary decision, critical, as a fixation which lies just over the region border will be counted the same as one which is on the other side of the image. In an extreme case, a fixation might be computed as more similar to a fixation that lies in the same region than one that is spatially closer but outside the region's bounds. In one sense, using more regions provides a more accurate representation with a higher resolution of the movements made. However, more regions also make the analysis less tolerant of small deviations that might otherwise seem negligible.

1.2 Other Methods

We have outlined two main methods for comparing scanpaths, but there are several other ways of analysing such patterns. Some researchers have used Markov matrices [1, 24, 25], which show the transition probabilities from one predefined region to another. However, while this may be useful for short scanpath segments, the matrices explode exponentially when longer chains are explored, making them impractical. They also require the same decisions regarding which regions to use as the edit-distance.

Fixation maps such as those shown in Figure 4 are a useful way of displaying eye movements, particularly those from large populations [26]. In these maps, fixations are represented by a two-dimensional Gaussian centred on the fixation location. The width and height of the Gaussian can be varied, and multiple fixations summed, forming an attentional landscape. Comparing two fixation maps is then possible, and as fixations are essentially distributed this may be an efficient way of computing the spatial similarity between two scan patterns which avoids some of the confounds associated with linear distance. Two maps could be correlated or a difference map could be produced (perhaps after normalising the height of the peaks). Spatially identical scanpaths would give a completely flat difference map. Standard fixation maps hold no information regarding sequence, although it might be possible to introduce a temporal element, either by combining maps derived from different time periods or by varying the height of fixation peaks over time. The fixation map approach also provides a way of identifying the regions of interest from the data (for use in the edit-distance method, for example). A threshold or critical value can be chosen, and all

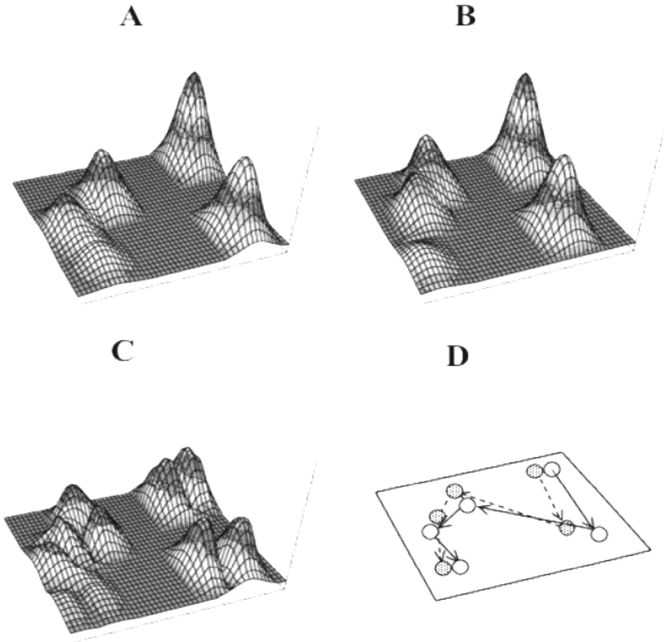


Fig. 4. Fixation maps can be used to represent and compare scanpaths. Each fixation is represented by a 3D Gaussian. The z-axis or height could represent fixation duration or another index. Alternatively, in order to encode sequence, height can represent when the fixation occurred. Early fixations produce a high peak, whilst later ones give a progressively lower distribution. Multiple fixations on the same area are summed together, producing an attentional landscape. A) and B) show two fairly similar scanpaths, with z normalised to range from 0 to 1. The absolute difference between the two can be represented in the same way as a "difference" map. Identical scanpaths would show a totally flat difference map, whilst peaks indicate areas where fixation allocation differed between the two. A 2D schematic of the two scanpaths is also shown (D).

areas receiving more attention, those whose peaks lie above the threshold, can be selected for use in further analyses. Alternatively the threshold could be gradually adjusted until the required number of discrete regions is selected. Tatler et al. [16] also avoid contaminating their measure of similarity by looking at the full distribution of fixations across the image. In their method fixations are binned into 2° by 2° squares and a spatial probability distribution derived. The difference between two scanpaths is then given by a measure from information theory, the Kullback-Leiber divergence, which computes the difference between the corresponding probability distributions. The Kullback-Leiber divergence gives the number of additional bits of information needed to describe one distribution given another. Thus a low value indicates similar scan patterns. A disadvantage of this technique is that it requires large amounts of data, so it is best used when the fixations from many trials and observers are being examined.

In summary, distance-based methods are useful for averaging commonalities in *where* people look, though they are confounded by differences in spatial variability and do not reflect the temporal order of serial scanpaths. The edit-distance approach captures sequence at the expense of spatial resolution but requires somewhat arbitrary decisions such as which region scheme to use. Levenshtein's [18] string editing method is implemented easily and has been used successfully elsewhere, and so will be used to compare fixation sequences in the following two experiments.

2 Experiment 1: Domain Knowledge and Fixation Scanpaths

Fixation scanpaths are more similar over multiple viewings of a picture than would be expected by chance. Independently it has also been found that low-level visual saliency has a large influence on the locations of the first few fixations. However, bottom-up processes such as these may be overridden by top-down cognitive knowledge in the form of domain proficiency, suggesting that fixation locations are determined by multiple factors. In the present experiment domain specialists were asked to look at a set of photographs in preparation for a memory test. They were then given a second set of pictures and were asked to identify each one as being from the previous set, or new (not seen during the encoding phase). The experiment investigates the stability of fixation scanpaths between the first and second viewing of a picture, and the influence of the viewer's own knowledge of the domain from which the pictures were selected.

Regular patterns of fixations may result from fixation on the most conspicuous regions. Each time the picture is inspected, perhaps the viewer looks first at the most conspicuous region, then at the next most conspicuous region, and so on. The conspicuity or saliency distributions do not change between viewings of course, and so the sequence of fixations would not change either. Itti and Koch's [27] algorithm enables the measurement of the visual saliency of an image on the basis of its physical properties, by the identification of peaks in the distribution of intensity and changes in colour and orientation. The algorithm builds an overall "saliency map" of the image to determine the ordinal allocation of attention to the regions of the display. The effect of salience on eye fixation locations has been supported by Parkhurst, Law, and Niebur [28] who showed participants a range of images, including photographs of home interiors, buildings, and natural environments. Saliency strongly predicted fixation probability during the first two or three fixations, and the model performed above chance throughout each trial. Parkhurst et al. concluded that a purely bottom-up account of visual attention was sufficient to account for fixation behaviour. Further support comes from Underwood, Foulsham, van Loon, Humphreys, and Bloyce, [29] who found that when viewers inspected the scene in preparation for a memory task, objects higher in saliency were more effective in attracting early fixations.

A bottom-up explanation for similarities in scanpaths at encoding and recognition could therefore be that fixation locations are at least partly determined by saliency, as this remains constant over viewings. Repeated patterns of fixations may be a product of viewers repeatedly looking at the most conspicuous regions of the picture, and so similar scanpaths may not result from a memory of the first viewing but from the visual characteristics of the picture itself.

Top-down influences are known to reduce the effects of visual saliency on fixation behaviour, and so it is possible that the bottom-up effect of saliency could be moderated by an increase in the viewer's top-down knowledge of the scene. If an effect of domain knowledge on saliency influences eye movements, it would be interesting to see if it was consistent over repeated viewings. This has not been specifically investigated in non-search tasks, although there have been studies that have found a cognitive override of saliency in search tasks [29-31].

The eye movements of experts differ from non-experts; for example, experienced football players have been found to have a higher search rate, involving more fixations of shorter duration than novice players [32]. However, there is little evidence of how the eye movements of domain-specialists and non-specialists vary with the saliency map of an image. Furthermore, if eye movements are related to memory, then the overriding effect of domain knowledge should be constant over time, producing similar scanpaths on multiple viewings of the same stimulus. Specialists are consistently more accurate with the recognition of domain specific targets [33] and they consistently produce scanpaths reliably different from non-specialists [34]. We hypothesise that the fixation scanpaths of non-specialists will be influenced by low-level visual saliency, but that domain specialists will produce different eye-movements to non-specialists on the same picture, and would provide support for the overriding effects of domain knowledge. If domain specialists look at images in ways that reflect their knowledge, then the possibility arises of recording a student's eye movements as an implicit assessment of their knowledge.

2.1 Method

Three groups of students were recruited: 15 Engineers, 15 American Studies students and 15 non-specialists (control group).

Eye position was recorded using an SMI iVIEW X Hi-Speed eye tracker. Ninety high-resolution digital photographs were used as stimuli, sourced from a commercially available CD-ROM collection. Thirty of the pictures were engineering-specific, 30 were Civil War specific, and 30 were of natural scenes such as gardens, parks and landscapes. Half of the pictures in each category were shown in both the first viewing (encoding) and in the second viewing (recognition) phases, while the other half were shown only as part of the recognition test test.

Itti and Koch's [27] model was used to generate saliency maps for the first five most salient regions for each picture (see Figure 5) – the regions of greatest aggregate intensity, colour and orientation variation. The only further criterion for stimuli was that all 5 salient regions were non-contiguous; those pictures where the same or overlapping regions were re-selected within the first 5 shifts were replaced.

Following a practice at the task, the first stage of the experiment began, with 45 pictures shown to all participants (15 engineering pictures, 15 Civil War pictures and 15 natural scenes), presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen, and each picture was presented for 3 sec. During this time participants freely inspected any aspects of the picture they chose. After all 45 stimuli had been presented, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was new or old, using the computer keyboard

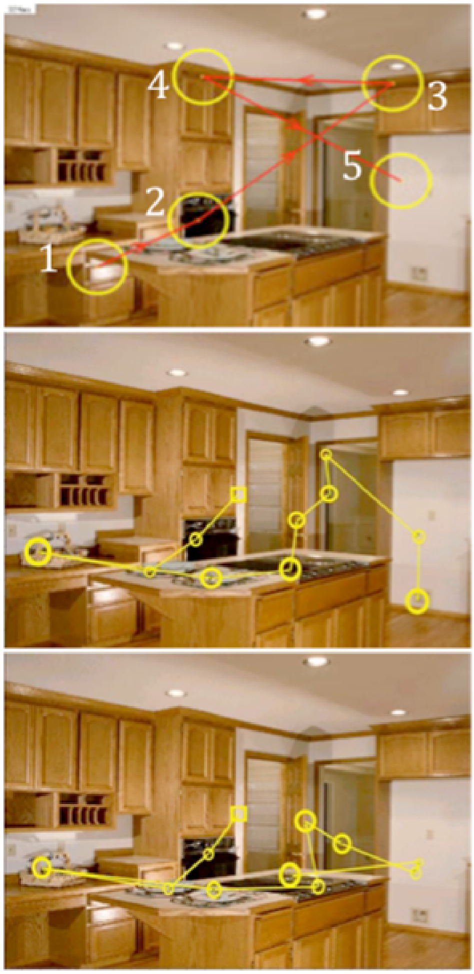


Fig. 5. Saliency map predictions of the first five fixations on a picture from the neutral category (top panel), based on the five most salient regions. The saliency map algorithm identifies the most conspicuous regions of an image on the basis of changes in colour, orientation and brightness. The most conspicuous region is first identified (the brightly illuminated draw in the lower left quadrant of the top picture, numbered "1" above), and then the next most conspicuous region, and so on. A process of inhibition of return prevents the same region being selected repeatedly. The rank orderings of conspicuous regions form the basis for a prediction of the order of the first few fixations on a picture. The top panel, showing the five most salient regions of the picture, and their rank orderings serve as the model-predicted order of fixations, for comparison with actual scanpaths. The centre panel and the bottom panel are exceptionally similar scanpaths from the first and second viewings by one participant.

to indicate whether each picture had been shown previously. During this phase, 90 stimuli were presented in a random order; 45 of these were old and 45 new. Each picture was again shown for 3 sec.

2.2 Results and Discussion

The analyses focused on string analyses, to compare scanpaths during encoding and the second viewing, and comparing encoding scanpaths to the sequences predicted by the saliency model. String editing was used to analyse the similarity between scanpaths produced on encoding and second viewing. A 5 by 5 grid was overlaid onto the stimuli (see Figure 2). The resulting 25 regions were labeled with the characters A to Y from left to right. Fixations were then labeled automatically by the program, according to their spatial coordinates, resulting in a character string representing all the fixations made in this trial.

The scanpaths generated from encoding of a picture were compared with the scanpaths during recognition, and with those predicted by the saliency model [27]. An example of a predicted scanpath is shown in Figure 5 (top panel). Each scanpath was given a score depending on the similarity of the eye fixations on first viewing each picture with the scanpaths predicted by the saliency model. The two lower panels of Figure 5 show examples of observed scanpaths, during encoding (centre panel) and recognition (bottom panel). The average similarity scores (maximum value of 1) for each group of participants were compared for each type of stimulus using an ANOVA.

There was a reliable between-groups difference in viewing Civil War stimuli [$F(2,42) = 4.068$, $MSe = 0.012$, $p<0.05$], with engineers (similarity score of 0.098) and control students (score of 0.107) matching the saliency model closer than American Studies undergraduates (score of 0.054). Paired contrasts indicated that both of these comparisons were reliable at $p<0.05$. There was also a between-groups difference in viewing Engineering stimuli, [$F(2,42) = 16.249$, $MSe = 0.026$, $p<0.05$], with American Studies (score of 0.122) and control participants (0.105) showing scanpath similarity scores closer to the predictions of the saliency model than those of the engineering students (score of 0.0426). The engineers had lower similarity scores than both of the other groups (both comparisons were reliable at $p<0.001$). There was no between-groups difference in viewing neutral stimuli [$F(2,42) = 2.739$, $MSe = 0.015$, $p<0.05$]. When students looked at pictures of artefacts from within their domain of knowledge they were less likely to look at the visually most salient features, and this held for American Studies participants inspecting Civil War photographs and for engineering students inspecting pictures of motors, turbines and industrial production facilities.

The scanpaths generated from encoding of a picture were compared with those on second viewing during the recognition test using an ANOVA analysis. Overall, there was a string similarity of 0.238 for non-specialists, 0.245 for Engineers and 0.268 for American Studies undergraduates. Randomly generated strings would give a value of approximately 0.0417, making the string similarities for all three groups of participants reliably greater than chance ($p<0.05$). There was no difference between the participant groups on the similarity of scanpaths at encoding and recognition [$F(2,42) = 0.522$, $MSe = 0.004$, $P = 0.597$].

Does knowledge of a domain interact with the influence of saliency on scanpaths when viewers look at images from within their domain? All the fixations made on a particular stimulus were compared to the five most salient areas of that picture. In previous research, saliency effects have been found when memory tasks were performed. In this experiment, it was found that the specialist groups made fewer fixations on the salient areas when the pictures were from their own domain – engineering students made fewer fixations on the visually salient areas of Engineering pictures, and American Studies undergraduates made fewer fixations in salient areas of Civil War pictures. There was no significant difference between groups when looking at neutral pictures, and non-specialists showed no significant difference across stimuli types. This lends some support for the saliency map theory, in that saliency is again shown to have a large influence on eye-movements. However, it does partly argue against the position that a *purely* bottom-up account of visual attention is sufficient to account for fixation behaviour. This is not the case, as salient features had less of an effect when the picture fell into the student's specialist domain. Engineers make reliably fewer fixations in highly salient regions when viewing engineering pictures and American Studies undergraduates made fewer fixations on the salient regions of Civil War pictures. This apparent cognitive-override of saliency may seem intensified because the interesting parts of the stimuli perhaps were of particularly low saliency, and thus it is almost like they were actively seeking out low-salient regions, which would not have been of interest to non-specialist.

The cognitive-override effect that has been found is consistent with previous investigations of saliency influences [29-32] in a search task, but when an encoding task was used, as here, the saliency map did predict fixation locations. This was only found for non-domain specialists here.

Overall, scanpaths produced on encoding of a picture compared to those produced on second viewing were more similar than would be expected by chance. This was consistent across all participants, despite group or stimulus type. Scanpath theory [10] suggests that visual patterns are represented in memory as a network of features and the attention shifts between them. This network is then replayed and compared to the external stimulus when recognising the image later. By this account, the scanpaths at encoding were similar to those at recognition because they were stored and recalled top down, to determine the scanning sequence. Although the similarity seen here is statistically reliable, the scanpaths are far from identical, and there is still a large amount of variance unaccounted for. Previous demonstrations of scanpath similarity have largely used simple patterns, with fewer and larger regions of interest. It is likely that the much more complex photographs used here resulted in reduced scanpath repetition, possibly due to a greater influence of knowledge-based inspection strategies.

In conclusion, saliency does have a strong influence on eye movements, shown by the similarity of actual scanpaths to those predicted by the saliency model [27]. However, domain-specific knowledge can act as an overriding factor, decreasing the influence of saliency on driving eye movements. This effect has been shown to be stable over time.

3 Experiment 2: Delaying the Interval between Viewings

In Experiment 1 the scan patterns recorded during a recognition test were more similar to the patterns seen during the first inspection of the picture than would be expected by chance. The saliency model also had success in predicting fixation locations, but only for viewers who were relatively unfamiliar with the domain from which the pictures were selected. Comparing performance during a recognition test against performance during encoding may have obscured the analysis, however, because encoding and recognition are different tasks with different cognitive processes. Viewers would have been inspecting the pictures for very different purposes during encoding and recognition. A different estimate of scanpath similarity might be obtained if a comparison was made of scan patterns on two successive viewings of a picture where the purpose of inspection is held constant. In Experiment 2 we again show students pictures from their own domain of knowledge and from another, with eye fixations recorded during encoding and recognition, but also tested recognition a week after the initial viewing, so that scanpaths could be compared during two recognition tests. The control participants revealed no interesting patterns of fixations in Experiment 1, and so only domain specialists were compared here.

3.1 Method

The participants were 15 American Studies and 15 Engineering students from the same source as those tested in Experiment 1, the same pictures were used, and eye position was again recorded using an SMI iVIEW X Hi-Speed eye tracker. Saliency maps were generated using Itti and Koch's [27] model with standard parameters. The experiment used a two-by-three mixed design, with two specialist groups of participants and three specific types of stimuli. All participants viewed the same stimuli under the same task conditions. Test pictures were inspected under one of three viewing conditions: encoding, immediate recognition, and delayed recognition.

Participants were not told to look for anything in particular in any of the pictures but were asked to look at them in preparation for a memory test. Following a practice phase with five pictures not otherwise used in the experiment, the first stage of the experiment began. There were 45 stimuli (15 engineering pictures, 15 Civil War pictures and 15 natural scenes) presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen. Each picture was presented for 3 seconds, during which time participants moved their eyes freely around the screen. This presentation format is the same as was used in Experiment 1.

After all 45 pictures had been presented, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was new (never seen before) or old (from the previous set of pictures) by making a keyboard response. During this phase, 90 pictures were presented in a random order; 45 of these were old and 45 new (though the participants were not informed of this fact). In order to facilitate an ideal comparison between encoding and test phases, each picture was again shown for 3 seconds. One week after the original recognition test, participants returned to the laboratory and were shown the 90 test pictures again, with task again being to say whether they had seen each picture during the original encoding phase.

3.2 Results and Discussion

The scan patterns generated from first viewing of a picture were first compared to respective scan patterns predicted by the saliency model [27] again using the string edit-distance method. Observed scan patterns were found to be more similar to those predicted by the model when stimuli were *not* domain-specific. When stimuli *were* domain specific, the similarity score dropped to the estimated chance level. There was a reliable between-groups difference in viewing Civil War stimuli [$F1,28=52.50$, $MSe=0.099$, $p<0.001$], with a string similarity score of 0.027 for American Studies and 0.142 for engineering participants. There was a reliable between-groups difference in viewing Engineering stimuli [$F1,28=48.75$, $MSe=0.067$, $p<0.001$], with a string similarity score of 0.033 for American Studies and 0.128 for engineering participants. There was no difference between groups when viewing neutral stimuli [$F1,28=1.40$]. Students were less likely to perform according to the predictions of the model when viewing pictures from their domain of interest.

The scan patterns observed during the encoding phase and during the immediate recognition test were also quantified by the edit-distance method, and the resultant similarity scores compared to the estimated chance level, using a one-sample t-test. All comparisons showed the string similarities between encoding and test to be reliably greater than chance. Strings of fixation locations were similar for American Studies undergraduates inspecting Civil War pictures (observed similarity score between encoding and recognition of 0.155, $t14=9.54$, $p<0.001$), engineering pictures (similarity score of 0.147, $t14=10.89$, $p<0.001$), and neutral scenes (similarity score of 0.181, $t14 = 9.89$, $p<0.001$). The two scan patterns were also similar for Engineers looking at Civil War pictures (score of 0.196, $t14=8.69$, $p<0.001$), at engineering pictures (score of 0.247, $t14=13.23$, $p<0.001$), and at neutral pictures (score of 0.203, $t14=8.11$, $p<0.001$).

A comparison was also made between scanpaths in the immediate and delayed picture recognition tests. A mixed-model ANOVA found reliable a main effect of type of picture [$F2,28=31.84$, $MSe=0.077$, $p<0.001$], and no effect of participant group: [$F<1$], but there was an interaction between these two factors [$F2,56=68.25$, $MSe=0.164$, $p<0.001$]. Pairwise comparisons found that for both American Studies undergraduates and Engineers, scan patterns were reliably more similar between encoding and delayed test when they inspected pictures that were within their domain of interest. That is, American Studies participants had higher string similarity scores for their two recognition viewings of Civil War pictures (similarity score of 0.220) than they did for engineering pictures (0.112) or for neutral scenes (0.099), with both comparisons reliable at $p<0.001$. Engineering students showed the opposite pattern, with greater string similarities for engineering pictures (score of 0.278) than for Civil War pictures (0.093) or for neutral pictures (0.091), and both comparisons were again very reliable ($p<0.001$). All six average similarity scores were greater than the value estimated for chance (all contrasts were reliable at $p<0.001$).

4 Discussion and Conclusions

When we look at a picture a second time, do we look at the same features, and in the same order? In each experiment a set of images of real-world scenes was shown to participants who were familiar with the domain from which they were selected, or not. If scanpath similarity depends upon the viewer's knowledge of the domain of the picture then scanpaths could, in principle, be used to assess a viewer's knowledge. The pictures were shown for a few seconds, and then a recognition test performed, with the students deciding whether they had previously seen each picture. Eye fixations were recorded throughout, and fixation scanpaths were quantified for comparison between the two viewings of the picture. In the second experiment there was a notable addition to the procedure: the recognition test was repeated after a week so that scanpaths could be compared across two viewings that had the same purpose. We also asked about the value of visual saliency in attracting fixations and fixation sequences: do viewers look at visually conspicuous regions in a scene?

The delayed recognition test was introduced because encoding and recognition engage different cognitive processes and so a comparison of scan patterns may reflect these differences rather than any differences between inspection processes. In both experiments, viewers' scanpaths at encoding were more similar to those made when inspecting the same picture at test than would be expected if fixation locations were made randomly. Importantly, scanpaths were similar in successive recognition tests.

The scanpath comparison in the first experiment was between fixation sequences made during encoding and during recognition, and because these are different tasks engaging different cognitive processes, we introduced an additional test in Experiment 2, a second recognition task. This resulted in a new finding, in which a comparison between two recognition tasks, performed a week apart, eliminated the differences between cognitive processes that are inherent in these memory tasks. When comparing the two recognition tasks, in which the same viewers looked at the same pictures on separate occasions, there was a similarity between the scan patterns of the first few fixations. The stability of the scan patterns over time and when the same task is used is of note here.

During both encoding and recognition in the experiments, the regions identified as being visually conspicuous by the saliency map model were fixated more often than other regions, but only for those participants who were not familiar with the content of the picture. Support for the saliency map model is qualified by the extent to which engineering and American Studies undergraduates looked at pictures taken from their own domains of interest and at other pictures. The same pictures were presented to both groups of participants, to eliminate the possibility of any results being attributable to differences between pictures. The saliency map model correctly predicted high proportions of fixations on salient regions for viewers looking at pictures from other domains of interest, but when they looked at pictures from within their domain, there was very little correspondence between the locations of their fixations and the locations of conspicuous items. When an engineer looked at a picture of an engineering plant or a turbine the tendency was to inspect the feature of domain interest rather than the bright, colourful components – the content dominated inspection and the picture's visual characteristics were secondary. Similarly, when American Civil War specialists looked at uniforms, weapons and other artefacts they also responded to the meaning of the items depicted.

The two experiments confirm the predictions of the saliency map hypothesis in the locations of early fixations on pictures of real-world scenes, at least for viewers unfamiliar with the content. Conspicuous regions are fixated more than other regions during the first few seconds of inspection when viewers were encoding the pictures on first viewing. It has been previously reported that fixations are guided to these regions during encoding but not in search tasks [29-31], and the present results extend these conclusions to take into account the prior knowledge of the viewer.

This result is important because it is evidence of individual scan patterns that are picture-based rather than being the product of general scanning strategies or of saliency-driven scanning. If scanpaths were the product of a habitual and stereotypical saccade-generator routine, or indeed of the low-level visual characteristics of an image, then we would expect invariant similarity scores. Instead, when comparing fixation sequences for pictures from different domains of interest, the similarity between scan patterns varied. There was sensitivity to the content of the picture here, with individuals varying their fixations behaviour according to what they were looking at. The knowledge of the viewer influences the way that they inspect a picture, and this raises the possibility of the assessment of their domain knowledge through the observation of their fixation scanpaths. A knowledgeable student will inspect a picture according to its content, whereas an unfamiliar picture will be inspected according to its low-level visual characteristics. Perhaps a student's level of knowledge could be assessed implicitly by observing their eye movements.

A second potential application of the findings is with the user-centered design of computer interfaces. To assess the usability of an interface developers may use rapid prototyping with the early involvement of end users who provide feedback that often involves providing verbal commentaries or reports [35]. Holzinger has argued that we cannot take users' verbal reports at face value, however, and that their actual non-verbal behaviour would provide a preferable measure of interface usability [36]. By comparing the fixation patterns of end users with the patterns provided by expert users, an evaluation could be developed that does not rely upon verbal reports or questionnaires, and that could provide a direct index of a system's intended usability.

References

1. Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., Crundall, D.: Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergon 46, 629–646 (2003)
2. Underwood, G., Chapman, P., Bowden, K., Crundall, D.: Visual search while driving: skill and awareness during inspection of the scene. Trans. Res. F: Psychol. Behav. 5, 87–97 (2002)
3. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychol. Bull. 124, 72–422 (1998)
4. Henderson, J.M.: Human gaze control during real-world scene perception. Trends In Cog. Sci. 7, 498–504 (2003)
5. Knoblich, G., Öllinger, M., Spivey, M.J.: Tracking the eyes to obtain insight into insight problem solving. In: Underwood, G. (ed.) Cognitive Processes in Eye Guidance, pp. 355–376. Oxford University Press, Oxford (2005)
6. Buswell, G.T.: How people look at pictures: A study of the psychology of perception in art. University of Chicago Press, Chicago (1935)
7. Yarbus, A.L.: Eye movements and vision. Plenum, New York (1967)
8. Mackworth, N.H., Morandi, A.J.: The gaze selects informative details within pictures. Percept. Psychophy. 2, 547–552 (1967)
9. Mannan, S., Ruddock, K., Wooding, D.: The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. Spat. Vis. 10, 65–188 (1996)

10. Noton, D., Stark, L.: Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vis. Res. 11, 929–942 (1971)
11. Stark, L., Noton, D.: Scanpaths and pattern recognition. Sci. 173, 753 (1971)
12. Brandt, S.A., Stark, L.W.: Spontaneous eye movements during visual imagery reflect the content of the visual scene. J. Cog. Neurosci. 9, 27–38 (1997)
13. Laeng, B., Teodorescu, D.S.: Eye scanpaths during visual imagery reenact those of perception of the same visual scene. Cog. Sci. 26, 207–231 (2002)
14. Mannan, S., Ruddock, K., Wooding, D.: Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. Spat. Vis. 9, 363–386 (1995)
15. Mannan, S., Ruddock, K., Wooding, D.: Fixation patterns made during brief examination of two dimensional images. Percept. 26, 1059–1072 (1997)
16. Tatler, B.W., Baddeley, R.J., Gilchrist, I.D.: Visual correlates of fixation selection: effects of scale and time. Vis. Res. 45, 643–659 (2005)
17. Henderson, J.M., Brockmole, J.R., Castelhano, M.S., Mack, M.L.: Visual saliency does not account for eye movements during search in real-world scenes. In: van Gompel, R., Fischer, M., Murray, W., Hill, R.W. (eds.) Eye Movements: A Window On Mind And Brain, pp. 537–562. Elsevier, Oxford (2007)
18. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. Soviet Physice-Doklady 10, 707–710 (1966)
19. Choi, Y.S., Mosley, A.D., Stark, L.: Sting editing analysis of human visual search. Opt. Vis. Sci. 72, 439–451 (1995)
20. Sankhoff, D., Kruskal, J.B. (eds.): Time Warps, String Edits And Macromolecules: The Theory And Practice Of Sequence Comparison. Addison-Wesley, Reading (1983)
21. Privitera, C.M.: The scanpath theory: its definition and later developments. In: Proc. SPIE 6057 (2006)
22. Privitera, C.M., Stark, L.W.: Algorithms for defining visual regions-of-interest: Comparison with eye fixations. IEEE Trans. Patt. Anal. Mach. Intell., 22970–22982 (2000)
23. Groner, R., Walder, F., Groner, M.: Looking at faces: local and global aspects of scanpaths. In: Gale, A., Johnson, F. (eds.) Theoretical And Applied Aspects Of Eye Movement Research, pp. 523–533. Elsevier, Amsterdam (1984)
24. Stark, L., Ellis, S.R.: Scanpaths revisited: cognitive models direct active looking. In: Fisher, D.F., Monty, R.A., Senders, J.W. (eds.) Eye Movements: Cognition And Visual Perception, pp. 193–227. Lawrence Erlbaum, Hillsdale (1981)
25. Underwood, G., Phelps, N., Wright, C., van Loon, E., Galpin, A.: Eye fixation scanpaths of younger and older drivers in a hazard perception task. Ophth. Physiol. Opt. 25, 346–356 (2005)
26. Wooding, D.: Eye movements of large populations: II. Deriving regions of interest, coverage and similarity using fixation maps. Behav. Res. Meth. Instr. Comp. 34, 518–528 (2002)
27. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506 (2000)
28. Parkhurst, D., Law, K., Niebur, E.: Modelling the role of salience in the allocation of overt visual attention. Vis. Res. 42, 107–123 (2002)
29. Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., Bloyce, J.: Eye movements during scene inspection: A test of the saliency map hypothesis. Euro. J. Cog. Psychol. 18, 321–342 (2006)
30. Foulsham, T., Underwood, G.: How does the purpose of inspection influence the potency of visual saliency in scene perception? Percept. 36, 1123–1138 (2007)

31. Underwood, G., Foulsham, T.: Visual saliency and semantic incongruency influence eye movements when inspecting pictures. Quart. J. Exp. Psychol. 59, 1931–1949 (2006)
32. Williams, A.M., Davids, K.: Visual search strategy, selective attention, and expertise in soccer. Res. Quart. Exer. Sport. 69, 111–128 (1998)
33. McCarley, J.S., Kramer, A.F., Wickens, C.D., Vidoni, E.D., Boot, W.R.: Visual skills in airport-security screening. Psychol. Sci. 15, 302–306 (2004)
34. Manning, D., Ethell, S., Donovan, T., Crawford, T.: How do radiologists do it? The influence of experience and training on searching for chest nodules. Radiography 12, 134–142 (2006)
35. Holzinger, A.: Rapid prototyping for a virtual medical campus interface. IEEE Software 21, 92–99 (2004)
36. Holzinger, A.: Usability engineering methods for software developers. Comm. ACM 48, 71–74 (2005)

Knowledge Assessment Based on Evaluation of 3D Graphics Annotation in Lesson Context

Dorian Gorgan, Teodor Stefanut, Madalina Veres, and Istvan Gabos

Computer Science Department, Technical University of Cluj-Napoca,
28 G. Baritiu st., 400027, Cluj-Napoca, Romania
{dorian.gorgan,teodor.stefanut}@cs.utcluj.ro,
{madalina_veres,gabosistvan}@yahoo.com

Abstract. Instead of classical user interfaces based on forms, the graphics annotation proposes another type of user interaction providing support for free expressions, creativity, manual ability, and artistic imagination. Moreover the graphics annotation directly on the surface of 3D eLearning objects requires the development of new graphics algorithms, annotation models, and interaction techniques. This paper reports the experiments developed through the eTrace eLearning Environment concerning with possibility of using the 3D graphics annotation in the context of eLearning lessons. The research addresses a few subjects such as the appropriate model of the 3D graphics annotation that support the automatic evaluation, lesson context based annotation model and processing, interaction techniques for annotation description, drawing and visualization. The work concerns with practical solutions for all the steps through which the teacher defines the annotation related constraints, the system records the annotation specifications within the lesson structure, the user carries out the graphics annotation based answer, the system processes the annotation, quantifies the answer and computes the final mark. This paper highlights mainly the first step of the 3D graphics annotation model description. The research opens new opportunities for graphics annotation based knowledge assessment in eLearning applications against the well known and limited multiple choice questions.

Keywords: 3D graphics annotation, eLearning, knowledge evaluation, lesson structure, automatic evaluation.

1 Introduction

Most of the present internet applications implement text based tools as user interaction techniques, in order to analyze the teaching material included in distance learning lessons: text documents in various formats, images, videos, 3D scenes of objects etc. Forms – which are sets of user interface controls such as lists, buttons, checkboxes, text areas, and menus – are widely used in these environments, allowing the teacher to specify some data format constraints and specific user input values. Sufficient and quite recommended in many circumstances, these kinds of interactions have many limitations when the subjects to be analyzed are images or various graphical objects in

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 145–160, 2008.

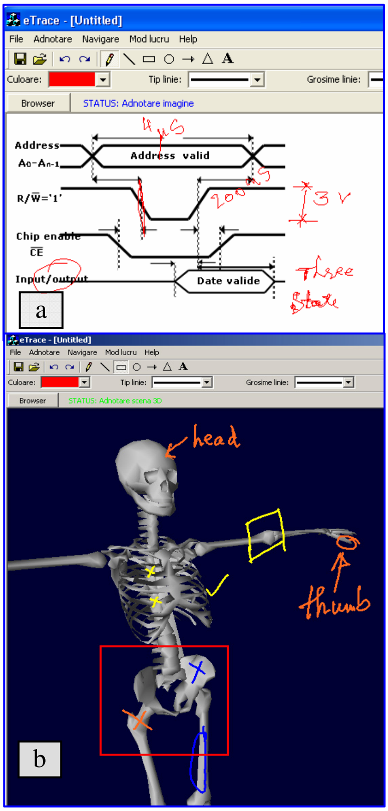


Fig. 1. (a) 2D graphics annotation on 2D document (2/2D), and (b) 2D graphics annotation for a 3D skeleton object (2/3D)

2D or 3D representation. Another major problem of text based communication is the high risk of communication errors due to incomplete or imprecise descriptions made under time pressure (online testing) or in a foreign language (from the user point of view) [1].

Enforcing very few constraints comparing to forms and other common user interface interaction techniques (e.g. menus, multiple choice questions, radio buttons, edit boxes), the annotation allows the user to concentrate more on the task completion activity and less on the input methods and devices and to express his knowledge in a natural and personal way, proving to be a very powerful and efficient user interaction method. Our research has primary focused on creating new interaction methods, developing appropriate tools for using annotations on different types of documents, implementing new algorithms for encoding, managing and assessing the graphics annotation, developing new eLearning lessons models, and moreover, on performing usability tests in order to establish the technical requirements for graphical annotation based user interfaces. The annotation may represent an important tool in problem solving as the user has the ability to be creative, imagine its own answers and use a personal approach to the issue, but in the same time it represents a major difficulty when comes to evaluate the quality of the answer.

In this paper we will present a classification of the annotation types, techniques and models that are necessary in the implementation of knowledge assessment lessons based on the graphics annotation on 3D objects (Fig. 1). These models are created having as a main goal to restrict as little as possible the user free expression through annotation and, in the same time, to allow a complete and correct evaluation of the answer [2]. A detailed example will be described in this paper, concerning methods to define the constraints that are necessary to enable automatic assessment of annotations.

The paper is structured as follows: Section 2 highlights some works related with pen graphics annotations in immersive and non-immersive environments. Section 3 describes the eTrace eLearning Environment. Section 4 presents briefly the main knowledge assessment methods for annotations on 3D objects, while Section 5 describes the basic concepts of the evaluation model description. Section 6 and 7 will

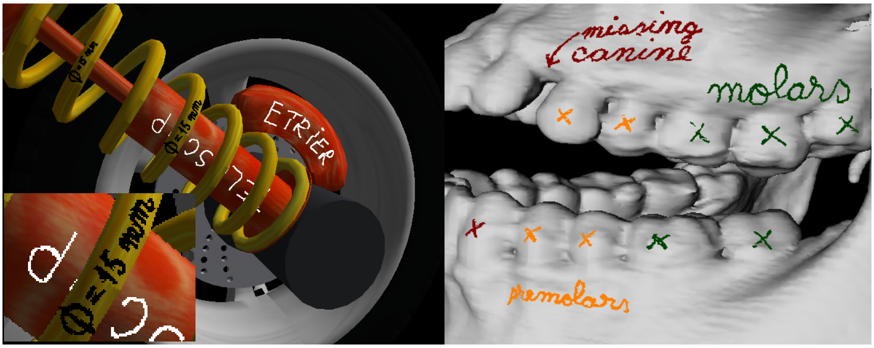


Fig. 2. Graphics annotations directly on the surface of 3D objects

present a detailed example of creating an exercise that will allow the evaluation of the annotation based answers and finally, Section 8 sketches a few conclusions.

2 Related Works

To date, researchers have developed a variety of sketch based interface systems in various domains, like 3D surface modeling [3] and [4], chemistry teaching [5], mathematics [6], music [7] etc., proving that graphical pen interfaces are more reliable, natural and effective than classical ones. All these achievements can be further developed and integrated into eLearning applications, allowing teachers to use these techniques in distance learning lessons.

Collaborative working sessions over the internet, based on annotations in non-immersive 3D virtual worlds, are already implemented in different applications from various domains, described in [8], [9] and [10]. The main advantages in using annotation based in place of text based communication, are the natural and easy ways of information exchange. The transmission of very complex shapes and concepts between working session participants is very much improved due to the better control over the communication errors.

As the information presented is more and more complex, in the present, dynamic media is increasingly used in eLearning environments, as it allows representations of complex concepts in a more natural way. As presented in [11] different studies have been performed in order to establish if the learning performance is or not influenced by the type of media used in the teaching process: dynamic or static. The results proved that learning performance using dynamic media was significantly higher than those of the static textbook lesson when the learning material had a certain level of complexity. In order to allow the user to efficiently analyze and interact with these complex materials, new interaction methods, more natural and intuitive, are necessary.

Rick Rashid noted during the inauguration of the Center for Research on Pen-Centric Computing, at Brown University (http://pen.cs.brown.edu/): "pen based computing has the potential to alter the way students and teachers interact ... positively affect not only the educational process but our working methods and our culture."

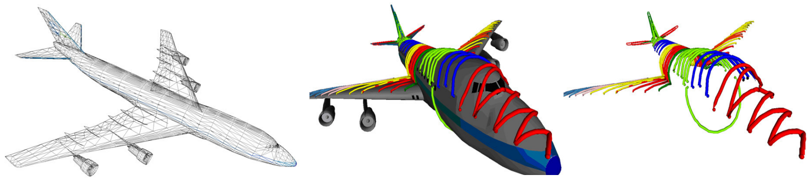


Fig. 3. 3D object is a polygonal model. Graphics annotation becomes an individual entity in the virtual space.

The systems presented above have the ability to draw, record, manage and display the annotations but none of them supports the automatic evaluation of these annotations. As the knowledge assessment process is a crucial component of the teaching process, we are concerned with the possibility of automatic evaluation of the annotations in the context of knowledge assessment in the eLearning applications.

3 eTrace eLearning Environment

The eTrace application has been developed as an eLearning platform centered on graphical annotation as the main user interaction technique [2] and [12]. The eTrace application mainly concerns with the development of new specific interaction methods and tools, the implementation of new annotation encoding and management techniques, and on performing usability tests in order to establish the technical input requirements for annotation based interfaces. This paper will focus on the requirements, solutions and experiments related with the automatic evaluation of the 3D graphics annotation in the process of knowledge assessment.

eTrace implements a basic learning objects modeling and management system, allowing the teachers to easily include HTML format lessons into the application and use its annotation interaction capabilities. It is a client server application that accepts any media file type in the lesson structure. The lesson resources could be: images, sounds, Flash presentations, video files, 3D objects etc.

3.1 Graphics Annotations in eTrace

In eTrace two criteria classify the graphics annotation: *object and document space*, and *the annotation space*. The object or the document could be represented in 2D space (i.e. images, plane text, and documents in various format types) or it could be represented as 3D object in virtual space. As well as the object the associated annotation could be in 2D or 3D space. Conceptually the 2D annotation is performed in a plane. For 2D objects and documents the plane is actually a transparent window placed parallel over the document surface (Fig.1a). For 3D objects the 2D annotation is performed on a projection plane moving around the 3D scene (Fig 1b). The 3D graphics annotation is accomplished inside the virtual space, and mapped directly on the 3D object's surface (see Fig. 2, 3 and 4).

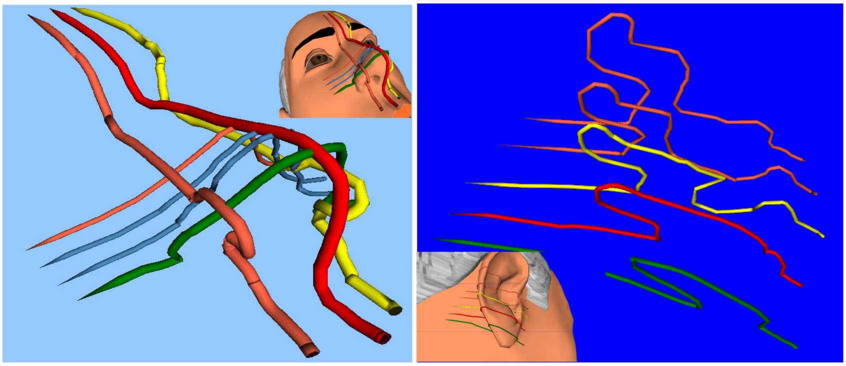


Fig. 4. Graphics annotations by different tools and attributes on 3D objects

3.2 Lesson Development in eTrace

The main generic users of the eTrace lessons are the teacher and the students. The teacher has the ability to create, modify and visualize teaching materials, while the students can only visualize and analyze these materials. Each lesson can be roughly divided in two parts: *presentation of technical considerations* and *knowledge assessment*. In both sections the teacher and the students may perform any type of graphics annotation (i.e. 2/2D, 2/3D, and 3/3D) according with the type of the learning object they are analyzing. Every annotation created into the eTrace platform is than saved into the database together with the necessary information about the annotated resource, the user that is the author of the annotation and the session when the annotation was created. Using this information, the application can correctly identify, display and manage the annotations in future working sessions (Fig. 3 and 4). These aspects will be presented briefly in this paper, only in the extent that they are related to the automatic knowledge evaluation based on graphics annotation. Details about the annotation encoding, storing and managing into the eTrace platform can be found in [12].

Using the eTrace application, the teacher has the ability to include in their lessons exercises that require the students to give their answers using graphics annotation. As mentioned above, the system will record and store students answers into the database and based on this information will make the evaluation and will compute the corresponding mark. Additionally, in order to enable the automatic evaluation support, the teacher must provide into the lesson's structure all the data required by the system, as described later in this paper.

The research aims to find out the answers to these questions: what is the appropriate conceptual model for 2D and 3D graphics annotation in order to support the automatic evaluation? What is the best way for the teacher to describe interactively the annotation model? What are the graphics algorithms that offer to the user the needed support to perform the interactive annotation on 3D surfaces? How are described, combined and executed the graphics and spatial constraints? How is quantified the quality of the annotation in a specific context? How is transformed a set of complex constraints into a final mark?

4 Knowledge Assessment Methods Based on Graphics Annotations on the Surface of 3D Objects

The answers represented through annotation can be evaluated by the teacher visually, if their number is not very large. In eLearning systems, since every teacher can have thousands of students that may attend the classes and complete tests through annotations, the visual assessment is not a reliable solution. Therefore we focused our research on studying different techniques and algorithms that are necessary in order to provide automatic evaluation support for graphics annotations in eLearning environments. The experiments have been achieved through eTrace application [12].

The graphics annotation is used by teacher and students to communicate additional information to the materials already integrated into the lesson. For example, when presenting a lesson to the students in a passive working session – when the students are able only to visualize the presentation – the teacher can use static annotations to explain better some concepts and aspects presented in the teaching materials. At the end of the presentation session these annotations are integrated into the content of the lesson, and can be visualized later by the students without any other intervention of the teacher. In interactive presentation parts of the lesson the teacher and the students are able to add annotations in order to communicate complex information in a natural manner.

When working separately, in individual sessions, every student may use annotations to analyze the learning content provided by teacher, making specific graphical notes. These annotations are recorded in the eTrace platform as separate entity, as they are not part of the lesson content, and is identified as well by the user's name who is actually the author.

In the situations presented above – passive presentation, online or individual learning sessions – there are no limitations for the annotations that can be performed by any user – teacher and student. In the case of graphics annotations that are the subject of the evaluation (e.g. through the knowledge evaluation sessions) a few constraints must be considered in order to avoid the computational errors that might occur, and to ensure a correct and uniform evaluation. Through this paper we will present a few possible solutions for the automatic evaluation of graphics annotations in the context of knowledge assessment. The types of annotations on 3D objects that have been considered for this research are: single and continuous curve, stroke, polyline, and freehand [2].

One of the most important elements in the annotation evaluation is the context of its creation. In the evaluation process it is imperative to take into account the purpose and the meaning of the annotation as identical annotations can have totally different meanings in different contexts. For example, students may use same annotation to draw some graphical items or to highlight different elements from the annotated resource (e.g. using question marks, ellipses, rectangles etc.). At the evaluation time, the system has to know precisely which are the features to be evaluated, and how to measure the parameters and which are the mathematical formulas needed to transform the measurements into grades and finally the grades into marks. Actually, the system should have the ability to compute the quality of the answer from quantitative measurements in specific contexts.

There are three distinct steps that must be outlined:

a) *The description of the evaluation model* – provides to the system all the information necessary in the context of the lesson to make possible the automatic evaluation for knowledge assessment at runtime;
b) *The evaluation of the annotation at runtime* – using the description of the evaluation model and the current graphics annotation, the system evaluates and grades the user's annotation;
c) *The interpretation of the results* – the application processes the results in the specific eLearning context of the current lesson and computes the quality of the answer.

When creating every exercise, the teacher must describe through the evaluation model all the data needed by the application to complete the runtime evaluation and result interpretation. In Section 5 we will present some concepts and solutions that can be used by the teacher in the description of the evaluation model at the lesson creation time. All the cases which are not the subject of the description through the evaluation model are to be analyzed visually by the human evaluator.

This paper consider but not concerns with detailed presentation of specific techniques of drawing on 3D surfaces, graphics algorithms, and user interaction techniques.

5 Evaluation Model Description

All the information needed by the system for automatic evaluation must be provided by the teacher at authoring time, when the exercise is created. First step is to establish the context and the purpose of the annotation to be assessed, to ensure that the students will properly understand the requirements and the system will compute correctly the answers.

5.1 Purpose of the Annotation

In the context of the lesson the annotation may be used as a user interaction technique by which the teacher or the students *draw* a shape or *select* a set of visible items. For instance, by drawing, the students describe a shape that has a specific meaning in the context of a particular lesson. The system considers a limited number of types for the shape the user is drawing, and evaluates the annotation according with a set of constraints already defined in the evaluation model. Actually the constraints address the dimension and the orientation of the graphics shape. The automatic evaluation algorithms consider the following drawing parameters: profile, time, direction and succession of the shape.

5.2 Shape Description

Along with the purpose of the annotation, one of the most important evaluation criteria is the *shape* of the annotation. The teacher must define the constraints in the evaluation description model in order to allow later the application to accurately

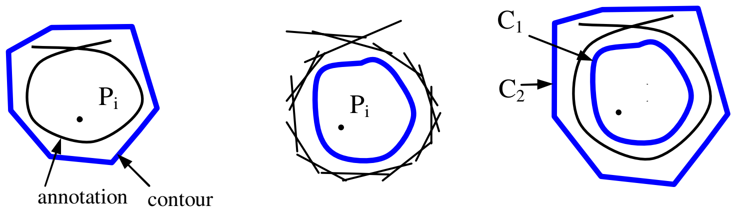


Fig. 5. Defining shape constraints by contour

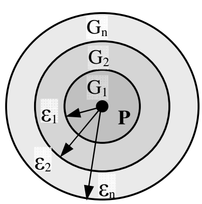
evaluate and determine the correct shapes for the answers, when the lesson is executed and the student performs the graphics annotation. Our research concerns with three different techniques, called *contour*, *key points* and *shape pattern,* which are used individually or in any combination, in order to describe these constraints.

5.2.1 Contour Constraint

The *contour* is defined by a closed poly-line directly drawn on the surface of the 3D object without picking up the pen (Fig. 5). The closed state of the contour is one of its most important properties, so the application will automatically connect the first and last point with a line generated on the object's surface. Although the line used to draw the contour can be any of the three types presented in Fig. 5, the application will approximate the smooth curve and the freehand contour by the polyline contour [2].

The contour based evaluation method can be used by two ways: (a) the contour is described by the teacher as a constraint in the evaluation model, and the student's graphics annotations must be inside or outside that contour; (b) the contour is the student's answer such a selection or a drawing, and it must comply with a given constraint defined by the teacher (Fig. 5). For instance, two different contours could be defined as a constraint, in order to describe a specific area on the surface of the 3D object, where the student's answer should be placed (see Fig. 5.c, where the student's annotation must be outside C_1 and inside C_2).

The *contour* related techniques are used to determine the quality of the student's answer by computing the percentage of the annotation points that complies with the given constraints (i.e. inside/outside the contour).



$\varepsilon_n = n * \varepsilon_1$ or
$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are distinct values

If the annotation is passing by a minimum distance d, $\varepsilon_{n-1} < d < \varepsilon_n$ the grade takes the value G_n.

Fig. 6. Key point description

5.2.2 Key Points Constraint

A second technique used to describe constraints over the annotation shape is the set of *key points* placed on the 3D surface of the object. For each key point can be defined different parameters that are used in the evaluation process (Fig. 6).

The global grade of the annotation evaluated by this technique is defined as a weighted average of the grades established individually for every key point. Through a mathematical formula the teacher can specify that all the grades are equally important to the final grade or can establish that some points are more important than others. By this way, for every point P_i the teacher can specify a fraction value v_i and can define how the grade will be computed. For instance:

$$G = \frac{G_1 * v_1 + G_2 * v_2 + \ldots + G_n * v_n}{v_1 + v_2 + \ldots + v_n} \tag{1}$$

Key point method could express complex requests in terms of precision, annotation direction and time intervals.

5.2.3 Shape Pattern Constraint

The teacher can define his annotation as a *shape pattern* constraint in the annotation definition model. This shape will be used to evaluate the annotations performed by students. The prototype annotation provided by the teacher is automatically processed by the system and transformed into a *shape pattern* using some parameters, such as ε values defined for the key point constraint (see Fig. 7).

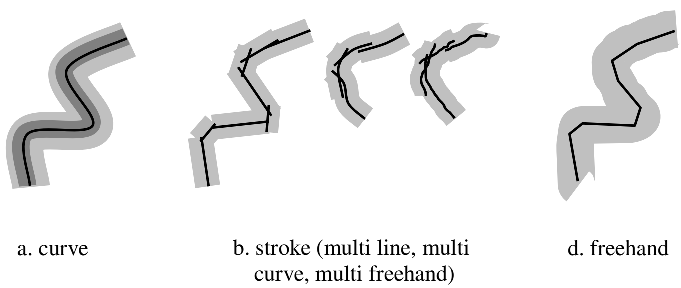


Fig. 7. Defining shape constraints by shape pattern

The student's annotation to be evaluated will be compared with the shape pattern using the given ε values, and the grade will be computed in the same manner as for a single key point.

5.2.4 Time Description

Time is an important criterion in evaluating the graphics annotation getting information on the drawing speed, succession of events, overall time, intermediate delays etc. It represents an important factor that enables the creation of exercises that simulate real situations. In surgery for instance, many interventions are time constrained and, further more, must be chronologically ordered.

For the evaluation of the annotation the teacher can define a global time value or, using shape constraint based techniques, he can refine the conditions and set up time intervals for different sections of the annotation. For instance, in combination with key points, the teacher could specify for each key point the relative time to the beginning of the annotation, when that point should be reached by the student, or the time interval between two consecutive points [2].

Additionally, the teacher has to include into the time constraints the description for the annotation grading. For instance the student must draw the annotation from P_A to P_B in 5 seconds. Some ways to define the delay penalty are as follows:

- for every 1 second delayed the penalty will be value V;
- for a delay between 1 and 3 seconds, the penalty will be V_1, for a delay between 3 and 5 second will be V_2 and for a longer than 5 seconds delay the mark is 0.

The techniques already described could be mixed by the teacher and transformed into a quite complex and particular answer assessment mechanism. For example, the key points, the contour, and the time can be combined to describe a very delicate incision that must be accurate, precisely located on the organ, firmly executed and done in a very short time.

6 Samples of Knowledge Assessment by 3D Graphics Annotation

Let us exemplify a few cases that highlight the 3D graphics annotation techniques in the context of eLearning objects, which are heavily achieved by the multiple choice questions or by text based approach:

6.1 Selection by Contour on the Surface of 3D Objects (Figure 8a)

The student should understand in the context the notion of "affected area" in order to achieve the annotation. By the annotation based technique the student may express a quality of the answer in a quite large range of the drawn contours, instead of the multiple choice question where he/she must choose one answer from a limited and explicitly given set (e.g. one of four);

6.2 Selection of an Area of Interest between Two Contours on the Surface of 3D Objects (Figure 8a)

The area of interest is defined by the inside of the first contour and outside of the second one;

6.3 Drawing Inside a Contour (Figure 8a)

The student answer must be inside the non-visible contour defined by the teacher.

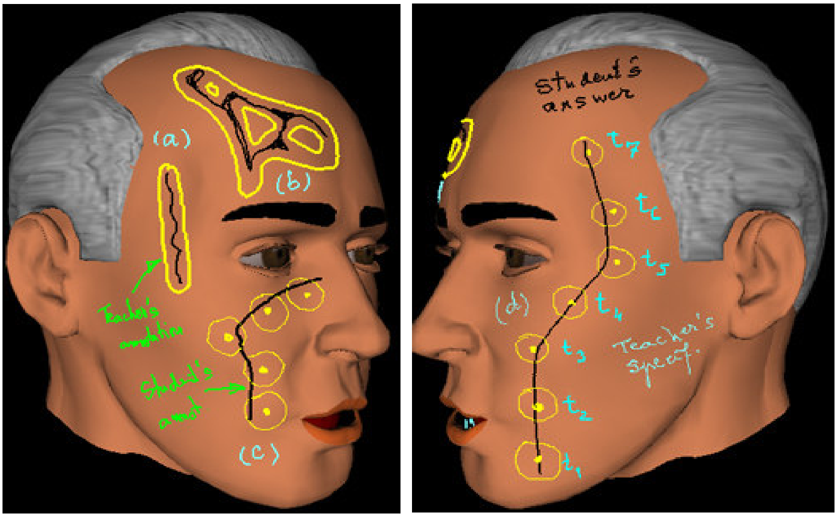


Fig. 8. Evaluation of the graphics annotation on the surface of 3D eLearning objects: (a) inside one contour; (b) checked by a set of contours; (c) checked by a sequence of key points; (d) checked by a sequence of chronological key points

6.4 Drawing Inside an Area Bordered by a Few Contours (Figure 8b)

The answer must be inside the area defined by the non-visible contours.

6.5 Drawing Checked by a Sequence of Key Points (Figure 8c)

The annotation should pass closer to a sequence of key points.

6.6 Drawing Checked by a Sequence of Chronological Key Points (Figure 8d)

The annotation should pass closer to a sequence of key points according with temporal constraints.

6.7 Drawing Checked by Combinations of Contour, Key Points, and Temporal Constraints

The teacher defines a set of constraints according to which the student's annotation must comply.

7 Example of Evaluation Model Description

Let us analyze the solution for 3D graphics annotation model description. Let us consider the following case:

On the 3D model attached please represent the incision line needed to remove the affected skin. Use a continuous annotation and confine all the needed criteria: position, direction, speed.

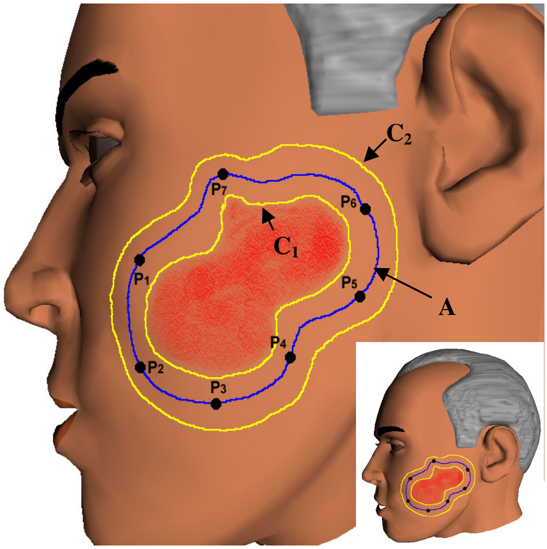


Fig. 9. Affected area removal exercise

The answer to this exercise can be very complex and difficult to evaluate from many aspects. To obtain the best result, the teacher has to define multiple criteria for annotation assessment and combine them in order to allow the system to perform a complete analysis and accurate mark computation.

First, the teacher should define the scale used to represent the marks. It can be from 1 to 10, 1 to 100 or any other numerical values. This scale will be used also to compute the grades obtained separately from every constraint applied. For this example we will use a scale from 1 to 10.

Area related constraints. For the first constraint, the teacher should describe for the system the area that is to be removed, by drawing a contour around it (C_1 in Fig. 9). The students should not annotate the answer inside this area as the entire surface contained in the contour will be removed. The grade for this contour (G_{C1}) can be defined as follows:

$$G_{C1} = \textit{the percentage of annotation points outside the contour}$$

This way, if 80% of the annotation points are outside the contour C_1 the grade computed for this constraint is 8 (according to the scale we have chosen).

To ensure that the students will not remove a very large portion of the skin, the teacher should define a second contour (C_2) with the specification that the points from the answer annotation should be inside. As this is a very important restriction that should not be crossed in any circumstance, the teacher can define the grade of this criterion as:

$$G_{C2} = \begin{cases} 1, \text{if } 100\% \text{ of annotation points are inside } C_2 \\ 0, \text{otherwise} \end{cases} \quad (2)$$

As the area between the C_1 and C_2 cannot be very narrow (because the C_2 constraint is a very restrictive one and should not be applied in minor deviations) and at the same time not all the annotations between the two contours are correct, the teacher should define another criterion that will allow the system to perform more precise evaluations of the answers. This can be achieved by drawing an annotation model (A_M) that represents the correct answer of the exercise and by defining the computation method for the grade of this constraint similar to the example below:

$$G_{AM} = \begin{cases} A_{max} - d, d < A_{max} \\ 0, \text{otherwise} \end{cases} \quad (3)$$

In this formula, A_{max} represents the maximum value from the marks scale chosen and d is the average distance measured in the metric chosen by teacher (i.e. cm, mm, pixels etc.) between the annotation model and annotation answer. We can consider 7

as a numerical example value for this constraint, as for the C_1 constraint we established a grade value of 8.

Direction and time related constraints. Two of the most important criteria that can be defined by the teacher in order to assess the answers for this example are the direction and the time interval needed for the creation of the annotation answer. There are many situations in real life when these two criteria are very important and must be taken into consideration for the assessment process, as in the case of incisions, sutures etc. One of the most flexible methods to describe these constraints is the use of key points. As it is presented in Fig. 9, the incision can be approximated by 7 points (or more if desired). These points will be placed by the teacher on the object's surface using the annotation, at the authoring time. The mandatory attributes that the teacher must define for every point in order to achieve automatic annotation evaluation are:

- ε_i values and their associated $G_{\varepsilon i}$ grades (see Fig. 8), that can be identical to all the points or different for some of them, i = 1..7
- the formula to compute the global grade of the evaluation realized with key points (G_P), from the points grades G_i

Optionally, some other attributes related to the key points may be defined:

- time t_i relative to the beginning of the annotation, when the student must reach P_i; the grade for this attribute at point *i* will be G_{Ti}
- time intervals $\{t_i, t_j\}$ specifying the maximum (or minimum) time the user has to reach from P_i to P_j; the grade for this constraint will be G_{Tij}

- the points order that user must respect (ex. $P_1 \rightarrow P_2 \rightarrow \ldots \rightarrow P_7$); the grade will have the symbol G_D

If these optional attributes are defined, the teacher must also specify the method to compute the point grade G_i and the global key points evaluation grade G_P, using the results from the assessment of every attribute.

For this example we define:

- each point P_1 to P_7 has 10 ε_i areas with ($G_{\varepsilon i} = i$). The maximum grade for every point is 10 and minimum 0 (if no ε_i vicinity is crossed by the student's annotation)
- G_7 and G_5 represent 30% of G_P (15% each)
- G_2 and G_3 represent 10% of G_P (5% each)
- G_1 ; G_4 and G_6 represent 60% of G_P (20% each)

$$G_P = 0.15*(G_7 + G_5) + 0.05*(G_2 + G_3) + 0.2 * (G_1 + G_4 + G_6) \tag{4}$$

If we consider numerical values $G_1 = 10$, $G_2 = 9$, $G_3 = 10$, $G_4 = 8$, $G_5 = 9$, $G_6 = 10$ and $G_7 = 1$ it will result $G_P = 8.05$.

Overall time constraint. As a final criterion for the annotation assessment, the teacher can define the overall time needed to create the answer, according to the time attributes specified for the key points. For this example we will consider that this time interval must be less than 7 seconds but not less than 4. A penalty of 25% will be applied for every second that passes before or beyond these limits. The grade for this constraint has the symbol G_T and we will consider a value of 7.5, as the annotation has been created in 3 seconds.

Final mark. After all the grades for the constraints are established, the system needs a rule to compute the final mark. This rule can be defined using mathematical (*, -, +, / etc.) operators and a weight value for every grade, or using Boolean operators to express the obligation of the student in obtaining a certain mark to a specific constraint. For this exercise we will define the following rules:

- G_{C1} has a weight of 10% from the final mark
- G_{C2} must be 1 in order to consider the mark valid
- G_{AM} has a weight of 40% from the final mark
- G_P has a weight of 30% from the final mark
- G_T has a weight of 20% from the final mark and must be greater than 0 in order to consider the mark valid

The final mark for this exercise will then be computed as:

$$M = (0.1*G_{C1} + 0.4*G_{AM} + 0.3*G_P + 0.2*G_T) \text{ AND } (G_{C2} = 1) \text{ AND } (G_T > 0) \tag{5}$$

With respect to the numerical values we established, the final mark is:

$$M = (0.1*8 + 0.4*7 + 0.3*8.05 + 0.2*7.5) \text{ AND } (G_{C2} = 1) \text{ AND } (G_T > 0) = 6.82 \tag{6}$$

As we can see from the formula of the final mark, if the student doesn't integrate his annotation in the area marked by C_2 or the time to create the answer is too long, the final mark is going to be 0, even if the computed value is actually 6.82.

8 Conclusions

As the information presented in eLearning environments is increasing in complexity, the usage of dynamic learning objects becomes a necessity. The classical interaction techniques, usually text based, are not adequate anymore for the interaction with these complex representations (animations, images, 3D objects etc.), so the development of new techniques is a must. The annotation is one of the most promising interaction methods developed in the present, offering to the users more natural ways of interaction with learning objects, but its implementation is limited by the difficulty of evaluation process.

The experiments developed and executed in eTrace have proved the efficiency and the usability of the automatic assessment for pen based graphics annotation. The future research will concern with the implementation of different methods and algorithms for runtime evaluation and processing of the results. The future work will be focused on gesture oriented annotation, semantics model of 3D graphics annotation, algorithms for drawing and processing the annotation for different types of 3D surfaces, usability and efficiency evaluation.

Acknowledgments. The work reported through this paper was supported by the I-Trace Socrates Project, funded by European Community, under the Contract 223434-CP-I-2005-IT-Minerva-M.

References

1. Moodle - A Free, Open Source Course Management System for Online Learning, `http://moodle.org/`
2. Gorgan, D., Stefanut, T.: Knowledge Assessment for Annotation Techniques on Medical eLearning Objects. In: CBMS 2008 – 21th IEEE International Symposium on Computer-Based Medical Systems, Jyväskylä, Finland, June 17-19, 2008, pp. 379–384. IEEE Computer Press, Los Alamitos (2008)
3. Kara, L.B., Shimada, K.: Construction and Modification of 3D Geometry Using a Sketch-based Interface. In: EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling (2006)
4. Masry, M., Kang, D.J., Lipson, H.: A freehand sketching interface for progressive construction of 3d objects. Computers and Graphics 29 4, 563–575 (2005)
5. Tenneson, D.: Chempad: Visualizing Molecules in ThreeDimensions. In: Workshop on the Impact of Penbased Technology on Education (2007)
6. Zeleznik, R., Miller, T., Li, C.: Designing UI Techniques for Handwritten Mathematics. In: EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling (August 2007)
7. Forsberg, A.S., Dieterich, M.K., Zeleznik, R.C.: The Music Notepad. In: Proceedings of UIST 1998, ACM SIGGRAPH (1998)
8. Jung, T., Gross, M.D., Do, E.Y.-L.: Annotating and Sketching on 3D Web models. In: Proceedings of the 7th international conference on Intelligent user interfaces, San Francisco, California, USA, pp. 95–102 (2002)
9. Goddard, T., Sunderam, V.S.: ToolSpace: Web Based 3D Collaboration. In: Proceedings of the 1999 International Conference on Intelligent User Interfaces, Redondo Beach, CA, USA, January 5-8, 1999, pp. 161–165 (1999)

10. Daily, M., Howard, M., Jerald, J., Lee, C., Martin, K., McInnes, D., Tinker, P.: Distributed Design Review In virtual Environments. In: Proceedings of the third international conference on Collaborative Virtual Environments, pp. 57–63 (2000)
11. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
12. Stefanut, T., Gorgan, D.: Graphical Annotation Based Interactive Techniques in eTrace eLearning Environment. In: Proceedings of the 4th International Scientific Conference eLSE 2008 - eLearning and Software for Education, Bucharest, April 17-18 (published, 2008)

The Influence of Instruction Mode on Reaching Movements during Manual Assembly

Sonja Stork, Christian Stößel, and Anna Schubö

Ludwig Maximilian University Munich, Department Psychology,
Leopoldstr. 13, 80802 Munich, Germany
sonja.stork@lmu.de

Abstract. The mode and design of instruction presentation have a high impact on human task performance. Especially in a working context, instruction modes have to be developed according to the properties of human information processing like attentional selection and action planning. Accordingly, tasks have to be divided in subtasks in order to find the best way to support each subprocess. The present study presents an experimental setup and first results in order to investigate the impact of different instruction modes on the first goal-directed movement during an assembly task. Results are discussed with respect to implications for instruction presentation and the development of a system for adaptive worker assistance.

Keywords: Instruction presentation, information processing, motor behavior, human performance, augmented reality.

1 Introduction

When confronted with several different working processes, task complexity and worker demand increase with respect to the afforded flexibility and efficiency. One way to improve work performance is the development of task instructions which optimally support the worker, especially while executing complex or new tasks [1]. Accordingly, educating novices in new working environments is a challenge for instruction design as especially workers who are unfamiliar with a task need precise , step-wise information. Usability of manuals can be improved by the use of style guides, e.g. while optimizing the graphical layout of instruction details. Nevertheless, besides the use of general principles, instructions have to be adapted to specific task properties and to the cognitive processes involved. Therefore, usability tests have to compare instruction modes in order to find the optimal way to support different subtasks during the work process. Traditionally, usability tests often consist of evaluations based on questionnaires or interviews, which are useful for getting information on the subjective satisfaction and acceptance levels of the user [2]. Moreover, action analysis and field observations should be included within the development process of user interfaces (For an overview of different usability evaluation methods see [3]). The analysis of task performance as an objective measurement can be compared with the subjective

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 161–172, 2008.

impression of the user. Especially for the usability criterion 'efficiency' the measurement of completion times and error rates are relevant for building a common metric for usability engineering. Analyzing task demands can provide insights in certain related subprocesses as opposed to investigating stress per se, which can have several causes and might be linked to very different subprocesses [4]. Performance parameters have to be extracted depending on the relevant task and subtasks, and the working process has to be segmented into useful action units. In general, performance times and error rates as well as dwell times on instruction details can deliver valuable information on human task performance. Moreover, the analysis of specific movement parameters of human motor behavior can help to evaluate benefits of different instruction modes as these can help to specifically differentiate what part of a task is difficult and needs closer investigations.

The working scenario used here consists of a manual assembly task where toy bricks have to be mounted according to detailed instructions. This task can be divided roughly into a selection phase where the relevant parts have to be found and grasped and into a mounting phase. The following section deals in more detail with the involved subprocesses.

1.1 Information Processing and Selective Visual Attention

In order to support a worker adequately, relevant task information has to be presented at the right time and place. Only adaptation to the task demands can yield to effortless information processing that does not overburden the worker. Therefore, understanding the relevant cognitive processes and their limitations as well as determining the resulting bottlenecks that are likely to occur during the working process is one key issue for the development of adaptive support [5,6]. Relevant cognitive functions related to information processing during instruction presentation and manual assembly involve the whole spectrum from perception, attention, memory to action planning [7].

The information processing framework can be used in order to describe the relevant processing stages in more detail, which are assumed to be passed through in a rather serial manner. The perceptual processing involves the stimulus preprocessing, feature extraction and stimulus identification. That is, information presented in an instruction has to be located first and attention has to be directed to the relevant details. During manual assembly, for example, the type and number of necessary parts for a workpiece need to be identified. When the first part has been chosen from the instructions, its type (e.g. form and color) and number have to be memorized and the respective part location has to be found at the region of the working environment where the parts are stored (e.g. in a box close to the working area). Next, the relevant action or response (e.g. grasping movement with the left or right hand) has to be selected internally and executed. If necessary the ongoing motor response has to be adjusted and corrected online (e.g. if the selected box for stored assembly parts is empty) [8]. These subprocesses should be supported adequately in order to facilitate the working process. Especially in highly demanding work steps or while learning new tasks can the efficiency of task execution be increased and errors be avoided by choosing the best way of support.

One way to facilitate the working process is to use the instruction presentation to direct attention to the relevant working area. Attention can be directed voluntarily (endogenously) by instructions or intention or reflexively, by external exogenous cues, like a salient object in the environment (e.g. a red assembly part among a large amount of blue and green parts). Although attention can shift covertly without moving the eyes to a certain spatial region, the number of objects and locations that can be attended simultaneously is limited. During manual assembly for example, it is not possible to attend to the workpiece and to the instructions at the same time if they are spatially too far apart. Therefore, attention has to be shifted between the instruction and the workpiece, and every attention shift takes time and effort. Accordingly, instruction presentation should guide the worker to attend to the correct locations and objects thus avoiding unnecessary time consuming attention shifts.

The task of selecting the next relevant assembly part among other irrelevant parts resembles a visual search task, where a target is surrounded by several distractors. The task difficulty increases with increasing similarity between target and distractors, that is, with increasing feature overlap where no target pop-out can occur [9,10]. A method to direct the workers attention is the use of salient exogenous cues, like highlighting the relevant box which is then surrounded by non-highlighted boxes. This method enables bottom up or stimulus driven selection in contrast to top down or task driven selection. ¿From fundamental research, it is well known that spatial cues can accelerate attention shifts, as well as eye movements and simple reaction times [11]. These properties of spatial visual attention and motor actions can be applied in order to make worker performance more efficient. Naturally, however, such support needs to be specifically designed in order not to yield the opposite effect and distract the worker. Thus, careful experimentation seems to be a requirement for the development of support systems.

1.2 Task Performance and Motor Control

There has been extensive research in order to investigate body and limb movements during human-computer-interaction (HCI). Accordingly, there exist several models that predict characteristics of human motor behavior. Fitts's law is a model of human movement execution which predicts the time required to rapidly move to a target area, as a function of the distance to the target and the size of the target [12]. Here, Fitts tried to link information theory with the difficulty of movement execution. With this model the speed-accuracy tradeoffs can be predicted. In this context aiming movements with rapid voluntary actions under different conditions have been investigated. For example, the model has been applied as an instrument for usability engineering while predicting text entry rates on mobile phones. With respect to human-computer interaction, motor behavior in working scenarios has also been investigated for the development of user interfaces. For example, decreased movement times and error rates as well as an increased accuracy have been shown for stylus input in contrast to finger input [13] and for touch screens in contrast to track balls, while standing and walking [14]. Moreover, Guiard's descriptive model of bimanual skill describes movement patterns of the preferred and non-preferred hand [15]. This model was used for example in a study of bimanual control and desktop computer affordances (for an overview see [16]). Such predictive models and parameters can be used to objectively evaluate human task performance. In addition to the total movement time, speed-accuracy tradeoff and hand selection, movement parameters like latency of movement onset and peak velocity can also provide useful information about task complexity and task difficulty. Therefore, in the present study, task performance with different instruction modes was analyzed on the basis of these movement parameters.

Besides more objective task parameters like distance and size of a to be grasped object, there exist task demands that cannot as easily be measured and evaluated because they result of the interplay of certain subtasks, which influence each others In his multiple-resource theory, Wickens proposed that mental resources can be divided to four dimensions associated with perceptual processing (visual versus auditory), processing codes (spatial versus verbal), processing stages (perception and central processing versus responding)and response modalities (manual versus verbal) [17]. Each of these resources has its specific capacity limitation resulting in decreased performance in case that two tasks needs the same resource. An analysis of those task aspects that might interfere with respect to a shared resource enables the prediction of multiple-task performance and mental workload [18]. The multiple-resource theory can be applied to the assembly situation, where different spatial areas have to be monitored (e.g. instruction area, parts area, work piece) and different manual actions have to be performed with both hands at the same time.

1.3 Support Via Optimized Instructions

Instruction modes can be classified into paper manuals, manuals on PC and different forms of Augmented Reality (AR) on a continuum with increasing support by the computer system [19]. Computer generated information can be dynamic as opposed to static text book or manual information. For example, transitions of information presentation are possible, where objects appear and disappear over time [20].

In AR applications, the environment is overlayed with additional information at the exact position where it is needed. In the context of assembly instructions it can show where the next part can be found and at which position it has to be mounted. For example, the aerospace company Boeing has developed a systems which facilitates the mounting and wiring of cables in the airplane [21]. One possibility to implement AR is the head-mounted-display (HMD), where the information is projected on a semitransparent mirror which is mounted via a device on the head in front of the eye. The HMD device has the drawback that it is rather uncomfortable to wear and that it has a relatively low resolution. Therefore, a more convenient method seems to be the augmentation of reality via projections on the working area. In general, the AR technology can reduce eye and head movements, accelerate attention shifts and support spatial cognition [22]. In one study, comparisons between different instruction modes showed an advantage of AR techniques for total assembly times and step times, especially for rather complex tasks like wiring (but not for intuitive or repetitive tasks) [19]. These results resemble those within educational research, where benefits of dynamic animated learning material have been demonstrated mainly with complex learning content [20]. Moreover, in the same study AR benefits were mainly observed in the part selection phase and when parts had to be positioned, but not during the manual execution itself [19].

On the basis of these applied results and spatial cueing experiments in fundamental research, the interesting question remains whether not only total step times but also different movement parameters show benefits from the AR technique.

2 A Working Scenario: Selecting and Grasping Objects for Manual Assembly

The present experiment was conducted within a working scenario consisting of the task to build an object with toy bricks according to detailed instructions which were presented in different modes. According to previous findings, the assembly times per part varied with the instruction modes: the mean assembly times for each part were about 10 seconds faster with contact analog highlighting of relevant boxes and projection of the instruction to the working area in comparison to pure monitor instructions [8].

Nevertheless, these times per part include all processes from scanning the instruction, attentional selection of relevant parts, planning and executing the reaching and grasping movements and the assembly itself, which takes the largest amount of time within one step. Therefore, the interesting question remains, whether the different instruction modes influence the early phase of part selection during one assembly step. Reaching and grasping movements can be separated to a transport phase, during which the hand moves towards the goal position, and a grasp phase, during which the fingers enclose the goal object [23]. If the attentional selection of assembly parts is really improved in the contact analog condition, than also the reaching movement and grasping movement (i.e. the transport and grasp phase) to the first relevant box should start earlier.

2.1 Methods

Participants. Thirty paid volunteers (10 women, 20 men, aged between 20 and 30 years, mean age 23.7 years) participated in the experiment. Most of them had a technical background due to studying engineering sciences.

Instruction Modes. Three different instruction modes were used which differed in the location and the content of instruction presentation. Depending on the condition, the instructions were presented on the monitor or via a projection on the working area. The experiment was conducted using a between-subjects design, that is each condition was presented to 10 participants. In all modes instructions consisted two areas: on the left, a list showing the amount of all relevant parts within one work step and rendered pictures of the parts and on the right, a picture showing the goal state of the object which has to be assembled. The assembly objects were designed by a CAD program (LegoCad), a software that allows constructing simple 3D models. Between two and seven LEGO® bricks were used to build one object. In order to make the stimulus set more realistic, they were additionally rotated in perspective. Viewing point, shadows and light sources were chosen in a way to show as many details as possible. In the monitor and projection condition, a schematic picture of the box positions was depicted on the top. Dots within these boxes were to deliver the position of the necessary object parts. Therefore, the only difference between monitor and the projection mode consisted in the location of the instruction presentation. In the monitor condition, participants were forced to switch attention between the instruction that was displayed on a computer screen and the work piece and box area over a larger distance than in the projection condition. With projections it was possible to arrange the work piece close to the depicted goal state, thus enabling an easier comparison during the assembly process. Moreover, also the depiction of the relevant boxes (indicated by the dot) was closer to the real boxes. Therefore, part positions are expected to be found more easily. In the contact analog condition, this schematic box view was replaced by the highlighting of the relevant boxes themselves. This highlighting of boxes was supposed to be helpful especially in the selection phase of the movement planning. Fig. 1, shows a bird view of the three resulting instruction views: contact analog (left), projection (middle) and monitor(right).

Setup and Apparatus. The setup consisted of a standard workbench equipped with a beamer which was placed on top. A front-surface mirror was mounted in 45° angle orientation in front of the beamer enabling the projection of

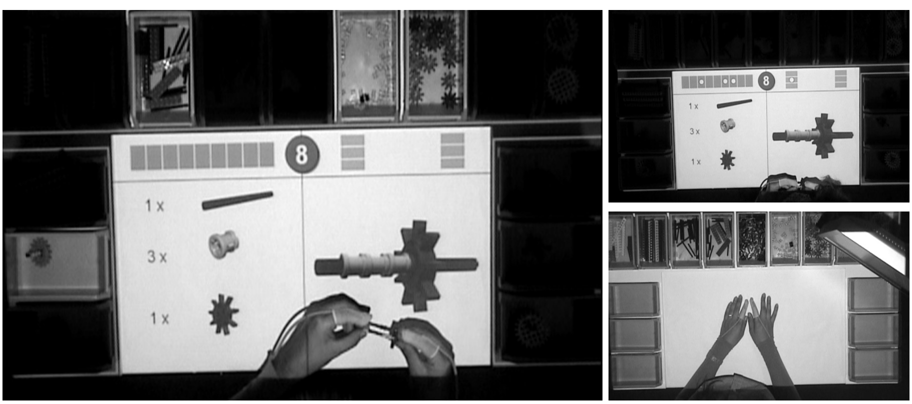


Fig. 1. The three different instruction modes: contact analog highlighting of boxes (left), projection on the work bench (top right) and monitor presentation (bottom right)

instructions directly on the working area. A Polhemus Liberty motion tracking system was utilized for tracking the movement of the right and left hands [24]. This electromagnetic system with 6 degrees-of-freedom consists of a source which generates the magnetic field and 2 markers. Marker positions are determined in x, y, and z coordinates as well as in rotational angles. The sampling rate used here was 60 Hz. Two footpedals were arranged at an easy-to-reach position in order to enable switching to the next (right press) or the previous (left press) assembly step without disturbing the manual assembly process. Footpedal presses could also be employed in order to segment the motion tracking data to meaningful step units. Additionally, the whole session was recorded by a DV video camera. This video stream could be used to evaluate the process according to parameters which cannot be inferred from the Polhemus data.

Procedure. The experimental session started with placing the markers on the tip of the right and left index finger. This was followed by a short training in order to make the participant familiar with the parts to be used and with assembling while having markers on the fingers. Subjects were standing in front of the workbench during the whole experiment. The experiment consisted of building three windmills of different complexity and different amount of steps (15, 18, 19 steps). After pressing the right foot pedal a step started with the next instruction presentation. Participants were instructed to work as fast as possible, but no explicit time constraint was given.

2.2 Data Analysis

After splitting the Polhemus data of each subject on the basis of foot pedal presses to task steps, these task steps were further segmented according to the hand positions. Here, we were interested mainly in the first phase of the step (see grey bar in Fig. 2), that is the first movement to a box after the instruction information has been processed and the first relevant part position has been located for planning and executing the grasping movement. In order to find the movement onset and the grasp onset (the time between start of instruction presentation and first object contact) the point in time when the hand leaves the working area and enters the parts area for the first time was taken as reference point. This was necessary because the first grasping movement could be preceded also by hand movements within the working area (e.g. for arranging assembly parts). Searching backwards from the point in time when the working area was left, the movement onset was calculated as the first point in time when the velocity per second went below 25 cm/sec. It was not possible to use a velocity of 0 cm/s to calculate the movement onset because in several cases the first grasping movement was preceded by little hand movements within the working area so that velocity only rarely decreased to 0 cm/sec. The minimum velocity between entering and leaving again the parts area was used in order to calculate the grasp onset. Here, typically the hands as well as the finger movements reached their minimum because the hand stayed for a moment at the goal position and the finger reached the object borders. Finally, the peak velocity and acceleration between the movement onset and the grasping were determined. The hand velocity

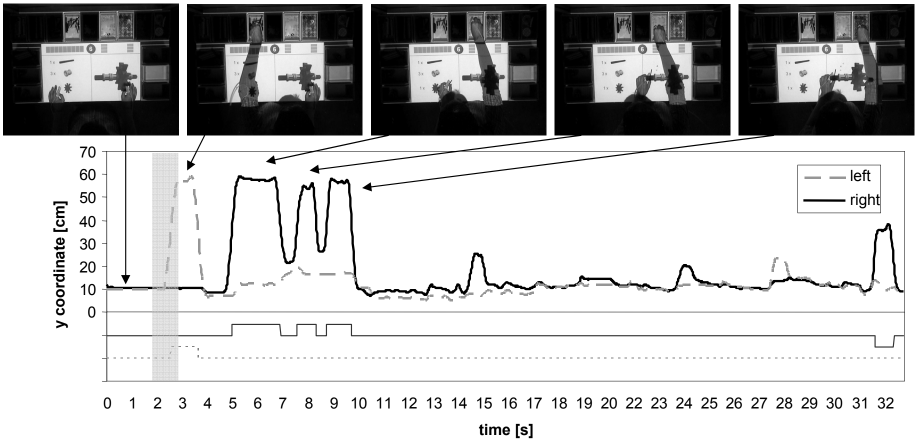


Fig. 2. Y coordinates of the left and right hand during one assembly step (middle), segmentation according hand position (bottom), pictures of respective reaching movments (top)

between movement onset and grasp onset was calculated as Euler distances of the three coordinates x, y and z. The acceleration profile was calculated as first deviation of the velocity. The mean values per subject were corrected of outliners by excluding all values below and over 2 standard deviations. Fig. 2 depicts an example for the segmentation of one task step on the basis of the motion tracking data. In the middle, the y coordinates of the right and left hand are plotted with respect to the time axis. At the bottom, it can be seen how the step can be segmented while using the points in time when the working area has been left or entered: Every upward segment in the plot stands for time within the parts area. At the top, a snapshot from the video stream demonstrates the single actions involved. Starting with the hands in a resting position in the working area, the first grasping movement was performed with the right hand followed by three grasping movements of the left hand.

2.3 Experimental Results and Discussion

On the basis of the cueing effect and mechanism of spatial attention we hypothesized an effect of instruction mode on movement parameters of the first reaching movement within a task step. Fig. 3 shows the mean values of the movement onset latency and the calculated point in time of grasping (left) as well as the peak velocity and acceleration (right) during this movement. Results are in accordance with the previous findings showing an effect of reduced assembly time per part for the contact analog condition [8]. The mean latencies for the movement onset and grasping are shortest in the instruction mode where boxes are highlighted in contrast to the monitor presentation. Moreover, the peak velocity and

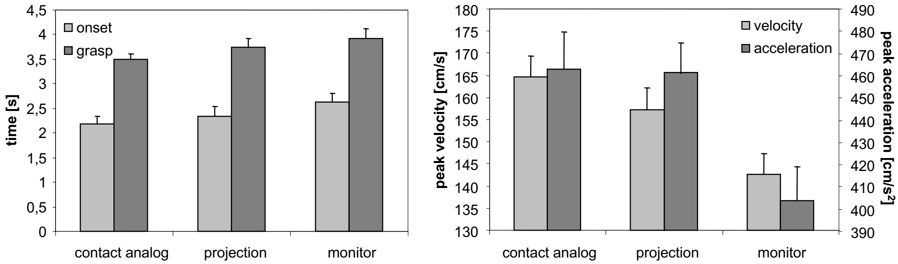


Fig. 3. Means and standard errors for movement onset latencies and time to grasp (left) and peak velocities and accelerations (right) with different instruction modes

acceleration were higher in the contact analog condition. An oneway analysis of variance (ANOVA) over all three conditions revealed a significant effect of condition for the peak velocity ($F(2, 27) = 5.626$, $p < 0.01$). However, the parameters onset, grasp and acceleration showed no significant main effect. Therefore, subsequent t-tests were computed for comparing performance for single conditions. Movement onsets were significantly faster for contact analog presentation compared to monitor presentation ($t(18) = 1.85, p < 0.05$, always one-tailed). A similar time benefit was observed for the grasp onset ($t(18) = 1.877, p < 0.05$). Also the peak velocity ($t(18) = 3.397, p < 0.01$) and peak acceleration ($t(18) = 2.617$, $p < 0.01$) were higher with the contact analog instruction mode in comparison to the monitor presentation. In the 'projection' condition peak velocities ($t(18) = 2.124$, $p < 0.05$) and acceleration ($t(18) = 2.782$, $p < 0.01$) were significantly faster than in the 'monitor' condition. The comparison of 'contact analog' and 'projection' showed no significant effect.

It seems that the contact analog condition enabled faster selection of the relevant part position and faster shifting of attention. However, also the projection without highlighting revealed higher peak velocities and accelerations than the monitor condition. Benefits seem to occur also due to the depiction of relevant box position and task instructions close to the real boxes. Further experiments will be executed using a within-subject design, in order to analyze the effect of box highlighting in more detail. In general, the monitor condition where information on the relevant box position and task information was spatially more separated from the box and working area, seems to lead to the worst performance. Possibly, attention shifts take more time in this conditions and the limitation of resources like spatial attention are more obvious.

Nevertheless, in comparison to the 10 s effect for the whole step time, the ratio of time improvement for the first movement is rather small, that is, there have to be time advantages also within the later phases of the assembly step. Consequently, future analyses will focus also on the subsequent reaching and grasping movements as well as on the assembly phase itself.

3 Conclusions and Future Directions

3.1 Implications for Usability Testing of Instruction Presentation

The present experimental study compared three different instruction modes with respect to their influence on the first reaching and grasping movement during manual assembly. Results showed that with contact analog highlighting of relevant boxes, latencies for the movement onset and grasping movement were shorter and peak velocity as well as peak acceleration were higher during the reach-to-grasp phase in comparison to the monitor presentation. Based on this, one can speculate that movements started earlier because the attentional selection of the relevant part was facilitated by the exogenous cue. Moreover, the spatial cue seems to accelerate the whole movement, possibly because the worker was more confident about having chosen the right box. Additionally, also the projection of instructions on the work area and close to the boxes lead to benefits for movement velocity and acceleration.

The present results give insights into the mechanisms involved in processing of information presented in (assembly) instructions. The results demonstrate that principles of spatial cueing can be used in order to support a worker during task execution. Moreover, the present study demonstrates the value of motion tracking as means for the segmentation of task processes and the evaluation of task performance on the basis of objective measurements. Additionally, the workbench setup used here demonstrates that it is possible to develop a system for worker support designed of a beamer and a front-surface mirror.

3.2 Future Directions for Adaptive Instruction Presentation

Future data analyze will focus also on movement parameters like onset, peak velocity, acceleration and variability as well as movement trajectories during the whole assembly step. Moreover, eye movement recordings will give detailed information concerning the dwell times on certain instruction areas, on search strategies and eye-hand coordination during assembly [25]. This method allows further segmenting the task steps, e.g. in order to test, whether a speeded up first reaching movement is preceded by for example faster eye movements or shorter latencies for the first fixation of the relevant box. In the long run, data collection with markers should be replaced by methods like image processing, but while such systems are still under development it is useful to combine information from different sources following a data fusion concept [26]. Additionally, in order to generate adaptive instructions the system has to be aware of the environment and the current state of the product in the manufacturing process [27]. All together, experimental results and further technical developments will be used in order to optimize an assistive system for manual assembly which adaptively supports the worker.

Acknowledgments

This work was funded in the Excellence Cluster "Cognition for Technical Systems" (CoTeSys) by the German Research Foundation (DFG). The authors would like to thank their research partners Alexander Bannat, Florian Friesdorf, Cordula Vesper, Frank Wallhoff and Mathey Wiesbeck in the CoTeSys project "Adaptive Cognitive Interaction in Production Environments" (ACIPE).

References

1. Sarter, N.: Coping with Complexity Through Adaptive Interface Design. In: Jacko, J. (ed.) HCI 2007. LNCS, vol. 4552, pp. 493–498. Springer, Heidelberg (2007)
2. Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K. (ed.) 11th International Conference on Computers Helping People with Special Needs. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
3. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
4. Andre, A.D.: The value of workload in the design and analysis of consumer products. In: Hancock, P.A., Desmond, P.A. (eds.) Stress, workload, and fatigue, pp. 373–382. L. Erlbaum, Mahwah (2001)
5. Proctor, R.W., Vu, K.-P.L.: Human information processing: An overview for human-computer interaction. In: Sears, A., Jacko, J. (eds.) The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications, 2nd edn., pp. 43–62. CRC Press, Boca Raton (2008)
6. Wickens, C.D., Carswell, C.M.: Information processing. In: Salvendy, G. (ed.) Handbook of human factors and ergonomics, 3rd edn., pp. 111–149. John Wiley, Hoboken (2006)
7. Landau, K., Wimmer, R., Luczak, H., Mainzer, J., Peters, H., Winter, G.: Arbeit im Montagebetrieb [Work in assembly factory]. In: Landau, K., Luczak, H. (eds.) Ergnomie und Organisation in der Montage, pp. 1–82. Hanser, München (2001)
8. Stößel, C., Wiesbeck, M., Stork, S., Zäh, M.F., Schubö, A.: Towards Optimal Worker Assistance: Investigating Cognitive Processes in Manual Assembly. In: Proc. of the 41st CIRP Conference on Manufacturing Systems (2008)
9. Tresiman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology 12, 97–136 (1980)
10. Schubö, A., Wykowska, A., Müller, H.J.: Detecting pop-out targets in contexts of varying homogeneity. Investigating homogeneity coding with event-related brain potentials. Brain Research 1138, 136–147 (2007)
11. Crawford, T.J., Müller, H.J.: Spatial and temporal effects of spatial attention on human saccadic eye movements. Vision Research 32(2), 293–304 (1992)
12. Fitts, P.M.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
13. Holzinger, A., Höller, M., Schedlbauer, M., Urlesberger, B.: An Investigation of Finger versus Stylus Input in Medical Scenarios. In: Luzar-Stiffler, V., Dobric, V.H., Bekic, Z. (eds.) ITI 2008: 30th International Conference on Information Technology Interfaces, June 23-26, 2008, pp. 433–438. IEEE, Los Alamitos (2008)

14. Schedlbauer, M.J., Pastel, R.L., Heines, J.M.: Effect of Posture on Target Acquisition with a Trackball and Touch Screen. In: 28th International Conference on Information Technology Interfaces (ITI), pp. 257–262. IEEE, Cavtat (2006)
15. Guiard, Y.: Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. Journal of Motor Behavior 19, 486–517 (1987)
16. MacKenzie, I.S.: Motor behaviour models for human-computer interaction. In: Carroll, J.M. (ed.) HCI models, theories, and frameworks: Toward a multidisciplinary science, pp. 27–54. Morgan Kaufmann, San Francisco (2003)
17. Wickens, C.D.: Multiple resources and performance prediction. Theoretical issues in ergonomics science 3(2), 159–177 (2002)
18. Johnson, A., Proctor, R.W.: Attention: Theory and practice. Sage Publications Inc., Thousand Oaks (2004)
19. Wiedenmaier, S., Oehme, O., Schmidt, L., Luczak, H.: Augmented Reality (AR) for Assembly Processes: Design and Experimental Evaluation. International Journal of Human-Computer Interaction 16(3), 497–514 (2003)
20. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
21. Curtis, D., Mizell, D., Gruenbaum, P., Janin, A.: Several devils in the detail: Making an AR app work in the aeroplane factory. In: Proceedings of the 1st IEEE International Workshop on Augmented Reality, Los Alamitos, CA, pp. 1–8 (1998)
22. Tang, A., Owen, C., Biocca, F., Mou, W.: Comparative Effectiveness of Augmented Reality in Object Assembly. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Ft. Lauderdale, FL, USA, pp. 73–80 (2003)
23. Rosenbaum, D.A.: Human Motor Control. Academic Press, Toronto (1991)
24. Polhemus, http://www.polhemus.com
25. Stork, S., Stößel, C., Müller, H.J., Wiesbeck, M., Zäh, M.F., Schubö, A.: A Neuroergonomic Approach for the Investigation of Cognitive Processes in Interactive Assembly Environments. In: Proc. of 16th IEEE International Conference on Robot and Human Interactive Communication (ROMAN 2007), pp. 750–755 (2007)
26. Wallhoff, F., Ablasmeier, M., Bannat, A., Buchta, S., Rauschert, A., Rigoll, G., Wiesbeck, M.: Adaptive human-machine interfaces in cognitive production environments. In: Proceedings of the IEEE International Conference on Multimedia and Expo. (2007)
27. Wiesbeck, M., Zaeh, M.F.: A model for adaptively generating assembly instructions using state-based graphs. In: Mitsuishi, M., Ueda, K., Kimura, F. (eds.) The 41st CIRP Conference on Manufacturing Systems, pp. 195–198. Springer, Tokyo (2008)

Usability Metrics of Time and Stress - Biological Enhanced Performance Test of a University Wide Learning Management System

Christian Stickel[1], Alexei Scerbakov[2], Thomas Kaufmann[3], and Martin Ebner[4]

Zentraler Informatik Dienst / Vernetztes Lernen
Graz University of Technology
Steyrergasse 30/I,
A-8010 Graz
stickel@tugraz.at, alexei@sbox.tugraz.at,
kaufma@sbox.tugraz.at, martin.ebner@tugraz.at

Abstract. This paper describes the modification and outcome of a performance test applied to a university wide learning management system under realistic conditions to identify usability problems and to compare measures such as success rate, task time and user satisfaction with requirements. Two user groups with 20 test users each took part in this study. During the whole test psycho-physiological parameters of the test persons were monitored and recorded, in order to find event related stress symptoms. Modifications of the original test allowed a faster analysis of relevant quantitative metrics and the collection of qualitative information.

Keywords: Usability Test, Performance Measurement, Self-Assessment, EEG.

1 Introduction

In a world where information systems are designed to support particular transactions and workflows, a primary concern should be the effect on the efficiency of the work they support. This paper describes a modified performance test applied to a university wide learning management system developed at the IICM of Graz University of Technology (Ebner & Walder 2007). Actually this system hosts about 200 lectures and serves 10000+ users, thus the requirement for efficient workflows and productivity is given, especially as the system has to support short and longtime routines, as well as enhance learning and teaching behaviors. Targets are only meaningful when they can be expressed quantitatively, so this study measured the efficiency, respectively productivity of existing workflows and practices. The applied performance test is mostly based on the NPL Performance Measurement Method proposed by Rengger et al. whereby the modification was a small two-question survey after every task. The test persons gave thereby statements on the subjective difficulty of the tasks and their emotional state. Surprisingly a correlation was found between the hard to extract performance data and the easy to analyze survey data, which led to some ideas for economic web based performance tests. The approach included further the combination

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 173–184, 2008.

with a light Thinking Aloud Method and additional psycho physiological measures in order to research in the enhancement of UE methods with biological data. This enabled the collection of quantitative performance data on the one hand and qualitative feedback on the other hand, thereby balancing the advantages and disadvantages of both methods.

2 Motivation

The primary intention for conducting this study concerns the improvement of the tested software. TeachCenter is a learning management system that combines course management, digital content distribution and interactivity between students and teachers. The software is actually running a huge amount of lectures and is subject of continuous improvement and trend setting features. Because it is a daily used system with 10000+ users, lecturers and students it seems obviously that improvement of usability is of highest interest. Graz University of Technology had a high expertise in usability testing and learning objects (Holzinger & Ebner, 2003; Holzinger, 2004; Holzinger et al., 2005).

3 Methods and Design

The NPL Performance Measurement Method was used, in order to derive the effectiveness and efficiency of some main tasks, respectively workflows. Further it was of interest, which other issues the persons would find and what would be the underlying reasons. Therefore a combination of two Methods, the NPL Performance Test and the Thinking Aloud Test was designed, in order to provide performance data on the one hand and deeper insight on user interaction processes on the other hand. Further motivation came from the hypothesis, that the system has a good learnability. Therefore a biological rapid usability approach (Stickel, Holzinger & Fink, 2007) was applied whereby the psycho physiological parameters EEG, SCL and HR were recorded. As cross check to the above-mentioned goals, the original performance test has been further enhanced by a user self-assessment of the difficulty and the arousal after each task.

3.1 NPL Performance Measurement Method

The NPL Performance Measurement Method (Rengger et al., 1993) focuses on the quality and degree of work goal achievement. It is a rigorous usability evaluation of a working system under realistic conditions to identify major usability problems and areas. The test person fulfills tasks, whereby the time and video are recorded. As this method depends on realistic conditions, participants are not allowed to talk with the facilitator, instead they are asked to accomplish the tasks as fast as possible. Measures of core indicators of usability can be obtained, as defined in ISO 9241-11 (ISO, 1998) e.g. user effectiveness, efficiency and satisfaction. It's then possible to compare these measures with requirements. These measures are directly related to productivity and business goals. In this study the metrics Task Effectiveness, User Efficiency, Relative User Efficiency and User Satisfaction were derived. *Task Effectiveness* (TES) determines how correctly and completely the goals have been achieved in the context of the task. In most cases there's more than one way to accomplish a task and every task has several steps, as it's not meaningful to test single click actions - instead the use of the systems main functions is compiled in a task. TES is a function of quantity and quality of the task. Quantity is measured objectively as the percentage of the control parameters, which have been altered from their default values by the end of the task. Quality consists of the definition of an optimal path, with weighted alternatives and penalty actions (e.g. help or explorative search). Quantity and Quality are measured as percentage values, so the resulting TES is also a percentage value. The value of TES is obtained by measuring quantity and quality and application of the formula TES = 1/100 (Quantity x Quality). *User Efficiency* (UE) relates effectiveness to costs in terms of time, e.g. if a task can be completed in a high quality AND fast, then the efficiency is high. UE provides here the absolute measure for the comparison of the five tasks of this study, carried out by the same users, on the same product in the same environment. It is calculated as the ratio between the effectiveness in carrying out the task and the time it takes to complete the task using UE = Task Effectiveness / Task Time. The *Relative User Efficiency* (RUE) is a metric that can be employed by the relation of a particular group of users compared to fully trained and experienced user of the product being tested. It is defined as the ratio of the efficiency of any user and the efficiency of an expert user in the same context RUE = (User TES / Expert TES) * (Expert Task Time / User Task Time) *100. The User satisfaction is derived with a standardized questionnaire like SUMI or SUS.

3.2 User Self Assessment Web Questionnaires

One question of this study was how the subjective user assessment of single tasks compares to the objective performance measurement. Therefore a web interface was developed, which was present for the participants on a second screen. At the beginning of every task a “Start” button had to be pressed and when the task was finished a “Stop” button. In this way the duration of the tasks were recorded. After pressing “Stop” the user had to rate the difficulty and his arousal, each on a five point scale. After rating the “Start” button appeared again, ready for the next task. All results were logged on a web server, with a unique id for every participant.

3.3 Thinking Aloud Method

The Thinking Aloud (TA) method reveals hidden thoughts and gives an insight in the users mental model. It helps understanding how the user wants to use the system and what kind of features might be optimized. Therefore the user is asked to verbalize all thoughts and actions during the test. The test is usually designed of different tasks, which represent the major application and functionality of a system. The whole procedure is recorded with a video camera and the session is transcribed afterwards. The generated protocol can be analyzed in order to reveal information about the users reasoning sequences and goal structures. This study used the thinking aloud technique in a block right after the main tasks. Thereby questions were asked that aimed to determine how the system works. All of the so-called "Minitasks" could be easily answered by analog transfer from the recently used functions. The intention of this approach was the compensation of the main disadvantage of the NPL Performance Measurement Method, which will not reveal the reason for problems.

3.4 Psycho-Physiological Measurement Methods

Biological measures of emotional states have been used by several researchers in Human-Computer Interaction (Picard, 1997), (Picard, 2000), (Riseberg et al., 1998), (Murgg & Nischelwitzer, 2004). Muter et al. (1993) found that psycho-physiological measures can be regarded useful for Usability; especially the Skin Conductance Level (SCL) seems to be a good indicator for the overall usability of software, as they found a correlation between user-hostile systems and an increase of SCL. In this study a combination of Electroencephalogram (EEG), SCL and the Heart Rate (HR) was applied.

4 Design of the Study

This chapter covers the design of the study, thus the schematic design, the user profiling and recruitment, the setup and the standardized procedure are described.

4.1 Schematic Design

Every trial of the test was split in three main parts, the control condition (K1), the major tasks performance test (L1) and the mini tasks TA test (L2). Figure 1 on the next page shows the schematic design of the test. In the beginning all test persons were asked to fill out a profile and at the end of the test a feedback form for user satisfaction. For user profiling the scheme from the Performance Measurement Handbook was used. The German version of the System Usability Scale (SUS) questionnaire from Brooke (1996) was used to derive User Satisfaction afterwards. During the blocks K1, L1 and L2 the psycho physiological parameters EEG, SCL and HR were recorded. In L1 and L2 additional videos and screen recordings have been done.

4.2 Test Procedure

In the control condition the test persons were asked to relax in order to get some basic biodata. The relaxation process was supported by a Brainlight system (http://www.brainlight.com). The second block was the actual performance test. First the users were given a sheet of paper with the task. When they understood the task they pushed the "Start" button, otherwise they were allowed to ask the moderator comprehension questions. Each task had a goal and sub goals. When either the user thought that the complete task had been accomplished, or the moderator finished the task due to completion or time out, the user had to push the "Stop" button and was instantly asked in the web interface, to rate the difficulty of the tasks and his state of arousal. This self-assessment took place after each major task in the L1 block.

In the "Mini Tasks" L2 block, users were asked questions and had 30 seconds each to accomplish the task respectively answering the question and showing how to do. According to the TA Method the users were thereby asked to verbalize their thoughts and actions. As last step User Satisfaction was derived with the System Usability

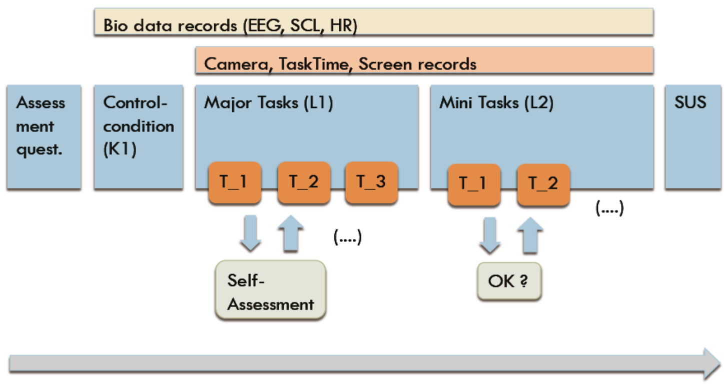


Fig. 1. Schematic design of the test

Scale (SUS), a standardized psychometric questionnaire for a high-level subjective assessment of Usability. The outcome is a rate on a scale of 0-100. Benefit and drawback is the general nature of the questions, however this allows the comparison of diverse systems (Brooke, 1996) and different parts of the system.

4.3 User Profiles

The profiling as described in the NPL PM handbook (Rengger et al., 1993) was applied. Thereby two major user groups were derived, which are students and lecturers. Target for both groups was getting novices, who had no or just little experience using the learning management system. The students had an average experience of 2 month, while the teachers had an average experience of 17 month, however the variance inside the teachers group concerning this variable was high. Beside this the profile contained several usability relevant issues on skills, trainings, mental-, physical- and job attributes. For the two user groups profile questionnaires were generated, which were then completed by every participant.

4.4 Hardware and Software Setup

Figure 2 shows the setup for the students group (left) and the teachers group (right). The test user (T) is sitting in front of two screens (M1, M2), on his left side is the moderator (M) Right beside the screens a mirror was positioned in order to capture facial expressions and the actions on the main screen (M1). The scribe (S) was operating the camera (C), Laptop L2 and taking notes. The operator (O) took care of the psycho-physiological recordings on laptop L1 and took notes too. The user test environment was a Shuttle PC with Windows XP, standard mouse and keyboard. As the learning management system is an online product Internet Explorer 6 was used as browser. Techsmith Camtasia Studio (http://ww.techsmith.com/camtasia.asp) was used for the screen recordings. A further laptop (L2) with the same software configuration was used to simulate online users. Video recordings were done with a HD Cam on a tripod right behind the user, capturing his actions on the screen, as well as the facial expressions in a mirror beside the screen.

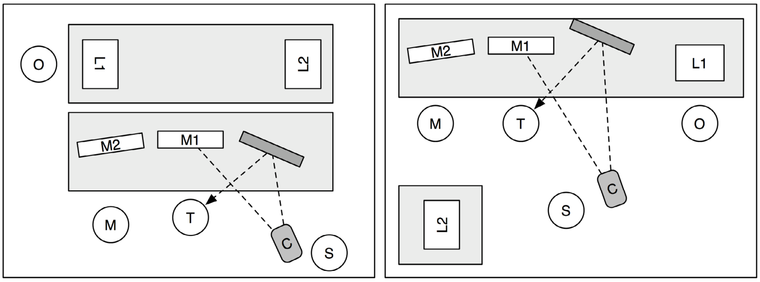


Fig. 2. Test settings for students (left) and teachers (right)

A Brainlight system (http://www.brainlight.com) was used as stimulation unit to induce relaxation by Steady State Visual Evoked Potentials (SSVEP) in a frequency range, changing between 8 and 12 Hz for 10 minutes. This system is a standalone product, which was programmed separately. It's based on SSVEPs, whereby a flickering light source elicits potentials of the same frequency in the brain, while the subject shifts the gaze to these stimuli (Müller-Putz et al, 2005). It works visually on the same base as the auditory frequency following-response, which states that most periodic complex sounds evoke low pitches associated with their fundamental frequency, also called periodicity pitch (deBoer, 1976; Evans, 1978; Moore, 1989). As different frequencies are linked to different mental respectively physiological states, relaxation can be induced using according frequencies. The EEG recordings were done with an IBVA 3 electrode headband EEG from Psychiclabs Inc. (http://www.psychiclabs.net) at a rate of 512 Hz. The equipment can be used in laboratory and field settings as well. As IBVA's headband uses only 3 electrodes, it can record an EEG of the frontal lobe only. A Lightstone from the wild divine project (http://www.wilddivine.com) was used for SCL and HR recordings. It acquires the data non-invasive by sensors, which are put on the fingertips. Unfortunately this might interfere with tasks that depend on extensive keyboard input.

5 Results and Discussion

The metrics Task Effectiveness (TES), User Efficiency (UE) and Relative User Efficiency (RUE) were derived for both groups to identify the problematic tasks. TES determines how correctly and completely the goals have been achieved, while UE relates effectiveness to costs in terms of time. The Relative User Efficiency (RUE) is the ratio between the efficiency of a user and an expert. First the RUE was calculated for every user, and then an average for all users was calculated per task. Most participants were new to the system, so the driving question for deriving this metric was the gap between experts and novices. TES, UE and RUE metrics are measured on a percentage scale, with 1 for the lowest and 100 for the highest value.

Figure 3 shows TES, UE and RUE for each task of the student trial. Tasks with low values determine usability problems. The results of the student trial show extremely decreased TES and UE values in the second task, which leads to the conclusion that

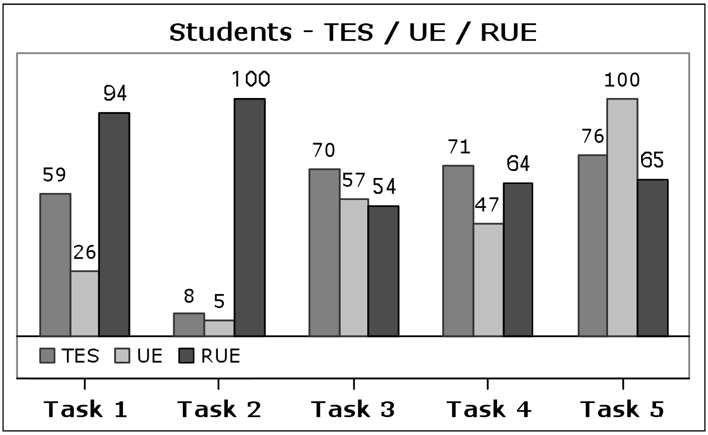


Fig. 3. TES, UE and RUE of the student trial

the students had the most problems with the second task. The goal of this task was gathering a set of information that was spread throughout the whole course, ranging from download files to discussion board threads. The use of the search feature was mandatory here in order to find all information concerning the specific topic. However, most test persons used explorative search, found one or two information and thought the task being finished. Actually the performance of the search window should have been tested; instead it was surprising to find that the search feature was hardly used.

This led to the conclusion to make the search more visible in terms of generalizabilty and place a search field with an according button in the page header. Figure 3 shows also a high RUE value for Task 2, which should usually be low. This is because the expert user wasn't able to solve the task, while some of the novice user accomplished the task.

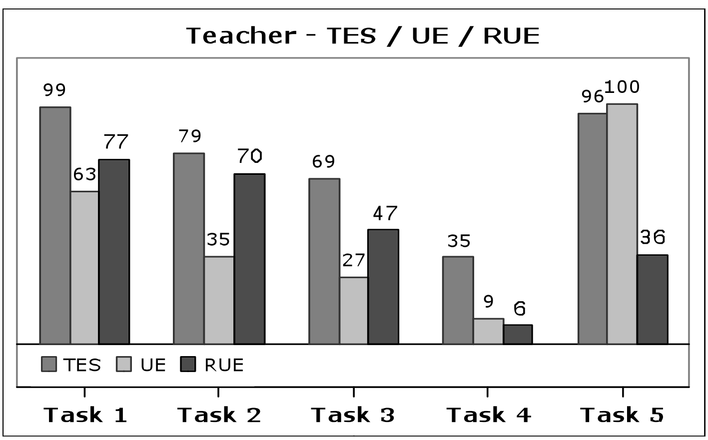


Fig. 4. TES, UE and RUE of the teacher trial

Figure 4 shows TES, UE and RUE for each task of the teacher trial. It can be seen that all values decreased in the fourth task so obviously this task contained an issue. Task 4 concerned the import of a description. The bottleneck in the workflow was the import button, which is just an icon with an upward arrow and not obvious to find. The fact that the screen recording software averted the tooltips made this task even harder. The second important issue in this case was a choice, where it was not obvious that the system expected further input. Overriding this choice cancelled the whole import process. Two changes were proposed in order to solve these issues. First the replacement of the import icon by an "IMPORT" button; second a clear message from the system that it expects another choice of the user.

Figure 4 shows also the strong decrease of RUE in Task 4 of the teacher trial as expected. This is because the expert user was able to solve the task much faster and with a higher quality. All other tasks show high values for RUE, even though they are averaged, this can be interpreted positive, because it mans that the average teacher user with one-year experience is able to accomplish main tasks with 80-90% efficiency of an expert user. Overall can be stated that the system has a very good learnablity. Simple functions are intuitive, while advanced functionality requires more training.

From Figure 3 and Figure 4 the question arose, why the Task Effectiveness was always higher than the User Efficiency. The results can be interpreted that in most cases the users of both groups were able to solve the tasks, however in a lower quality and taking much more time as expected and possible.

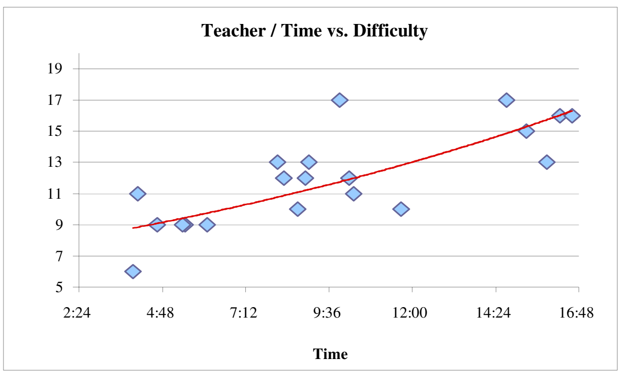


Fig. 5. Teacher assessed the difficulty according to the task time

Figure 5 shows that there's a positive correlation between task time and self assessed task difficulty, within the teachers group. The longer the tasks took the more difficult they were rated. Within the students group this can also be pointed out, but not significant. We suppose that it didn't occur within the students group, because the students have been faster, due to simpler tasks in the front end.

Figure 6 shows a positive correlation between the self-assessed task arousal and the task difficulty was found in the students group. The more difficult the tasks were rated the higher was the arousal. This was not found in the teachers group, although the distribution of the difficulty was similar. The positive correlation between difficulty and arousal in the students group could be interpreted as stress. We suppose students are more sensitive to test related stress than teachers.

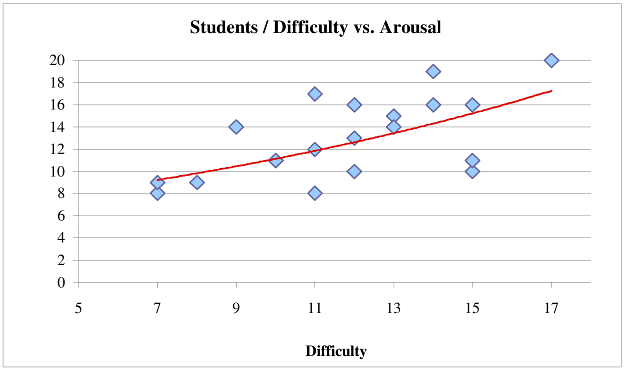


Fig. 6. Students were more exited on higher difficulty

One of the most important questions concerning the modification of the test was if there is a negative correlation between the subjective task difficulty assessment and the performance data. We hypothesized that the task difficulty should be high for low task performance. In order to visualize this, average performance and task difficulty data were normalized. Additionally the difficulty was inverted to show a positive correlating curve, which can be seen in figure 7 for the students group and in figure 8 for the teachers group. The curves for TES/UE and the subjective inversed difficulty (SID) correlate, so does also the curve for the subjective Arousal assessment per task, which was spared in the charts.

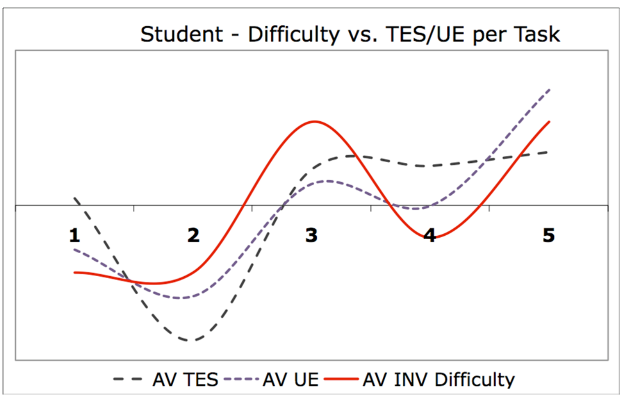


Fig. 7. Students SID vs. TES/UE per task

It is important to note that, as these curves correlate, they provide the same information on the usability of the system, although obtained in a different way. TES/UE values have to be generated by a standardized extensive procedure, while the SID is just the result of a simple question after fulfillment of a task. This leads to interesting web testing paradigms, which may provide similar conclusions as extensive testing procedures in a faster and more economic way.

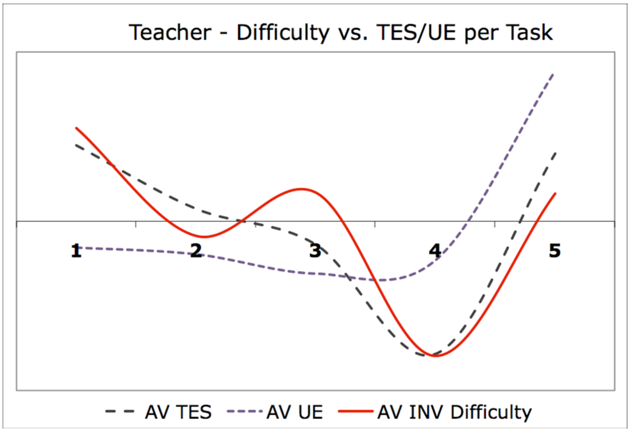


Fig. 8. Teacher SID vs. TES/UE per task

The System Usability Scale (SUS) rating for the system was equal for both groups (Students: 70, Teacher: 66), although they had different tasks, according to their role. The average difficulty of the self-assessment after each of the main tasks was similar (Students: 11.8, Teacher: 12 on a 25 point scale, whereby 25 is the highest difficulty). The SUS rating of the system can be considered as "usable". As there were two different groups with different tasks, it provides the feedback that most parts of system are satisfying and usable. Figure 9 shows a Gaussian distribution for user satisfaction is around 70 - 80 %.

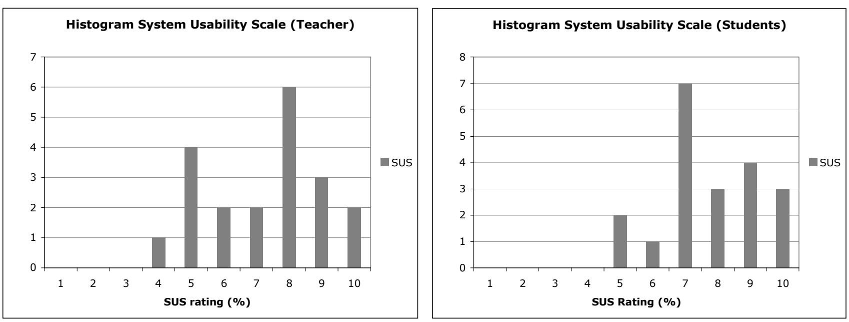


Fig. 9. Histograms of the SUS questionnaire for both groups

5.1 Further Observations

SCL peaks occurred when the user was confronted with problems, e.g. problems with uploading, login problems, explorative problems (can't find). Further peaks were noted when the users read the task and tried to understand what they had to do. With some users there were also peaks for every question in the Mini Task block. The SCL during working condition in L1 and L2 was relatively doubled compared to the SCL recorded in the relaxation condition K1.

5.2 Conclusion

The modifications of the original NPL Performance Measurement Method provided the performance metrics and additional qualitative data on reasons of problems. The self-assessment after each task gave insight into task specific user perception of difficulty and arousal. Furthermore, it was shown that the normalized performance and subjective inversed difficulty data correlate positive. So far we suppose that a simplified, more economic version of the NPL PM Method, can also provide data on the performance of a system. A further interesting paradigm can be the application of this procedure as automated web survey. As the subjective arousal assessment also correlated with the performance data, it will be interesting to analyze how this data relates to the recorded biological data and reveal further objectives and approaches to biological rapid usability testing.

References

1. Brooke, J.: SUS: a "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, A.L. (eds.) Usability Evaluation in Industry. Taylor and Francis, London (1996)
2. de Boer, E.: On the residue and auditory pitch perception. In: Keidel, W.D. (eds.) Handbook of Sensors Physiology, pp. 479–583. Springer, Berlin (1976)
3. Ebner, M., Walder, U.: e-Learning in Civil Engineering – Six Years of Experience at Graz University of Technology. In: Rebolj, D. (ed.) Bringing ITC Knowledge to work, Proceeding of 24th W78 Conference Maribor 2007 & 14th EG-ICE Workshop & 5th ITC@EDU Workshop, pp. 749–754 (2007)
4. Evans, E.F.: Place and time coding of frequency in the peripheral auditory system:some physiological pros and cons. Audiology 17, 369–420 (1978)
5. ISO The international Organization for Standardization : Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability (ISO 9241–11) (1998)
6. Holzinger, A., Ebner, M.: Interaction and Usability of Simulations & Animations: A case study of the Flash Technology. In: Proceedings of: Interact 2003, Zurich, pp. 777–780 (2003)
7. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
8. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
9. Holzinger, A., Nischelwitzer, A., Meisenberger, M.: Mobile Phones as a Challenge for m-Learning: Examples for Mobile Interactive Learning Objects (MILOs). In: Proceedings of: Third IEEE International Conference on Pervasive Computing and Communication (PerCom 2005), Kauai Island, HI, pp. 307–311 (2005)
10. Macleod, M.: Draft of chapter. In: Jordan, P. (ed.) Usability Evaluation in Industry. Taylor and Francis. Crown publishing, London (1994)
11. Moore, B.C.J.: Introduction to the Psychology of Hearing, 3rd edn. Academic Press, London (1989)

12. Müller-Putz, G.R., Scherer, R., Brauneis, C., Pfurtscheller, G.: Steady-state visual evoked potential (SSVEP)- based communication: impact of harmonic frequency compo-nents. Journal of Neural Engineering 2(4), 123–130 (2005)
13. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
14. Picard, R.W., Healey, J.: Affective Wearables. MIT Press, Cambridge (1997)
15. Picard, R.W.: Perceptual user interfaces: affective perception. Communications of the ACM 43(3), 50–51 (2000)
16. Rengger, R., Macleod, M., Bowden, R., Drynan, A., Blayney, M.: MUSiC Performance Measurement Handbook, V2. NPL, DITC, Teddington, UK (1993)
17. Riseberg, J., Klein, J., Fernandez, R., Picard, R.W.: Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state. In: Conference on Human Factors in Computing Systems, Los Angeles, CA, pp. 227–228 (1998)
18. Murgg, E., Nischelwitzer, A.: Physiological Usability Testing: A Biological Approach to Detect and Measure Usability Problems. In: Multimedia Applications in Education Conference (MApEC) Proceedings 2004, pp. 122–127 (2004)
19. Stickel, C., Fink, J., Holzinger, A.: Enhancing Universal Access – EEG based Learnability Assessment. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 813–822. Springer, Heidelberg (2007)

Users' Experience with a Recommender System in an Open Source Standard-Based Learning Management System

Olga C. Santos and Jesus G. Boticario

aDeNu Research Group, Artificial Intelligence Dept. Computer Science School,
UNED C/Juan del Rosal, 16. Madrid 28040, Spain
{ocsantos,jgb}@dia.uned.es
http://adenu.ia.uned.es

Abstract. The paper describes a model for recommendations in learning scenarios which has been designed from empirical findings following usability and accessibility criteria. This model supports course designers in describing recommendations and presents additional information to the user to explain why the recommendation has been provided. A prototype of a recommender system based on this model has been integrated in an open source standard-based learning management system. The main goal of the recommender is to improve the learning efficiency. Examples of recommendations defined with this model are provided. Moreover, a users' experience is reported.

Keywords: Recommendations, eLearning, usability, accessibility.

1 Introduction

Recommender systems (RS) can be applied to many areas where users are to be supported in their decision-making while interacting with large information spaces. They support users in finding their way through the possibilities offered in web-based settings by pre-selecting information a user might be interested in. Recommender technology has traditionally focused on e-commerce activities to select and suggest extra potential purchase to users/consumers, trying to ease the information search and the decision process. Another area where this support is very much demanded is in the eLearning field, where it would be desirable that learners are offered the most appropriate activities and resources to achieve their individual learning goals and support their needs in the most efficient way. Traditional approaches to computer-based instruction have followed a basic strategy to support learning in terms of objectives and learning resources, along with assessments focused on the measurement of learners' performance [1].

Some works have suggested applying recommendation strategies to learning scenarios. In this context, RS should help and support both learners and tutors during the course execution [2]. Learners should be supported in the performance of the course tasks by i) avoiding blockages, ii) improving the performance of the learning process by facilitating the most appropriate course contents and learning paths adapted to the

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 185–204, 2008.

learner's needs, and iii) promoting collaboration among peers. Tutors should be supported i) in the design of ad-hoc recommendations that can be delivered to the learner in the appropriate moment, and ii) in the follow-up of the learners' work by being alerted of troublesome situations [3].

Learning scenarios share the same objective as recommenders for e-commerce applications (i.e. helping users to select the most appropriate item from a large information pool) but have some particularities that have to be taken into account [4, 5, 3]: 1) the requirements (recommendations should be pedagogically guided and not only by learners' preferences, and accessibility barriers should be overcome by considering the user preferences and device capabilities), 2) the user predisposition (learners are not so motivated to continuously provide explicit ratings for each item they access as in e-commerce systems, but in turn they are used to fill in advance information requested by the institution), and 3) the structural context (educational specifications allow to situate the learner in the course). Moreover, the approach here is of lower granularity. We do not intend to recommend a course from a list of available courses regarding the users' preferences –as done in typical recommending systems, where movies or songs are recommended to a user–, but to recommend actions to the learner while performing the activities designed for a given course.

From our experience in aLFanet project (IST-2001-33288), we came to the conclusion that eLearning scenarios should combine design and runtime adaptations to better support users in the full life cycle of the learning process [6]. Following these ideas, we have designed a model to manage recommendations at design time to support the runtime operation. In this way, modeling at design time provides the needed scaffolding to offer recommendations that take into account and dynamic support required for each user.

This model allows defining different types of recommendations, which are available actions in a learning management system (LMS) to be done by learners and tutors. To apply them, different types of conditions are defined at design time, which are later computed at runtime against the current context (user, course and device). Although this model has been designed based on the requirements of learning scenarios –no matter the pedagogy applied– the ideas presented in this model can be reused in other domains.

In this paper, first we describe the model, present how usability and accessibility criteria have been considered in the user interface and comment on the sources for the data gathering. Next we present some types of recommendations that can be provided by typical LMS and show examples of recommendations defined with that model. A prototype of the system has been integrated in an open source standard-based LMS called dotLRN, which is presented in section 4. Afterwards, we describe the results of an experience with users, where learners and tutors were given a subset of recommendations in a course run in dotLRN platform and asked for their feedback. Finally, we comment on some pedagogical aspects of our approach and end-up with some concluding remarks and future works.

2 The Recommendations Model

Works in lifelong learning scenarios show that recommendations in them should be based "on most relevant information about the individual learner and the available activity, history information about similar learners and activities (learning path), guided by educational rules and learning strategies, aimed at the acquisition of learning goals" [5]. Moreover, other works show that accessibility requirements and device capabilities have also to be taken into account [7].

To support the definition of recommendations, we have worked out a model where a set of elements have to be defined to facilitate the runtime process. This model has been defined based on empirical findings and covers the following objectives: 1) supporting the course designer in describing recommendations in learning inclusive scenarios, 2) presenting additional information to the user to explain why the recommendation has been offered, and 3) requesting explicit feedback from the user when she has shown interest in the recommendation process to improve the recommender [3]. The followed methodology includes brainstorming sessions with psycho-pedagogical experts and evaluation experiences with end-uses, as the one reported in section 5.

2.1 Elements of the Model

The elements proposed for the model are: the categories, the techniques, the origin, the explanation, the timeout restrictions and the conditions. The later relates to the context information. The interrelations are shown in Fig. 1.

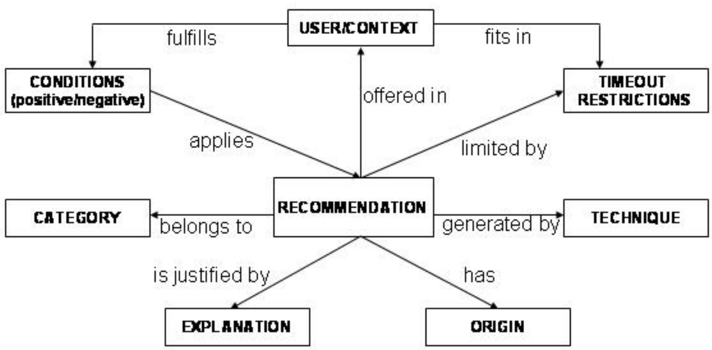


Fig. 1. A model for recommendations in learning scenarios

The Figure 1 summarizes the modeling options that characterize a recommendation. A recommendation belongs to one of the eight categories defined and can be generated by a single technique –or a combination of– techniques. At design time, the course designer selects the category to which the recommendation applies, defines the conditions and timeout restrictions and adds an explanation. The structure of the recommendation can point to the available recommendation types, which are services in the LMS. The origin and technique are dependent on the way the recommendations are generated. At runtime, the conditions and restrictions are checked against the user model and the context at hand. If applicable, the recommendation is offered to the user in an explicative, usable and accessible user interface (see below).

2.1.1 The Category

The category element classifies the scope of the recommendations offered. Our experience shows that depending on the course situation some categories (scopes) of recommendations will be more useful than others to help the user carry out her task in a more efficient, effective and satisfactory way. In particular, two states have been defined. First, the transitory state, when the user is new to the system and the course. Second, the permanent state, when the user is used to both platform and course methodology [8]. Furthermore, this information can be used to compare the performance of different RS with an ideal behavior by defining the ideal Kiviat figure for each situation. We have identified the following eight categories:

- Motivation (Mo): provides messages to motivate the learner when working in the course so she does not get frustrated if the results are lower than expected or if she requires a lot of time to carry out the given tasks.
- Learning styles (LS): suggests a way of learning the contents or the appropriate alternative content which applies best to the user preferred way of learning.
- Technical support (TS): provides hints for using the LMS functionalities or the browser.
- Previous knowledge (PK): takes into account what the user already knows, so some parts of the course are given more emphasis (if the user has not previous knowledge on them).
- Collaboration (Cl): fosters sharing contributions, communicating with course members, given the opinion on the peers work, etc.
- Interest (In): recommendations focus on those issues that the user has interest.
- Accessibility (Ac): deals with accessibility issues, such as recommending an alternative format that matches the user accessibility preferences.
- Scrutability (Sc): promotes self-reflection by telling the user what the system knows about her.

The above list of possible values for the categories is open to changes if the results from evaluations show that there are some overlaps among categories, or missing values are identified.

2.1.2 The Technique

This element refers to the recommendation technique used to generate the recommendations. According to [5], memory-based recommendation techniques are the most appropriate. Following that approach, we propose to combine the most appropriate for each situation. We consider the following techniques:

- Matching conditions (MC): selecting suitable recommendations from those defined ad-hoc by the professor in terms of conditions that apply to the current context.
- User-based collaborative filtering (UbCF): users that rated the same item similarly probably have the same taste and thus, can be recommended similar actions.
- Item-based collaborative filtering (IbCF): items rated similarly by users are probably similar, and thus, can be recommended for users who liked related items.
- Demographic collaborative filtering (DCF): users with similar attributes are matched, and similar actions recommended to them
- Ratings-Attributes mix (RAm): positive rated items by learners are recommended to similar learners.

- Case-based reasoning (CbR): if a user likes a certain item, she will probably like similar items in terms of the attributes they own.
- Attribute-based rules (AbR): item attributes are match to the preferences stored in the user profile.

In this way, social-based filtering (specially user-based and item-based techniques) are combined with information-based filtering. If the former can be applied, learners benefit from the experience of others.

2.1.3 The Origin

The origin identifies the source that originated the recommendation that has been given to the learner. We have defined the following four in order to motivate the learner to follow them:

- Preferred (Pf): it matches the user preferences.
- Popular (Pp): similar users have already found useful that recommendation.
- Tutor (Tu): the recommendation has been added by the tutor of the course.
- Course design (CD): the recommendation was specified in the course design.

Evaluations with users may result in new values for the origin element.

2.1.4 High Level Explanation

A high-level explanation of the recommendation is given for each recommendation, which explains to the user in detail why the recommendation was offered to her. This (as many of the above elements) is intended to promote trust in the user, as commented later.

2.1.5 Timeout Restrictions

Timeout restrictions to inform about the validity of the recommendation, and can be defined by an absolute or relative date, or a certain condition that can take place.

2.1.6 Applicability Conditions

Conditions for offering the recommendations define what values should take place for a user and her context at runtime to be given the corresponding recommendation. This conditions can be specified by identifying the set of values that should take place to be given the recommendation (positive conditions) and/or the set of values that should not take place to be given the recommendation (negative conditions). These values consider the user model attributes and the context.

Regarding the user information, the following values can be considered:

- A particular user (Us). Usually, recommendations are defined generic to any user that matches the required condition, but it may happen that a certain recommendation is to be given to a particular learner.
- Similar to another user (SU). Comparison on the user model information among the given user and the user using being recommended will be done.
- The role the user has in that moment (Ro). It can be a member of a community, the professor or a learner in a course.
- Learning styles of the user (LS). The recommendation may be appropriate for users with strong values in a particular learning style dimension.

- Technology level (TL). Depending on the technology level, a recommendation can be suitable for a user or not.
- Collaboration level (CL). The collaboration level is very relevant in virtual settings to assure the knowledge and experiences are interchanged among members.
- Accessibility preferences (AP). The accessibility preferences are critical in inclusive scenarios and can be considered in relation to the content of the recommendation or the interactions taken place in the course.
- Interaction data (ID). It considers the data regarding the interaction style, state and actions of the user
- Knowledge level (KL). It takes into account the knowledge level the user has regarding the learning objective (or competence) she is working in the current context.
- Interest level (IL). It takes into account the interest level the user has regarding the learning objective (or competence) she is working in the current context.

In turn, for the context information, the values are as follows:

- Platform (Pl). Recommendations are given to the learner when she is using the LMS facilities and can be applied when the user is working on a specific environment with its given resources. It can also take into account actions (i.e. read, write, create, etc.) done on objects (e.g. forum message, file, course assessment, learning object, etc.) of the LMS.
- Device capabilities (DC). Some recommendations may be or not suitable depending on the capabilities of the device being used.
- Instructional design (ID). Provides the situation the user is in the course, especially, the learning objective (or competence) being addressed in the current activity.

When designing a recommendation, not all the above properties should be filled in, only those that should be checked at runtime.

2.2 The User Interface

As in any RS, a critical issue is how to present recommendations to the user. A sensible approach is to offer a subset with the most relevant recommendations for the user, which she has the freedom to follow or not. The information shown for the recommendations and the way it is presented in the user interface can influence the attitude of the user towards the system. Providing an explanation on how the recommendation has been produced increase the user trust of the system [9].

In this section, first, we discuss the structure defined to present a recommendation on the screen. Next, we comment on the usability and accessibility criteria followed.

2.2.1 Structure of the Recommendation

The information presented to the user consists in a list of one or more recommendations. The recommendation list consists of an introductory text for the user (called 'greetings') plus a list of suggestions of actions to do by the user. Each of these suggestions (called 'recommendation') is a sentence ('content') describing the suggested action to do by the user, where a part of the sentence may be defined as a hyperlink (or more commonly called, 'link'). Since the link may most of the times be within the suggestion sentence, this (the content) will be divided into two parts, separated by a placeholder which is placed where the link should go. In more detail:

- The content: the sentence that is shown to the user and includes the link (which is placed instead of a placeholder).
- The text: the text shown on the link
- The title: the title attribute of the link
- The pointer: the URI that opens the link, it can be a URL or an object identifier from the LMS, depending on the type.
- The type: it can be internal to the LMS (if the pointer is an object identifier from the LMS) or external (if points to a URL, both from inside or outside the LMS).

The Figure 2 (in section 4) shows how the graphical user interface of the recommending system looks like for four recommendations.

2.2.2 Usability Criteria

Usability relates to the clarity with which the interactions with the RS are designed in order to make it easy and intuitive to use. Studies have shown that in order to evaluate a RS, the user satisfaction is as important as the results obtained from accuracy metrics [10]. Providing good explanations can increase the user's trust on the system's recommendations. In the learning scenario, if the user follows the recommendations (and assuming that they are appropriate) it should be quicker and easier to achieve her learning goals. Therefore, providing good explanations on why recommendations have been provided to the user can improve the user satisfaction, and thus, the quality of the system.

Adapting [11] to the particularities of learning inclusive scenarios, we have designed the user interface to cope with the following aims:

- Transparency: explaining how the system works, that is, why the recommendation have been given to the user. To achieve this, we offer a link to the user profile (see Fig. 2 in section 4) and have created an explanation page (see Fig. 3 in section 4).
- Scrutability: allowing users to tell the system that it is wrong. This functionality is available from the explanation page (see Fig. 3 in section 4).
- Trust: increasing users' confidence in the responses given. For this reason, we try to offer good explanations for the recommendations given (to allow the user understand the system behavior when the system offers a wrong recommendation).
- Effectiveness: helping users achieve the learning goals. To achieve this, recommendations are given to make the user achieve the learning goals.
- Efficiency: helping users achieve the learning goals faster or with fewer resources. In this respect, recommendations are given to make the user achieve the learning goals in a quicker and easier manner.
- Persuasiveness: convincing users to follow the recommendations and do the actions suggested. For this reason, icons are added to identify the origin of the recommendation, so the user can select the preferred one.
- Satisfaction: increasing the ease of usability when interacting with the recommender. In this sense, recommendations appear when they are relevant for the context.

For usability reasons, the number of recommendations to be provided has also been limited. The number will depend on the device capabilities and the user accessibility preferences (such as the screen size and the font size). In any case, the user is given the option to get the full list of available recommendations. Moreover, she can also access to all the recommendation that has followed.

2.2.3 Accessibility

Accessibility deals with designing user interfaces that are flexible to meet different user needs, preferences, and situations. This flexibility makes possible for people with disabilities perceive, understand, navigate, and interact with the Web, but also benefits people without disabilities in certain situations, such as people using a slow Internet connection or driving a car, people with temporary disabilities such as a broken arm, and people with changing abilities due to aging.

Regarding accessibility, the user interface is compliant with the accessibility guidelines from the W3C Web Accessibility Initiative [12]. To support those guidelines the structural information of the recommendation asks for a title description in the link to clearly identify the target of each link (checkpoint 1.3 1 from the Web Content Accessible Guidelines 1.0). Moreover, the icons presented in the user interface to make easier to understand the reasons for the recommendation are described with an alternative text (checkpoint 1.1). Header elements have also been used to convey document structure (checkpoint 3.5) and the lists structure to present the recommendation has been properly marked-up (checkpoint 3.6). The language used has also been written clearly, to facilitate the understanding by deaf from birth and cognitive disabled users.

However, the accessibility level does not depend only on the output from the RS, but also on the way the LMS presents the HTML elements.

2.3 Data Gathering

The data managed by the model is obtained from different sources and at different phases. At design time, the recommendations can be defined in terms of the category, the explanation, the origin, the applicability conditions and the time out restrictions. This information can be filled in by the course administrator via a administration graphical user interface or automatically by intelligent agents that apply some recommendation strategies [7].

At runtime, the user features of the user at hand are obtained from the user model and data from interactions (active and passive) can be obtained from the tracker component of the LMS.

3 Recommendations

Recommendations refer to the different actions that can be recommended, which are available functionalities in the LMS. Currently the system is configured to provide certain types of recommendations (but it can be easily extended for new ones, provided that the proper information is defined). First, we present the type of recommendations available. Next, we present a table of some recommendations defined following the model.

3.1 Types of Recommendations Available

The following recommendations are configured in the system:

- Learning styles (LS): points to the learning style inventory package to compute the learning styles of the user.
- Help (Hl): points to the general help page of the platform or to any of its subsections that is relevant to the current user context (contextual help)
- Post message (PM): points to a particular message of the forum and tells the user to provide a response to it.
- Read message (RM): points to a particular message of the forum and tells the user to read it.
- Upload link (UL): points to a particular folder in the file storage and tells the user to create a link there. The link can be internal to the platform or a external URL.
- Upload file (UF): points to a particular folder in the file storage and tells the user to upload a file to it.
- Read file (RF): points to a particular file and tells the user to read it.
- Read FAQ (RQ): points to a section in the FAQ and tells the user to read it.
- Fill assessment (FA): points to a particular assessment and tells the user to fill it in.
- Post a message in a blog (PB): points to a blog and tells the user to write a post.
- Comment a blog message (CB): points to a message of a blog and tells the user to comment it.
- Participate in chat (PC): points to a chat room and tells the user to participate.
- Make rating (MR): suggest to explicitly rate some elements of the environment, (the pointer will be to the object to be rated, e.g. a file), the collaboration level of a user, the difficulty level of an activity, the relevance of an activity, etc.
- Make comment (MC): suggest to comment a platform object, including items from the instructional design (the pointer will be to the object to be rated, e.g. a file)
- Read comment (RC): suggest to read the comment done on an object by a user.
- Read external link (RE): points to an external URL and tells the user to read it
- Enter a course (EC): points to a course space and tells the user to enter.
- Do activity (DA): points to an activity and tells the user to do it.
- Read resource (RR): points to a resource and tells the user to read it.
- Work on objective (WO): suggests focusing the work on a particular learning objective (or competence), whose description is given in the link.
- Online classmates (OC): shows the users from the same course currently connected and recommends a synchronous communication.
- See user model (SU): suggests seeing information inferred about the user from her interactions in the system.
- See user statistics (SS): suggest to see the statistics of the user in the system (or another user –learner– in the case of the professor)
- Alert deadlines (AD): reminds a close deadline and points to the corresponding instructions.
- Accessibility features (AF): provides advice on using the accessibility features of the browser and the platform.
- Follow-up user (FU): suggests following the contributions done by one of the learners.

- Plain text (PT): shows plain text to the user, there is no link. Mainly to be used to offer motivational messages.

New types of recommendations can be easily added to the system, provided that the corresponding functionality is available in the LMS. In order to manage this, we have proposed an administration user interface that is currently being implemented.

3.2 Instances of Recommendations

To validate the model, we have compiled instances of recommendations from instructional design experts and mapped them into the model proposed. Some of the samples gathered are mapped in the following table. The technique used and the origin will be assigned at runtime, when the recommendations are produced.

Table 1. Description of recommendations according to the model (acronyms are defined in the corresponding sections)

Recommendations	Type	Cat.	Conditions (*)		TO restrictions	Orig.
			User	Context		
if Learner.inPlatform_tool=X & Learner.technology_level=low → link to help.toolX	Hl	TS	+ID(low number of sessions), +TL(low)	+Pl(action on tool X)	Until clicked OR up to 5 sessions	Pp
if Learner.averageTimePerSession = low → "Sure you have sometime today to stay a bit longer in course!"	PT	Mo	+ID(low average time per session)	+ID(activity.difficulty Level=high)	Average time per session increases 20%	Tu
if Learner.LearningStyle=global → Show first the contents table	RR	LS	+LS(global)	+Pl(course space) +ID(course)	Table content seen	CD
if Learner.knowledge_level=low → Give additional material	RR	PK	+KL(low in objective)	+Pl(course space) +ID(course)	Resources accessed OR knowledge_level increased	CD
if Learner.PreferredFormat=auditive → course presentation in audio	RR	Ac	+AF(format auditive)	+Pl(course space) +ID(resource has alternative)	Resource accessed	Pf
if LearnerA.user_model ~ LearnerB.user_model → tell LearnerA to evaluate LearnerB.contributions	MR	Cl	+SU (user B)		Ratings done by learner A	Pp
if LearnerA.rating.inObject1=0 & LearnerB.rating.inObject1=5 → tell Learner.A & LeanerB to justify rating in thread	PM	Cl		+UserA.Pl(low rating in object1) +UserA.ID(courseA) AND +UserB.Pl(high rating in object1) +UserB.ID(courseA)	Message posted	Pp
if Learner.interest.objectiveA=high→ give items on objectiveA	WO	In	+IL(high in objectiveA)	+Pl (course space) +ID(courseA)	Next activity available	Pf
if Learner.participation=high → Show user model	SU	Sc	+ID(high participation)	-Pl(user model accessed)	User model seen	Pp
if Tutor inCourse.A → follow-up learners low interactive level in courseA	FU	TS	+Ro(tutor)	+Pl (course space) +ID(courseA)	Clicked on learner info	CD

(*) A "+" means a positive condition on that feature; a "-" means a negative condition on that feature (see section 2.1.6).

For instance, the first recommendation from the table will show the help page of the forum tool to the current user in the LMS if she has used the system less than five times, her technology level is low and she has not read the forum help yet.

4 Integration in dotLRN Learning Management System

Current developments have focused on providing the infrastructure for the RS to allow offering recommendations in the LMS user interface. In particular, an open source infrastructure for open standard-based LMS has been implemented to enrich LMS functionality with a dynamic support based on users' interactions. This initial prototype has been integrated in dotLRN [13]. The following snapshots present the user interface of the recommending system.

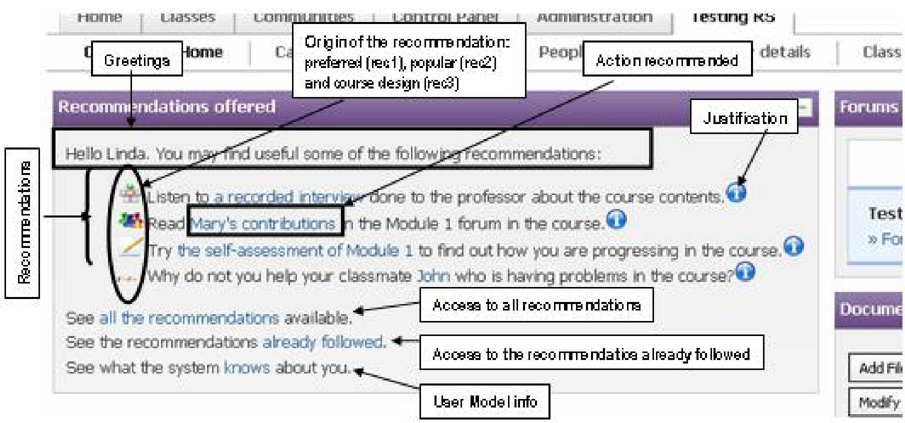


Fig. 2. Snapshot of a recommendation presented in the LMS

The icons from Fig. 2 represent the origin of the recommendation and are used to entice the user to follow the recommendations. Thus, depending on her profile and experience, she may be more confident to follow recommendations either useful for similar learners or those provided by the tutor.

The page with the explanation of the recommendations (Fig. 3) intends to get explicit feedback from the user (so the users can tell the system if it is wrong). If the user has clicked for getting more details on the recommendation, we can assume that she can find reasonable to be asked for her opinion. Therefore, she is presented with two links to close the page, one for the case that she has found useful the recommendation, and the other one for the opposite case. There is another link to provide more detailed feedback. It was found that users are ready to provide more input to the system in order to receive more relevant recommendations [9].

To avoid to overly clutter the interface and overwhelm the user with information, the number of recommendations has been limited, but the user is given the opportunity to access all the recommendation that were followed in the past as well as the

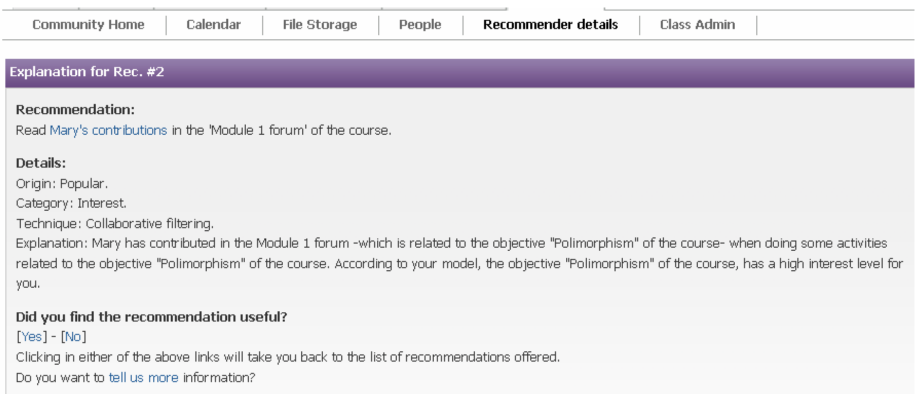


Fig. 3. Explanation for the Recommendation #2

whole list of recommendations available. Giving access to the user profile and the explanation for the recommendation offers transparency on the recommender output and system logic and therefore participate in increasing trust [9].

5 The Users' Experience

At this point, we describe the results of an experience ran during a course on Accessibility and Information and Communication Technologies, organized by aDeNu research group in July 2008 as part of the UNED offer of summer courses[1]. The focus of the experience was not put on whether the users found useful the recommendations given, but if the way recommendations were defined and presented to the users was useful and understandable for them. The set of recommendations offered was relevant to show the users how the recommending system would behave, and that this behavior would differ for each user, depending on each user's individual needs and preferences.

Fourteen users took part in the experience, although only nine of them reported valid results. One of them was visual impaired and another one was physically impaired. The other seven had no declared disability, but had worked with disabled users and were aware of their difficulties when using web-based environments.

The users were given a subset of recommendations (from those types described in section 3.1) and asked for their feedback. After an hour using the system, they were given a questionnaire. This questionnaire addresses functionality, usability and accessibility issues of the recommending system. Now we comment on the most relevant questions.

Most of the students considered that having a RS integrated in the LMS is very useful.

---

[1] Session on 'Future perspectives: towards a University fully accessible. Experiences in research projects of UNED (II)': http://apliweb.uned.es/cverano/cursos.asp?idcurso=126.

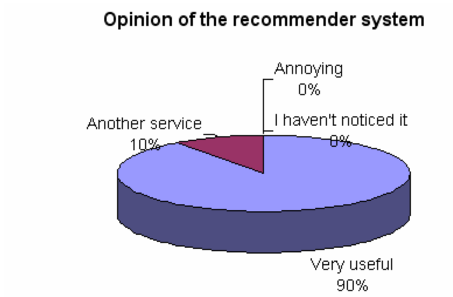


Fig. 4. Opinion of the RS by the users

Users were presented a list of recommendations (see below) to select the types they would prefer. From that list, seven types of recommendations were selected by more than half of the students: 1) "Fill in a learning styles questionnaire, so the system can be adapted to me", 2) "Read some section of the help, if there is a service in the platform that I don't know", 3) "Get alerts on deadlines to hand in an activity", 4) "Read a message in the forum that has information that may be relevant to me", 5) "Students that are on line, to set up an asynchronous connection", 6) "Advice to take advantage of the accessibility features of the platform and the browser" and 7) "Read a file uploaded by the professor or a classmate".

In turn, there were three recommendations that were selected by less than a quarter of the students: 1) "Rate some contribution done by a learner", 2) "Fill in a self-assessment questionnaire" and 3) "Access an external link of the platform".

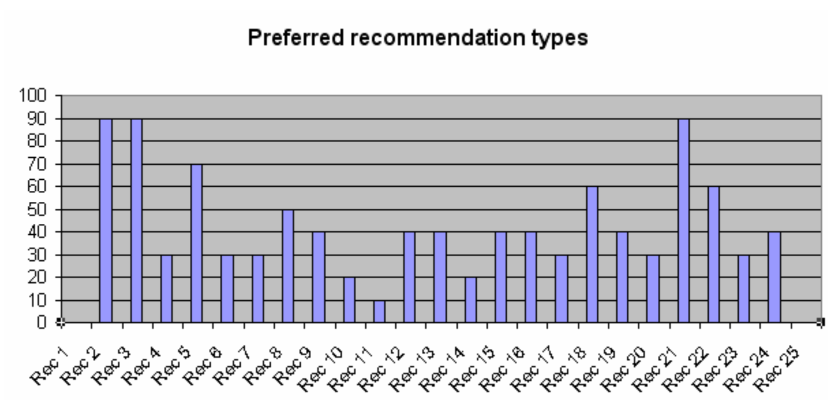


Fig. 5. Preferred recommendation types for the users

The recommendations offered in the experience were not computed by the recommending system itself, but defined ad-hoc for the experience (applying recommendation strategies to compute them is an on-going development work). However, we were interested in getting their feedback on which types of recommendations they would expect more likely, in order to guide our development efforts. The list of recommendations given to the users to rate for relevance is the following:

- Rec 1: None (I decide what to do)
- Rec 2: Fill in a learning styles questionnaire, so the system can be adapted to me
- Rec 3: Read some section of the help, if there is a service in the platform that I don't know
- Rec 4: Put a message on the forum to ask or share information about the learning objective I am working on the course
- Rec 5: Read a message in the forum that has information that may be relevant to me
- Rec 6: Upload a file with my contributions, that can be useful for others (e.g. the professor to evaluate it)
- Rec 7: Upload a link that can be useful to my classmates
- Rec 8: Read a file uploaded by the professor or a classmate
- Rec 9: Read a section of the FAQ
- Rec 10: Fill in a self-assessement questionnaire
- Rec 11: Rate some contribution done by a learner
- Rec 12: Make a comment to some contribution
- Rec 13: Read some comment of a classmate
- Rec 14: Access an external link of the platform
- Rec 15: Carry out some particular task from the course design
- Rec 16: Read a specific material from the course design
- Rec 17: Work on a specific learning objective of the course
- Rec 18: Students that are on line, to set up an asynchronous connection
- Rec 19: Access my user model, to see what the system knows about me
- Rec 20: See usage statistics from the platform
- Rec 21: Get alerts on deadlines to hand in an activity
- Rec 22: Advice to take advantage of the accessibility features of the platform and the browser
- Rec 23: Messages without any action (e.g. motivational messages)
- Rec 24: Follow-up of a classmate contributions
- Rec 25: Other

Results showed that the learning style information is considered critical, and they agree to fill in a learning style questionnaire in order to have the system adapted to it. Since the recommender is applied to a course, and courses are established by milestone actions that users have to accomplish, learners liked very much the idea of getting reminders of activities deadlines. Being aware of information added to the platform is also relevant for them (messages or files uploaded). They are also willing to establish synchronous contact with on-line classmates. Advice to take advantage of the accessibility features of the platform and the browser is also welcome. However, they gave less importance to evaluating other learners' contributions and being recommended to fill in assessments. Both are quite relevant to get feedback from the user to improve and evaluate the performance of the RS, since it provides very useful information, especially to check the validity of a property learnt by the system from the users' actions (which the system considers as implicit ratings). However, this has to be pedagogically driven (e.g., the learning design considers useful to let the learner assess a particular learning activity when she performs badly) since the RS follows a non-intrusive policy. Our view is that users did not like this type of recommendations thinking on being intrusively asked for feedback. Finally, accessing external links is not considered very relevant, either.

With respect to usability, a majority agreed that icons were clear. However, it has to be noticed that 20% of them had not paid attention to them.

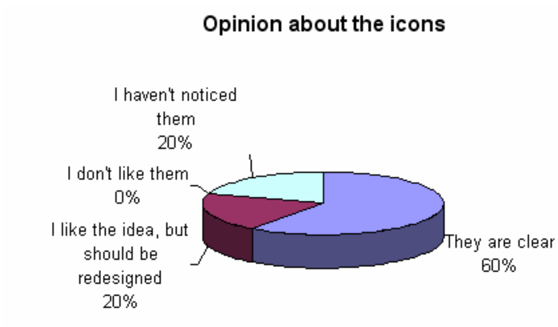


Fig. 6. Opinion about the icons in the user interface

They were also asked what kind of information they would like to receive about the recommendation. The four information sets given was found relevant for the majority of users. Especially, the high level explanation of the recommendation.

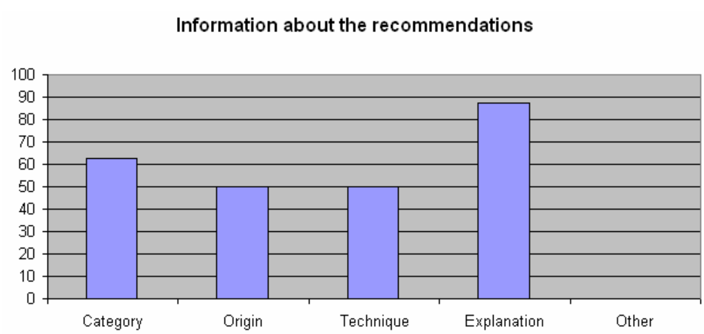


Fig. 7. Information to be given about the recommendations

Students were asked for their preferred category, among those defined. The most relevant for them was learning styles, which is consistent with the fact that being recommended to fill in the learning styles questionnaires to obtain more adapted recommendations was the most relevant for them. The collaboration category is selected as the less relevant. This is also consistent with the fact that not many of the recommendation types associated to collaboration had been highly selected.

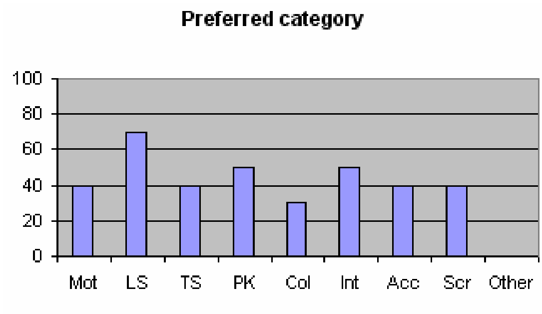


Fig. 8. Preferred categories for the students

Finally, users were asked if they had found any accessibility problem during the experiment. None of them reported any.

6 Pedagogical Considerations

In the previous sections we have presented a model to design recommendations for inclusive learning scenarios. The recommendations cover different scopes (categories) and our more than nine years in teaching on-line courses (and mainly teaching instructors how to teach on-line courses) has shown us that different support is required in the different situations of the course. Moreover, a good design done in advance – which prepares the hooks for the dynamic adaptations– improves the quality of the learning experience [6].

From a small scale experiment, we have detected that those recommendations based on the users' learning styles are highly relevant for them. Literature shows that there may be a direct relationship between learning styles and cognitive traits [14]. Based on that we could extend our recommendations model to cover cognitive traits such as working memory capacity, inductive reasoning ability and associative learning skills from the Cognitive Trait Model [15], as well as related information (field-dependency and thinking style). However, the modeling granularity and the corresponding recommendations should be manageable from the designer viewpoint.

Another issue to consider with respect to the recommendation model is how to deal with dynamic media. Recent studies show that dynamic media can support learning when limited cognitive resources, cognitive load and learners' mental representations are taken into account during the design and development of learning material. By managing recommendations, dynamic media can be tuned to the learners' experience, expertise and previous knowledge [16].

The design process for a recommendation should consider similar steps to those defined to develop a reusable, platform independent, objective based and adaptive course: 1) creation of the course material, 2) annotation with metadata, 3) define the instructional design to be applied guided by learning objectives, and 4) build the adaptive scenario that consider the runtime environment and allow delivering the adaptations to the individual learner needs [17].

Studies show that the key for designing interfaces for eLearning inclusive settings is the concept of 'ease of use', which requires the focus on the end users to address usability aspects [18]. This keeps with the methodology followed in our experiments, which is user-oriented and involves users in the very early stages of the design and development process, the so-called constructive technology assessment [19].

Cultural factors also influence the look and feel of interactive systems and every single individual develops a specific culture in terms of characteristics, behaviors, attitudes and values that affect all levels of human computer interaction (surface, functionality and interaction) [20]. If a system knows the cultural preferences of the end-user, it can adapt to them to reduce the mental workload, prevent mental distress and increase expected conformity [16].

7 Concluding Remarks

A RS in a learning inclusive scenario should provide the LMS with a set of recommendations personalized for the current user (learner or tutor) and context (i.e. the situation in the course –learning objective been worked– and the capabilities of the device used) to help the user select the most appropriate task in order to improve her learning efficiency. Thus, they require a particularized approach from those applied to e-commerce services.

We have developed a model of recommendations that considers the type of recommendations, the technique used, the origin that produced the recommendation, the category that the recommendation belongs to, the conditions that should take place at runtime to offer the recommendation, the timeout restrictions and a high-level explanation for the user justifying why the recommendation was produced.

Recommendations have been modeled according to evaluations with experts and users and are intended to facilitate the dynamic and inclusive support to learners during course execution. This model has been designed to i) support the course designer in describing recommendations in learning inclusive scenarios, ii) present additional information to the user to explain why the recommendation has been offered, and iii) request explicit feedback from the user when she has shown interest in the recommendation process to improve the recommender.

We have developed a user interface to present the recommendations in the LMS, which has been designed to be accessible, usable and explicative. In this sense, we have defined two levels of information given to the user. The first level is the information shown in Fig. 2. The user is given the list of recommendations and she can directly follow any of them. However, if the user would like to know more about the recommendation process, she is provided with more links where details are given. If she clicks in any of them, she is explicitly showing some interest on the recommendation process, and thus, it may be likely that she is receptive to give us explicit feedback on the recommendation process. For this reason, in the explanation window (Fig. 3) we explicitly ask her for feedback that is very helpful to improve the performance of the learning system.

A prototype has been tested with fourteen students from a summer course at our university. The number of users that have participated in the evaluations is small since the required developments are still on-going and not yet ready to carry out large scale experiments, so the results cannot be considered as concluding. But they provide valuable feedback that is useful for the next development phases of the system. In fact, we are applying an agile development methodology that involves users from the early stages of development, allowing changing the direction of the developments if outcomes from evaluations are not as expected. We are currently tuning the system before facing large-scale evaluations, including all the above issues identified such as the pedagogy support and graphical user interface.

8 Future Works

Our next step regarding the user interface is to embed the recommendations within the LMS services: instead of having all the recommendations clustered together within one module (or portlet), recommendations would be spread out and displayed in the most relevant locations, specially the modules that structure the learning design of the course. The objective is to offer highly contextual recommendations to the user, in a way of a contextual help, making them more obvious and relevant and therefore hopefully more efficient. Moreover, we have proposed a graphical user interface to administer the recommendations model, which is currently under development. This interface will be useful to prepare the following experimental settings since it will allow non technical instructors manage the recommendations.

The RS presented here is being integrated with the developments of ADAPTAPlan [21] and EU4ALL [22] projects. ADAPTAPlan provides a multi-agent architecture that offers adaptation based on three different user characteristics: i) competences, ii) learning styles and iii) context. EU4ALL is working on developing an open service architecture for accessible lifelong learning that accommodating the diversity of ways people interact with technology and the content and services it delivers taking into account the individual user needs. Moreover, these developments will be used to run experiences in other projects, such as ‘Accessibility for All in digital alphabetization’ (PAV-020000-2007-171) where accessible standard-based courses designed following the ALPE methodology [23] are offered to reduce the digital gap of all, especially people with disabilities and elderly people.

Another field to explore is to best adapt the RS to the new evolution of LMS, inspired from the Web2.0 trend, which leads toward more user collaboration, social interactions and user generated content. The flexibility required to address this developments is available through the use of web services technology.

In the meantime, we are also working on the prototype, to implement the recommendation techniques proposed in the model and facilitate the automation of data gathering to be able to run experiments with a larger number of users.

To increase coverage results additional evaluations will take place during the Fall term in several courses at the Computer Science School and under the program for ongoing education at our University.

Acknowledgement

Authors would like to thank the European Commission and the Spanish Government, for funding the research projects of aDeNu group, as well as the experts and users that have participated in the experience reported. Specially mentions are to be given to Carmen Barrera for her detailed comments on the recommendations modeling, Emmanuelle Raffenne for the suggestions for the design of the database model and for the implementation of the client side in OpenACS/dotLRN, Jorge Granado for the support in the development of the RS web service interfaces, Héctor Romojaro for setting up the required environments in the server, Emmanuelle Gutiérrez y Restrepo for her useful comments on the users' perception of the system and for presenting the experience to the participants of the summer course, María José Aguilar for her suggestions regarding the user interface, and Olga Revilla and Ludivine Martin for reviewing the paper. In addition, we highly appreciated the suggestions made by the anonymous reviewers.

References

1. Day, R., Payne, L.: Computer managed instruction: An alternative teaching strategy. Journal of Nursing Education 26(1), 30–36 (1987)
2. Zaine, O.R.: Building a Recommender Agent for e-Learning Systems. In: Proceedings of International. Conference on Computers in Education, Auckland, New Zealand, December 3-6 , pp. 55–59 (2002)
3. Santos, O.C., Boticario, J.G.: Modelling recommendarions for lifelong learning. In: Proceedings of the workshop on RecommenderSystems and Personalized Retrieval at ICADIWT 2008 (in press, 2008)
4. Tang, T., McCalla, G.: Smart Recommendation for an Evolving E-Learning System. In: Workshop on Technologies for Electronic Documents for Supporting Learning, International Conference on Artificial Intelligence in Education (2003)
5. Drachsler, H., Hummel, H.G.K., Koper, R.: Personal RS for learners in lifelong learning: requirements, techniques and model. International Journal of Learning Technology (in press, 2007)
6. Boticario, J.G., Santos, O.C.: An open IMS-based user modelling approach for developing adaptive learning management systems. Journal of Interactive Media in Education, JIME (September 2007)
7. Santos, O.C., Boticario, J.G.: Improving learners satisfaction in specification-based scenarios with dynamic inclusive support. In: Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), pp. 491–495 (2008)
8. Santos, O.C.: Recommending in inclusive lifelong learning scenarios: identifying and managing runtime situations. In: Proceedings of the WI-IAT Doctoral Workshop at Web Intelligence and Intelligent Agent Technogoly conference (in press, 2008)
9. Swearingen, K., Sinha, R.: Beyond Algorithms: An HCI Perspective on RS. In: ACM SIGIR 2001 Workshop on RS (2001)
10. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: How accuracy metrics have hurt RS. In: Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems (2006)

11. Tintarev, N., Masthoff, J.: A Survey of Explanations in RS. In: Uchyigit, G. (ed.) Workshop on RS and Intelligent User Interfaces associated with ICDE 2007, Istanbul, Turkey (2007)
12. W3C. Web Accessibility Initiative. Retrieved on May 23 (2008), `http://www.w3.org/WAI/`
13. Santos, O.C., Raffenne, E., Granado, J., Boticario, J.G.: Dynamic support in OpenACS/dotLRN: Technological infrastructure for providing dynamic recommendations for all in open and standard-based LMS. In: Proceedings of the International Conference and Workshops on Community based environments (2008)
14. Graf, S., Lin, T., Kinshuk, J.: The relationship between learning styles and cognitive traits - Getting additional information for improving student modelling. Computers in Human Behavior 24(2), 122–137 (2008)
15. Lin, T., Kinshuk, P.A.: Cognitive trait model – a supplement to performance based student models. In: Proceedings of international conference on computers in education, pp. 629–632 (2003)
16. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
17. Santos, O.C., Boticario, J.G.: Meaningful pedagogy via covering the entire life cycle of adaptive eLearning in terms of a pervasive use of educational standards: The aLFanet experience. In: Innovative Approaches for Learning and Knowledge Sharing, Proceedings, pp. 691–696. Springer, Berlin (2006)
18. Ebner, M., Holzinger, A., Maurer, H.: Web 2.0 Technology: Future Interfaces for Technology Enhanced Learning? In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 559–568. Springer, Heidelberg (2007)
19. Heimgärtner, R., Holzinger, A., Adams, R.: From Cultural to Individual Adaptive End-User Interfaces: Helping People with Special Needs. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 82–89. Springer, Heidelberg (2008)
20. Rip, A., Misa, T.J.: y Schot, J. Managing Technology in Society. The approach of Constructive Technology Assessment, London, Pinter (1995)
21. Baldiris, S., Santos, O.C., Huerva, D.J., Fabregat, R., Boticario, J.G.: Multidimensional Adaptations for Open Learning Management Systems. In: Proceedings of the 2008 IEEE/WIC/ACM Joint International Conference on Web Intelligence and Intelligent Agent Technology 2008, International Workshop TUMAS-A 2008: Web Intelligence & Intelligent Agent Technology in eLearning (in press, 2008)
22. Santos, O.C., Rodriguez Ascaso, A., Boticario, J.G., Martin, L.: User modeling for attending functional diversity for ALL in Higher Education. In: Weske, M., Hacid, M.-S., Godart, C. (eds.) WISE Workshops 2007. LNCS, vol. 4832, pp. 301–312. Springer, Heidelberg (2007)
23. Santos, O.C., Boticario, J.G., Fernandez del Viso, A., Perez de la Cámara, S., Rebate Sanchez, C.: Basic skills training to disabled and adult learners through an accessible e-Learning platform. In: 12th International Conference on Human-Computer Interaction, pp. 796–805 (2007)

The Role of Usability in the Design and Evaluation of Dynamic Traffic Displays

Gerhard Leitner, Martin Hitz, and Rudolf Melcher

Department of Informatics Systems, Interactive Systems Research Group
Klagenfurt University
9020Klagenfurt, Austria
{Gerhard.Leitner,Martin.Hitz,Rudolf.Melcher}@uni-klu.ac.at

Abstract. Dynamic traffic displays are continuously replacing solid traffic signs [cf. e.g. 1]. Besides their obvious benefits, dynamic traffic signs may cause many potential problems, which are to be considered in the design and implementation of new signs. A project is presented by describing the basic considerations and methods applied in the development of a dynamic traffic display which was designed with the goal to optimally present information to traffic participants and support a behavioral change. The steps performed within the development process of design alternatives are presented as well as the different usability engineering methods applied in a web survey and an exploratory study in order to evaluate the design alternatives.

Keywords: Usability engineering methods, conceptual design, lab based evaluation.

1 Introduction

Traffic signs have a strong influence on our daily life. In the last years solid signs with static information are more and more replaced by dynamic displays which present changing information related to the actual context. Altough dynamic traffic displays like VMS (variable message signs) have their obvious benefits such as higher actuality of information and flexibility they also bear a lot of potential problems. Whereas solid signs have to be designed based on clear legislative regulations the majority of which is internationally standardized, it is difficult to apply these regulations to the design of dynamic displays. The reasons of this are multidimensional, e.g. related to the constraints of display technology (brightness, contrast) or perceptual aspects (e.g. the possibility of dazzling). Because of these constraints signs on dynamic displays are, in general, implemented in inverse color mode (bright text/graphics on dark background) which differs from the typical design of solid road signs (dark text/graphics on bright background). Another potential problem could be that, depending on the context, dynamic displays provide the possibility (which is frequently exploited) to use them for other purposes than pure traffic information, e.g. for advertising upcoming events (e.g. shown in Figure 4).

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 205–220, 2008.

This could lead to the problem that traffic participants ignore signs which are too similar to advertisements. In this paper we present a project the central task of which has been to control and manage the flow of traffic within and around a city which suffers from frequent traffic jams. The project was lead by the traffic department of the provincial government. The project partner responsible for technical issues was the Austrian motorway and expressway operator ASFINAG. The technology employed was based on number plate recognition and estimation of transit time.

The work package of our research group presented in this paper was to develop the design of a dynamic display that should optimally inform drivers on the current traffic situation and motivate them to choose an alternative route over the by-pass (a divided highway around the city) rather than using the (apparently) shorter route through the city center which in fact, due to traffic congestion, was more time consuming.

What made the project interesting but also a big challenge was, that due to a tight schedule and limited budget a fast and effective application of usability engineering methods has been of crucial importance. Based on the results we gained, we think that this goal could be achieved to a certain degree.

The paper is structured as follows. After a short review on related work in different scientific disciplines in Section 2, the methods we used to reveal the best possible results for the project are described in Section 3. In Section 4, the results we gained are presented, followed by a critical discussion on the lessons learned in the project.

2 Related Work

In the beginning of the project, a research on pertinent literature in different disciplines relevant for the project has been performed. The research areas we considered relevant were cognitive psychology, semiotics, information visualization, icon design, and display technology. The main focus, however, was laid on different aspects of human-computer-interaction, especially with respect to the usability engineering lifecycle and the applicability of different usability engineering methods (cf. e.g. [2, 3]).

Traffic research has a long history, and standardization of traffic signs has been an issue for about hundred years, cf. e.g. [4]. From the classical theories and concepts in psychology (memory and attention, cf. e.g. [5, 6]) it is well known that the capabilities of human information processing are quite limited especially when several tasks have to be performed simultaneously [7]. Driving a car is a typical situation where this is necessary. The quantity and quality of information presented in traffic must therefore meet certain requirements. Another well known fact of importance for the project is the form in which information is presented. Pictorial information is processed more effectively (cf. e.g. [8]) than textual information, but pictorial information is, in general, more ambiguous than its textual equivalent.

On the basis of results from different disciplines, researchers have formulated criteria traffic signs should fulfill. According to Dewar, traffic signs have to be easily detectable, attract attention, be legible (even in short exposure or from a greater distance), and must be easily understood [9].

Each aspect mentioned could bear problems, e.g. a sign should draw attention but not distract the driver's attention from the road [10, 11]. Several studies [9, 12, 13] revealed the problem that knowledge about the design (shape and color code) and meaning of traffic signs is quite low, in some studies only 50% of signs were named and described correctly. Factors influencing the recognition of traffic signs could be e.g. related to cultural issues [14, 15] or the age of traffic participants [16, 17]. A problem that adds to the difficulty is that advertisements are intentionally designed similar to traffic signs to attract drivers attention [18].

Many interesting insights can be gained from disciplines like information visualization (cf. e.g. [19]), but the standardization of traffic signs limits the degrees of freedom in designing new signs. However, as mentioned, when designing a sign based on standards one cannot be sure that all "consumers" understand the information. A possible solution for that problem can be taken from iconography. In addition to the above mentioned criteria a traffic sign should meet, the dimensions guessibility, learnability, and experienced user performance proposed by Moyes and Jordan [20] could be adopted for the design of traffic signs.

Another important aspect is related to display technology. The device chosen for the final implementation was a LED display which provides a resolution of 2048*1536 Pixels. Therefore space for displaying information was rather limited. Thus, we also considered alternative display techniques like RSVP (rapid serial visual presentation). However, as mentioned above, drivers have limited attention and information processing capabilities and phenomena like change blindness could also occur in traffic situations [21, 22]. Additionally, De Waard et al. could show that electronic road signs are rated worse than real signs in the dimensions clearness, legibility, good/bad, and pleasantness [23].

Besides the consideration of aspects relevant for the content of the traffic sign to be designed another important issue for the project has been to select the appropriate set of methods for the user (i.e., driver) based evaluation of the designs.

Dynamic traffic signs are a kind of ubiquitous computing system the evaluation of which is difficult. Based on the categorization of Preece et al., a differentiation between quick & dirty, lab based and field based evaluations can be made [24]. The selection of the right evaluation method is a challenge especially for systems which go beyond the classical user / desktop-computer interaction, or as Streitz et al. formulate it – "systems operating more or less as an element of the fabric of the environment around us" [25]. An overview of traffic signs evaluation methods is given by Martens [26]. Methods that are frequently applied are eye-tracking, thinking aloud or retrospective recognition of signs.

3 Methods

3.1 Conceptual Design

Based on the outcome of the literature research, four dimensions have been chosen which served as a basis for categorizing design alternatives to be drafted for further evaluation. The dimensions enumerated below are a result of a concept meeting carried out by the team members and were identified as the most relevant in the project´s context.

- Emotionality:
  To what degree does the design include emotional aspects, e.g feeling guilty about ones contribution to pollution (caused by congested traffic).
- Complexity:
  How many elements or combinations of elements does the design include.
- Information Content:
  How much information is presented in the design.
- Familiarity:
  Can the design as a whole or in parts be recognized based on experience.

Whereas some of the dimensions could be operationalized quite easily, we were aware of the problems that it is quite difficult to consider a dimension like "emotionality", however, emotional aspects play an important role in road information (cf. e.g. [27]). Therefore we designed our signs on the basis of examples of emotional signs we collected in the first phase of the project. Other dimensions like attraction of attention, readability, guessability discussed above were planned to be evaluated within the user based evaluations of the project. Due to the mentioned project constraints we could not develop one concept for each cell of the resulting four dimensional grid, but only a subjectively derived reduced set of alternatives to have a reasonable (non-exhaustive) variety of design concepts. To give a realistic impression of the look of the final display we applied a LED filter on the design mock-ups. Applying this filter also helped to see whether a display with the specification mentioned above would be at all capable to display the intended designs. With many designs it turned out that we had to adapt text parameters like font size, font type, or color as well as graphics. The following examples show thumbnails of the designs (with the LED filter applied).

The first design category was classified as *"traffic sign oriented"*, characterized by a high level of familiarity, high complexity, high information content but low emotionality. Examples of the designs are shown in Figure 1. The second category of designs, shown in Figure 2, was called *"schedule oriented"* and contained information typically displayed on VMS (variable messag e signs[1]) which are frequently used e.g. on airports or on highways in various countries. The designs are characterized by a medium degree of familiarity, high complexity, low emotionality, and high information content.

The third category of signs was called *"icon oriented"* and based on knowledge drawn from iconography. The designs are characterized by medium familiarity, low complexity, low emotionality, and medium information content.

Finally, we designed a category of signs, which we called *"emotion oriented"* due to an assumed higher degree of emotionality, but without the differentiation of basic emotions used in different models [c.f. e.g. [28]). This category is divided in two sub-categories. One sub-category is based on graphical elements the inhabitants of the city should be familiar with. One element is a character (Pulcinella, originally from Comedia dell´Arte) which is used by the city for advertising purposes, especially for cultural events. The other element is the logo of the city. We assumed that these characters could be recognized and associated with relevant information. The designs intentionally had a high degree of complexity, because – as mentioned – we oriented

[1] Cf. e.g. http://www.dambach.de/fileadmin/be_user/img/els_Diwista_WM.jpg (2008-09-10).

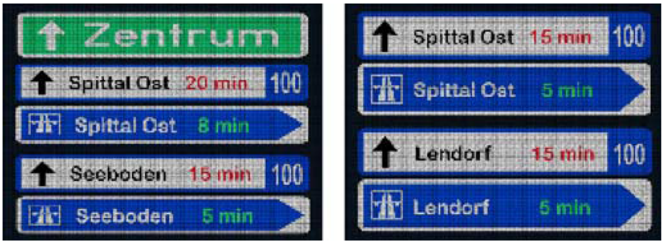


Fig. 1. Designs similar to conventional traffic signs. The design *"signpost with city center"* (*WMZ*) on the left, and *"signpost without city center" (WOZ)* on the right.[2]

Fig. 2. Design alternatives of the category *"schedule oriented"*. On the left the design with text *(VMT)*, in the middle with icons *(VMI)* and on the right with icons and signposts *(VMW).*

Fig. 3. Design alternatives of the third category "icon oriented". On the left an alternative with additional information on the saved time *("You save 15 minutes / Thank you!") (UMZ)*, in the middle an alternative with a general hint *("You are faster via the by-pass / Thank you!"), (UMT)*, on the right just the symbol with an icon and "*Thank you!" (UOT).*

our concepts on real signs we found in the preparation phase of the project (examples are shown in Figure 4).

The other sub-category was developed based on the idea to generate negative emotions which are frequently used in traffic information, as shown in Figure 6.

[2] For this example and in the following sections, acronyms consisting of capital letters represent designs developed within the project described, while acronyms with lower case letters represent designs taken from other sources. For the sake of ease of handling we kept the original German acronyms in this paper, because we used them in all the computations.

Fig. 4. Examples of multipurpose road signs with high complexity used for traffic information and community information, e.g. *city logo with text* on the left, information about the *european soccer championship* in the middle and *a reminder on immunization* on the right.

Fig. 5. Design alternatives of the category *"emotion oriented"*. On the left a design with *Pulcinella and a green car ("Faster via by-passA10") (PMG),* in the middle *Pulcinella in red ("Avoid the city!") with traffic jam in the background* (FAR) and on the right *the silhouette of the city with logo and traffic jam sign ("Ten minutes faster via by-pass A10!")(SLW).*

Fig. 6. Examples of emotion inducing information on Austrian highways. On the left a campaign against *speeding ("survive, reduce speed")*, on the right a campaign against *drinking and driving*.

Concepts like pollution, reduced life quality etc. should be induced. Designs of this category were not focusing on the personal benefit (save time) but on altruistic aspects (enhance life quality in the city by avoiding to drive through). The emotion-oriented designs are charaterized by medium familiarity, high complexity, high emotionality, and medium information content.

Fig. 7. Designs of the category “emotion-oriented”. On the left and in the middle a *buggy in a cloud of exhaust fumes*, once with additional information on transit times via the by-pass and trough the city, once only with a slogan *(“Please take the by-pass”)(VMK*). On the right the picture of an exhaust with a cloud of fumes, with the same slogan.

3.2 Online Survey

We were well aware that the signs we designed could contain a high amount of subjectivity and the transported message might– although clear for us– not be understood by other people. Therfore we conducted an online survey to eliminate designs which do not fulfill our requirements.

The design concepts were evaluated in an online survey where 132 persons participated (83 of them completed the whole questionnaire). The survey consisted of 20 items, about half of which were the designs we developed, the second half contained a sample of existing road signs and advertisements, some of which are shown in Figure 8. By using this combination we wanted to simulate a degree of cognitive load which is realistic for traffic situations and to have measures for the comparison of our designs with others.

The signs were presented to the subjects for just 10 seconds each in full screen mode. As mentioned, the survey was carried out to eliminate bad design alternatives.

Fig. 8. Selection of designs additionally used in the online survey. From top left to bottom right: *information bar (irs), closed road (snu), deviation / pedestrians (ufr), road works / deviation (bsu) and advertisement burger/soccer (wbf).*

Therefore conditions unrealistic for traffic situations, such as focused attention, no distraction, long presentation time and singular tasks were accepted.

After the presentation the sign disappeared and the subjects had to answer questions on the design, e.g. to describe the sign they have seen in their own words, what they would do if they saw this sign in reality, how they liked the design and how they would rate it on the basis of school grades.

The results of the online survey are presented in Figures 9 and 10. As can be seen in Figure 9, our design *UMZ* was rated highest, followed by two already existing signs (*irs, snu*) and a group of our designs from the categories *"traffic sign oriented"* and *"schedule oriented"*. Signs with complex pictorial content such as advertisements and our designs of the category *"emotion oriented"* were rated worse. The sign with the least rating was the real sign *(bsu, see Figure 8)* showing a deviation on a roadwork site.

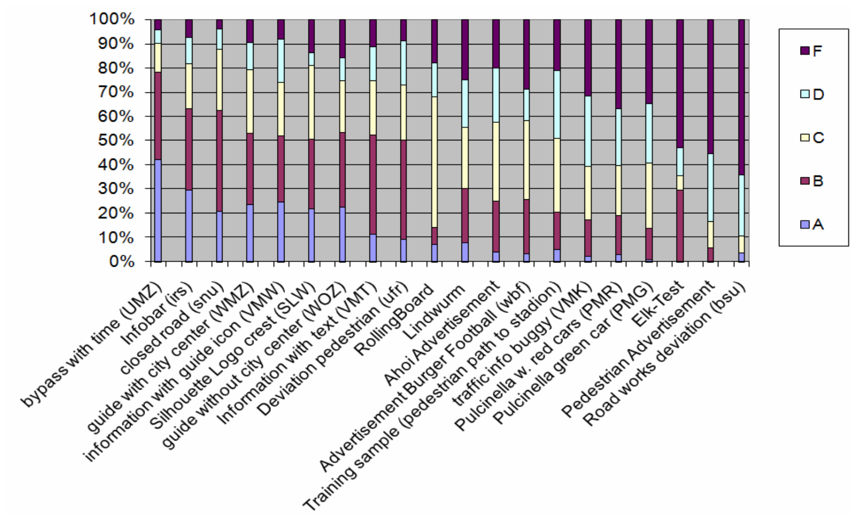


Fig. 9. Signs and scores given by the subjects of the online survey. Acronyms with capital letters *(e.g., UMZ)* represent designs developed within the project, acronyms with lower case letters *(e.g.,irs)* represent signs taken from existing examples.

Also of interest are the estimated (subjective) probabilities with which the subject would follow the messages on the signs. Here are two of the existing signs rated with the highest probability, followed again by the design *UMZ* and other designs from the categories *"traffic sign oriented"* and *"schedule oriented"*. The signs with complex graphics are again rated worse.

Based on the findings of the online survey, a representative sample of signs has been selected to be integrated into the design of a lab based evaluation.

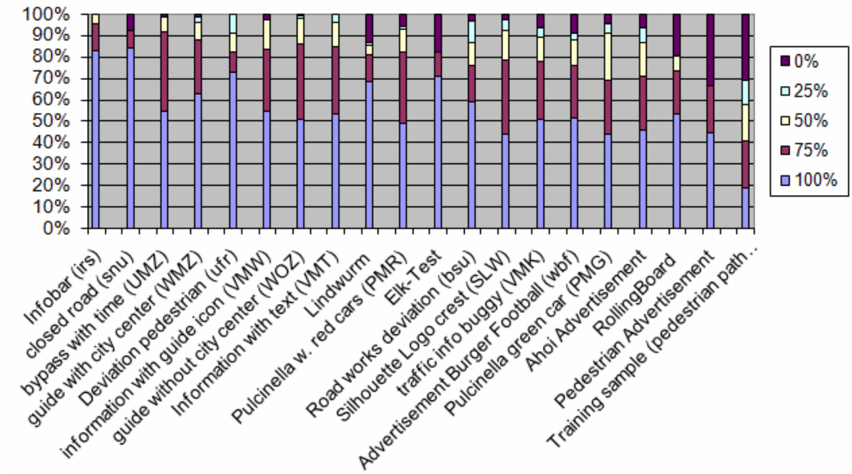


Fig. 10. Signs and subjective probabilities with which the subjects would follow the messages on the signs. Acronyms with capital letters *(e.g., UMZ)* represent designs developed within the project, acronyms with lower case letters *(e.g.,irs)* represent signs taken from existing examples.

3.3 Lab Evaluation

The major challenge for the final evaluation has been what method (field study or lab test) should be applied to get the most out of it. The first possible solution seemed to be to make digital printouts of the design alternatives on plates, install them on site and drive past them with a number of subjects (say, 12) usually used in such evaluations. The subjects could have worn mobile eye-tracking equipment and could have been observed with other equipment like video cameras.

This option has been withdrawn because of the high effort that would have been necessary and the potential danger, that other traffic participants could have been distracted by our signs which could perhaps lead to an increased probability of accidents.

The next possibility which was discussed and tested on site was based on the AR Toolkit[3]. The idea was to produce a solid wooden sign with a marker printed on it, whereas the marker is serving as a placeholder on which arbitrary information (different design alternatives) can be rendered at runtime. Therefore the marker had to be placed on locations which seemed to be suitable places for the final sign and filmed with digital video while driving. The toolkit has the capability to extract the marker from the video at runtime and to render our design alternatives on it.

The sign with the marker was crafted and tested on the campus of our university. The first results were satisfactory, however, rendering the design alternatives on the marker did not have the expected quality in relation to the involved effort.

[3] http://www.hitl.washington.edu/artoolkit/

By collecting the video material on the real locations we found that although the dimensions of the marker were 1,5m * 1,5m it was too small for the real world environment and the toolkit had difficulties in recognizing the marker from a distance.

Because of these findings we chose a classical lab based design for the evaluation with the benefit of having the full observation and recording equipment of our usability laboratory. To overcome the technical problems with the ARToolkit we switched to a video authoring software with which we could render our design alternatives on arbitrary rectangular objects into videos of locations of the city that were recorded in a preceding step.

During the test, the video sequences (3 sequences of about 8 minutes each, presented in random order) were presented to subjects. Figure 12 shows an example scene with one of our designs rendered on a road works sign.

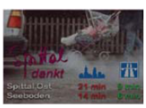
Fig. 11. Marker for ARToolkit positioned in a roundabout as preparation for video recording

Twelve persons with differences in demographic data (gender, age, driving experiences) were invited and participated in the test by sitting in front of a simulated car cockpit, however, without the possibility of interaction (e.g. influencing speed by pressing a brake or accelerator pedal). The subjects were instructed to view the following scenes as if they were in the position of the driver. The scenes were presented on a Sony 22” Wide Screen CRT.

Three different video sequences of about 8 minutes which consisted of different driving scenarios in and around the city were presented in alternating order. Some of them were recordings from the city center, containing a higher traffic density, street signs, crossings, traffic lights etc. Other scenes featured country-side or peripheral areas of the city with lower traffic and a lower density of signs. In some locations

Fig. 12. Traffic scene with the design *UOT* on the right and a speed limit end sign on the left

Table 1. Signs and corresponding answers for Video Sequence 1

Design	Present in video sequence	Correct answers	Design	Present in video sequence	Correct answers
	Yes	11		No	0
	Yes	6		No	11
	Yes	0		No	8

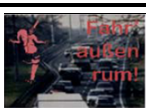
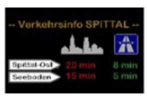
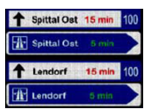
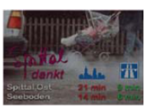
(which were identified as realistic positions for the signs) our designs were rendered onto other existing signs, advertisements or other rectangular objects in the video. The eyegaze of the subjects was recorded for further analysis.

In between the video sequences, the subjects were asked to answer questions whether or not certain signs were in the videos they saw. Moreover (as in the online survey) they were asked to describe the meaning of the signs, to rate them on the basis of school grades and whether they would follow the hints given on the signs.

Table 2. Signs and corresponding answers for Video Sequence 2

Design	Present in video sequence	Correct answers	Design	Present in video sequence	Correct answers
	Yes	11		Yes	2
	Yes	5		No	1
	Yes	4		No	11

Table 3. Signs and corresponding answers for Video Sequence 3

Design	Present in video sequence	Correct answers	Design	Present in video sequence	Correct answers
	Yes	7		No	3
	Yes	2		No	3
	Yes	0		Yes	0

The results were similar to the results gained from the online survey. Designs of the categories *"icon oriented"*, *"schedule oriented"* and *"traffic sign oriented"* were rated higher than the signs form the category *"emotion oriented"* and advertisements which were also rendered into the video sequences. However, the correctness of answers whether or not a sign was present in a sequence was not significantly different between the different categories.

4 Discussion and Conclusion

The results revealed by the different evaluation methods show a heterogeneous picture. Whereas both, the online survey and the lab test, revealed the result that the icon oriented signs are subjectively rated best, followed by schedule oriented and traffic sign oriented designs, the correct recognition of the signs is not clearly different between the different design categories. We got a lot of false positives and false negatives, the interpretation of which is quite difficult.

What could be observed, however, is that the icon oriented signs have been the clear (subjective) favorites, because they are unique and can be remembered and interpreted accurately. All other design alternatives are commonly used and therefore seem not to be suitable to fulfill two kinds of needs, i.e. supporting a long term change in the behavior of traffic members who are familiar with the situation (inhabitants of surrounding villages, commuters, or truck drivers) and, because one of the biggest economic factors of the region is tourism, to provide a sign which is fast and easy to understand even for persons with foreign mother-tongues. The icon shown in the figure below was "invented" by ourselves and seems to contain the crucial elements a sign should have (according to the criteria of [9]).

Fig. 13. Traffic scene with the design *UOT* on the right

Another observation we made was that although we invested a lot of effort in the rendering of the signs into the videos, the signs had some salient features which made them easier to identify, especially when the subjects had answered the first set of questions after the first run. Some of the subjects even developed a special behavior because they did not observe the traffic but looked on locations where they expected signs. Under real conditions, e.g. in a field study or in a realistic driving simulator this behavior would have led to critical situations. In summary, we are quite satisfied with the outcome of the project in relation to the effort of about two person months.

However, a lot of issues were identified which should be considered in future activities. The dimensions used for the development of the design alternative are not exhaustive. Further analyses are necessary. We used a pragmatic approach to evaluate the designs. However, especially within the usability test the subjects knew and felt that the situation was not real and they did not have to concentrate on really driving a car. One possible solution could be the implementation of a prototype which provides a more realistic interaction by the user (eventually based on the API of a gaming console). This solution would have the benefit of easier implementation, however, probably containing the same problem regarding realism of the scenario like in the approach described in this paper.

The other solution could be to mount equipment in a car and carry out the evaluation on a private car circuit, where all conditions can be controlled and manipulated and the danger of accidents (as discussed in Section 2) is reduced. This approach could bare the problem of much higher effort (e.g. programming and configuring a dynamic display, preparing mobile equipment). From our actual point of view the most important and most interesting activity would be to carry out a comparative study of the different approaches discussed, eventually with a baseline setting (e.g. two monitors, one with the traffic scene, one with a sign) and compare the outcome in relation to the necessary efforts.

Acknowledgment

We are grateful to Marcus Haring for his contributions to the project.

References

1. Wardman, M., Bonsall, P.W., Shires, J.D.: Driver response to variable message signs: a stated preference investigation. Transportation Research Part C: Emerging Technologies 5(6), 389–405 (1997)
2. Mayhew, D.J.: The Usability Engineering Lifecycle: a Practitioner's Handbook for User Interface Design. Morgan Kaufmann Publishers Inc., San Francisco (1999)
3. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
4. Krampen, M.: Zur Geschichte der amtlichen Verkehrszeichen. Zeitschrift für Semiotik 17, 3–8 (1995)
5. Baddeley, A.D.: Working memory. Oxford University Press, New York (1986)
6. Neumann, O.: Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen. Psychologische Rundschau 43, 83–101 (1992)

7. Engström, J., Johansson, M., Östlund, J.: Effects of visual and cognitive load in real and simulated motorway driving. Transportation Research Part F: Traffic Psychology and Behaviour 8(2), 97–120 (2005)
8. Paivio, A.: Mental Representations. Oxford University Press, New York (1986)
9. Dewar, R.E.: Verkehrsschilder: Forschungsstand und Verbesserungsmöglichkeiten [Traffic signs: state of the art and possibilities for improvement]. Zeitschrift für Semiotik 17, 35–63 (1995)
10. Wallace, B.: Driver distraction by advertising: Genuine risk or urban myth? Municipal Engineer 156, 185–190 (1995)
11. Martens, M.H., van Winsum, W.: Measuring distraction: The peripheral detection task (last access: 2008-09-10), `http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/Welcome.htm`
12. Al-Madani, H., Al-Janahi, A.: Role of drivers' personal characteristics in understanding traffic sign symbols. Department of Civil and Architectural Engineering, University of Bahrain, Accident Analysis & Prevention 34(2), 185–196 (2000)
13. Siegener, W., Träger, K., Färber, B.: Forschung Straßenbau und Straßenverkehrstechnik; 916 Wirtschaftsverlag NW (2005)
14. Ng, A.W.Y., Chan, A.H.S.: Culture Issues in Traffic Sign Usability. HCI (10), 379–387 (2007)
15. Heimgärtner, R., Holzinger, A.: Towards Cross-Cultural Adaptive Driver Navigation Systems. In: Holzinger, A., Weidmann, K.-H. (eds.) Empowering Software Quality: How can Usability Engineering reach these goals?, pp. 53–67. Austrian Computer Society (2005)
16. Kurniawan, S.H., Zaphiris, P.: Investigating the Age Effects on Subjective Assessments of Traffic Signs. In: Proceedings of the IIE Annual Conference, Dallas, TX, USA, pp. 20–23 (2001)
17. Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating Usability Metrics for the Design and Development of Applications for the Elderly. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) 11th International Conference on Computers Helping People with Special Needs. LNCS, vol. 5105, pp. 98–105. Springer, Heidelberg (2008)
18. Transports Quebec: Highway Safety CodeSection 306 Advertising creating confusion with road signs (last access: 2008-08-31), `http://www.cbsaffichage.ca/eng/creative/files/Article_306_English.pdf`
19. Card, S.K., Mackinlay, J.: The Structure of the Information Visualization Design Space. In: IEEE Symposium on Information Visualization, Phoenix, pp. 92–99 (1996)
20. Moyes, J., Jordan, P.W.: Icon Design and its Effect on Guessability, Learnability, and Experienced User Performance. In: Alty, J.L., Diaper, D., Guest, D. (eds.) Proceedings of the Eighth Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers VIII, August 7-10, pp. 49–59. Loughborough University, UK (1993)
21. Velichkovsky, B.M., Dornhoefer, S.M., Kopf, M., Helmert, J., Joos, M.: Change detection and occlusion modes in road-traffic scenarios. Transportation Research Part F: Traffic Psychology and Behaviour 5(2), 99–109 (2002)
22. McCarley, J.S., Vais, M.J., Pringle, H., Kramer, A.F., Irwin, D.E., Strayer, D.L.: Conversation Disrupts Change Detection in Complex Traffic Scenes. Human Factors 46, 424–436 (2004)
23. De Waard, D., Mesken, J., Brookhuis, K.A.: Recognition of electronic road signs. In: Brookhuis, K.A., de Waard, D., Weikert, D.M. (eds.) Human Factors in Telecommunications –Implications for the aged and disabled, pp. 37–48. Shaker Publishing, Maastricht (2000)

24. Preece, J., Rogers, Y., Sharp, H.: Interaction Design. John Wiley & Sons, New York (2002)
25. Streitz, N.A., Kameas, A., Mavrommati, I.: The Disappearing Computer, Interaction Design, System Infrastructures and Applications for Smart Environments. In: Streitz, N.A., Kameas, A.D., Mavrommati, I. (eds.) The Disappearing Computer. LNCS, vol. 4500. Springer, Heidelberg (2007)
26. Martens, M.H.: Assessing Road Sign Perception: A Methodological Review. Transportation Human Factors 2(4), 347–357 (2000)
27. Lewis, I., Watson, B., Tay, R.: Examining the effectiveness of physical threats in road safety advertising: The role of the third-person effect, gender, and age. Transportation Research Part F: Traffic Psychology and Behaviour 10(1), 48–60 (2007)
28. Plutchick, R.: The nature of emotions. American Scientist 89, 344–350 (2001)

Affordances in Conversational Interactions with Multimodal QA Systems

Andreea Niculescu

HMI Group, University of Twente
niculescuai@ewi.utwente.nl

Abstract. Implementation of adequate conversational structures is a key issue in developing successful interactive user interfaces. A way of testing the adequacy of the structures is to prove the correct orientation of each communicative action towards a preceding action. We refer to this orientation leading to a certain response as the affordance of the communicative action. In this paper we present a case study where affordances of implemented conversational structures (including verbal and graphical elements) in a multimodal medical QA system are identified applying Conversation Analysis (CA) tools and tested using the Cognitive Walkthrough (CW) method. The CW method was modified to fit the conversational approach and tested with five expert evaluators. Results showed that the affordance analysis helps detecting inefficient constructions leading to disruptions in the dialog flow, spots unnecessary functions and provides important insights on systems easy-of-use.

Keywords: Affordances, Conversational Interactions, Multimodal QA systems, Usability Inspection, Cognitive Walkthrough.

1 Introduction

This paper discusses the design evaluation of a multimodal QA system from the perspective of the affordance concept.

We claim that many problems associated with natural language based interactions are originated from the lack of deeper understanding of underlying conversational structures. Typical Graphical User Interfaces (GUI) use preponderant visual interaction elements. Therefore, the affordance design focuses mostly on visual elements, regarding verbal units as simple cognitive support for the graphics.

Multimodal QA systems are a special case of GUIs in which graphical elements are combined with large text units. The rapport between verbal and graphical elements is here reversed: the text units are the one supported by graphical elements being the quintessence of the QA interaction. Consequently, we believe that interaction designers should include more extensive analyses of verbal affordances when designing interactive systems based on conversational structures, like QA or dialog systems.

Since the concept of affordance has been often subject of intense controversies in HCI debates once introduced by Donald A. Norman in [1], many researchers

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 221–236, 2008.

such as Gaver, Hartson and Noman himself struggled with several definitions and categorizations in an effort to clarify the concept making it operational for evaluations. Section 2 comments in details these theoretical considerations. In section 3 some short examples of practical applications of affordances are presented followed by section 4, where the affordance concept is integrated in the framework of the conversational analysis protocols. The case study, including methodology, short system overview, questionnaire and scenario design are presented in section 5. The results are largely discussed in section 6. This paper ends with conclusions containing the result summarization and improvement suggestions for the QA interface.

2 What Are Affordances?

The concept of affordance was developed by the American psychologist J.J. Gibson in [2] and [3] it is a significant part of his ecological theory of direct perception. Gibson defined the term "affordance" as latent action possibilities existing in an environment independent of the individual's ability to perceive them. These action possibilities are in relation with the actor's capabilities of action being independent of his culture, prior knowledge or expectations. Any substance, any surface, any layout has some affordances with respect to a certain actor. According to his theory, the action possibilities indicated by affordances are perceived visually in a direct way that does not require mental information-processing activity, i.e. the immediate perception of the environment will inevitably lead to a certain action.

Gibson's theory of non-conscious information pick-up was criticized as it fails to explain how actors assign meaning to what they see deciding whether to perform an action. Even if the existence of affordances is independent of the actor's experience and culture, the ability to perceive such affordances may be dependent on them. Therefore, the actor may need to learn first to discriminate the information he gets from the environment in order to perceive it directly [4]. Although Gibson rejected the involvement of mental activities in the process of direct perception, he conceded that even for the most "basic" affordance perception might need to somehow develop - a clear suggestion that learning could be involved.

Other criticisms to Gibson's theory refer to the fact that affordances are defined as being relational but the nature of the relation between actors and environment are not further discussed. Even though the theory presents some illustrative examples of affordances, it doesn't provide an analytic way of identifying affordances[5].

Despite all the criticism, Gibson's theory of affordance brought radical changes in the field of perceptual psychology and was successfully adopted as a key concept in other fields, such as cognitive science, robotics, artificial intelligence, design etc.

In the HCI field the term "affordance" was introduced by David A. Norman in his book "The Psychology of Everyday Things" [1]. Norman adopted Gibson's concept and used it to address design aspects of artifacts considering technology as part of the environment. He defined affordances as physical, "perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used. [...] Affordances provide strong clues to the operations of things. Plates are for pushing. Knobs are for turning. [..]. When affordances are taken advantage of, the user knows what to do just by looking: no picture, label, or instruction needed" [1].

However, even though Norman borrowed the term from Gibson he disagreed with his theory about whether the mind processed or simply "picked up" information (see [1], [6] and [7]). Consequently, Norman departs from Gibson's theory and considers affordances perceived properties of objects that may or may not exist. The perception is determined not by the action capabilities of the actor as Gibson says, but by his mental and perceptual capabilities. Moreover, past knowledge and experience are tidily coupled with affordances in Norman's view [4].

For Norman, the notion of affordance becomes a mixture between actual (shape, material, color) and perceived properties of the object where the perceived properties are in fact the suggestion of how the object might be used.

The lack of separation between physical properties and perceptual information about their use created a rather ambiguous definition of affordance and generated large discussions about the meaning of the term in the HCI community.

A substantial contribution to clarify the concept bringing the Gibsonian thinking back in the HCI world was made by the interaction designer William W. Gaver. Similar to Gibson, Gaver argued that affordances do exist independently of their perception, being ontologically but not epistemologically relevant. He defined the affordance concept as the property " [...] of the environment relevant for action systems" [8] and proposed a taxonomy where the affordance concept is separated from the perceptual information available about it.

In Gaver's framework affordances (aff.) and their perceptual information (per.inf.) are defined as entities taking binary values such as *"yes"* and *"no"*. Their combinations result in four types of affordances called: perceptible, hidden, false and correct rejection.

a. Perceptible affordances: (aff.: *yes*, per.inf.: *yes*) offer a link between perception and action by signalizing a possible action in a visible way.
b. Hidden affordances: (aff.: *yes*, per.inf.: *no*) offer no link between perception and action; actions are possible but there is no signal acknowledging their existence (e.g. a hidden door).
c. False affordances: (aff: *no*, per.inf.: *yes*) offer a link between perception and a non-existing action possibility; actions are mistakenly signalized as being possible (e.g. a door might appear to afford opening, but it won't afford if it's locked).
d. Correct rejection: (aff: *no*, per.inf: *yes*) refers to the situation where no action is afforded or signalized (i.e. no affordance).

While conveying the categorization to interaction design field designers should avoid false and hidden affordances as being a sign of weak design: false affordances bring users on a wrong path while hidden affordances waist resources, as users will probably encounter difficulties on detecting their existence. Instead, designers should concentrate on making affordances perceptible or creating situations where the lack of affordance is correctly rejected.

Another attempt to extend and refine Norman's concept of real and perceived affordance came from Hartson [9]. Hartson proposed four complementary types of affordance in the context of interaction design and evaluation: cognitive, physical, sensory and functional affordance.

Cognitive affordance -corresponding to Norman's perceived affordance- is associated with the semantics of the interfaces and refers to design features that help users knowing something (e.g. the label of a button indicating what will happen if a user clicks on it).

Physical affordance -corresponding to Norman's real affordance- is associated with characteristics concerning the "operability" of the interface, and refers to design features that help users to accomplish accurately a physical action in the interface (e.g. the size of a button that is large enough to allow users to click on it).

Sensory affordance is related to "sense-ability" characteristics of the interface and targets design features that help users perceiving (e.g. seeing, hearing, feeling) something (e.g. the font size of a label). Sensory affordance plays a critical supporting role to cognitive and physical affordances.

The last category, functional affordance addresses design features that help users to accomplish work (e.g. the internal system ability to sort numbers invoked by a user who clicked the 'sort' button [9]).

There are several other interesting interpretations and formulations of the affordance concept but due to obvious space limitation we discussed in details only a few of them that we considered as being the most relevant to our analysis.

3 Practical Dimensions of Affordances

The study of affordance goes beyond theoretical speculation; authors like Vainio et al. [10] validated the affordance concept as part of an interesting empirical study. They showed that participants during several tests could identify objects faster if they were congruent with an observed action prime (e.g. power grasp - power grasp compatible object) rather than incongruent (e.g. power grasp - precision grasp). They concluded that motor knowledge plays an important role in object identification and, consequently, action-related information associated with an (graspable) object is an inseparable element of that object's representation.

In the HCI field the affordance concept has found its practical settings as design model and analysis tool for physical and graphical user interfaces.

Sheridan & Kortuem [11] proposed an affordance-based design model of physical interfaces for ubiquitous environment. They proposed an experimental method to study object affordances showing how the method can be applied to the design of concrete physical interface artifacts.

Luo et al. [12] used the affordance design model as theoretical basis and methodological underpinning to evaluate an e-learning program on mammogram reading.

Hartson [9] explored the relationship between the affordance types associated with usability problems and provided examples and a methodology scheme for practitioners on how to identify affordances issues involved in flawed design cases.

However, the theory of affordance and its practical applications concern mainly the visual perception of environment objects analyzing verbal element only marginally, e.g. only when they are meant to support visual elements (see Hartson's cognitive affordance). Since new media technologies such as interactive information systems like QA or dialog systems are design artifacts that use elaborated conversational structures with preponderate verbal text elements there is a strong need to consider such elements as integral parts of the interaction design. Therefore, we propose a design evaluation that uses the affordance concept to analyze text units and graphical elements.

4 Affordances in Conversational Interactions

An important characteristic of conversational interactions in general is the fact that they are deeply anchored in the cultural context of use; that means they are based on conventions and constrains of the socio-cultural environment following a rigorous protocol course.
Contrary to Norman who argued that the socio-cultural world is placed outside the domain of affordance [6], Gaver emphasized the role of the culture together with other factors such as experience and learning involved in the process of perception of affordances [8]. We also believe these factors are not to be considered affordances but have an important function in making affordances visible.

A way of detecting these factors and implicitly affordances in conversational interactions is to apply conversational analysis (CA). In CA, conversations can be considered "environments" where certain types of actions such as gestures, mimics, verbal statements are afforded in certain circumstances - in terms of non-violating coherence principles and cultural constrains. Interlocutors can express their communicative intentions both verbally (through speech) and non-verbally (through gesture and mimic). Analyzing the organization of each conversational sequence it can be determined what kind of action possibilities (affordances) have the participants in a certain moment of the conversation. We consider these affordances from a pragmatic perspective, i.e. action possibilities oriented to achieve a certain goal.

Conversations are by nature interactive and follow a relatively strict turn-based protocol. We address in this paper only the case of closed-domain question-answer interactions since the system analyzed in our case deals with this type of interaction.

In general, question-answer conversations are fully structured in adjacent pairs, meaning that all exchanged turns are functionally related to each other in such manner that the first turn requires a certain type (or range of types) of second turn [13]. The adjacent pairs are grouped in three separate categories, corresponding to a conversation initialization, termination (both containing greeting -greeting pairs) and body sequence (containing question-answer pairs). These three categories form together what we call conversation protocol.

4.1 Conversation Initialization and Termination

Starting and ending a conversation are levels of phatic communication with social functions: they are responsible for establishing rapport or quitting the interaction "circle" in a polite way.

A conversation usually starts with a signal showing the readiness to engage in a conversation. Such signals are salutation forms, self-presentation (if the interlocutors haven't met before), non-verbal gestures (hand shake, hugging, kissing, hand waving, gazing), mimic (smiling), changing the corporal position towards the interlocutor, etc.

A similar protocol for ending the conversation includes farewells and thanks exchanging, waving arms, hugs, kisses, glancing away, re-orienting body posture away from interlocutor etc.

Each performed action affords in principle a similar one in return: a greeting, a smile, a self-introduction affords symmetrical responses from the interlocutor. However, the realization of conversation initialization might differ across culture taking into account participants' gender, age, social position and degree of acquaintance. For example, in Western cultures a stretched out hand will afford hand-shaking; in Muslim countries such a gesture will afford hand-shaking usually if both interlocutors are men; women instead will press their hand upon the chest to signalize salutation response avoiding at the same time the direct contact with the opposite gender.

4.2 Conversation Body

The conversation body contains the essential part of the interaction, namely the information exchange (also called informational communication).

The information exchange may start with a short explanation of the intended nature of the conversation-to-be preceding (the first question-answer exchange). At this point a common ground (a set of propositions that make up the contextual background for the utterances to follow) can be established.

From a pragmatic point of view a question may afford following responses: a matching answer, acknowledgment of ignorance, suggestion for asking someone else (re-routing), intermediary questions to clarify a previous question, postponement, refusal to provide an answer, feedback showing that the question was understood or a request for time to process the question, etc.

In case of miscommunication repair strategies occur in form of explanatory adjacent question-answer pairs.

The turns can be accompanied by non-verbal cues such as gestures and mimic used to emphasize the content; e.g. gaze signalizes attention and readiness for interaction, rising eye-brow shows surprise, smile acknowledges agreement, etc.

Question and answer pairs must respect the coherence principle by being semantically and meaningfully related to each other. In order to achieve coherent information exchange, syntactical features such as anaphoric, cataphoric and deictic elements may be used. Also logical tense structure, as well as presuppositions and implications connected to general world knowledge are deployed to coherently connect answers to questions[14].

In common practice interlocutors do not perform their utterances at the same time. Speakers usually take turns to talk. Overlapping and simultaneous talk is generally seen in Western cultures as unpleasant. The turn-taking usually occurs at the utterance ends, often signalized by silence.

Interruption might be allowed if one of the interlocutors signalized verbally or by gesture the wish to take the turn[15].

5 Method

Even though conversational interactions with multimodal systems differ in many aspects from their human counterpart, they generally follow the same conversation protocol consisting of initialization, body sequence and termination. This similarity is intentionally simulated by designers in order to increase the system's easy-of-use and to make users' answers predictable.

As theoretical framework we adopted Gaver's taxonomy to identify affordance values and Hartson's scheme to establish affordance types. The analysis followed the conversational protocol steps described above.

The test was carried out using the Cognitive Walkthrough (CW) method. The method is a usability engineering tool meant to help designer teams to quickly evaluate interaction systems from early stages of development. It does not require a fully functioning prototype (as it is the case with the system on which the test was performed) nor users involvement. CW emphasizes cognitive aspects, such as learnability by analyzing users' mental processes required for each step[16]. Design experts perform the test taking into account the potential user perspective with the purpose of identifying problems that might arise during the interaction. After each scenario the experts are asked to answer questions stated in a questionnaire.

5.1 The IMIX System

Our study was performed on IMIX, a multimodal interactive QA system for medical queries with extended follow-up questions functionality.

IMIX was developed for educational purposes and can deal with medical encyclopedic questions, i.e. general questions that do not require expert knowledge like diagnostic questions or complex medical analysis [17]. The system's users are expected to be people with no professional knowledge of the medical domain. They will probably make use of such services only occasionally. No special training is required to interact with IMIX.

The system is primary text-based but allows users to use optionally speech[1]. Attached to IMIX is a talking head called Ruth.

[1] During the test only the text-based modality was deployed.

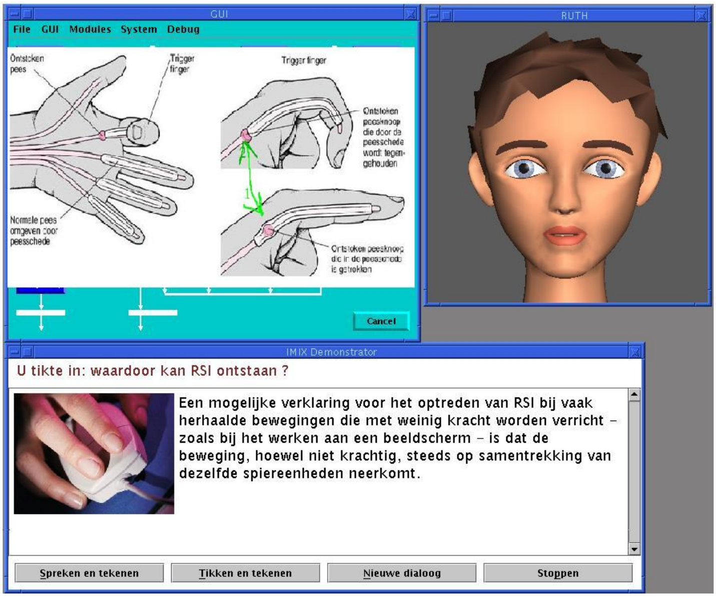


Fig. 1. Screenshot of the IMIX system

Cognitive research into multimedia has shown that the use of text combined with pictures may substantially contribute to the user's learning process of the presented material [18],[19]. Therefore, the IMIX answers contain text combined with suitable static images. The answers are made-up by matching the query to document fragments from the data base. The text answers may be both spoken by Ruth or displayed on the screen. Optionally, follow-up questions can be formulated as text, speech or drawing [20].

5.2 Prior Issues to the Test

Before starting the test the evaluators got paper sheets containing preliminary information about the test goal, a short description of the term "affordance", a detailed explanation of the human conversation protocol and a general description of the IMIX system.

Afterwards, the evaluators received the scenarios containing specific tasks to accomplish, a list of correct actions required to complete each of these tasks and a separate questionnaire for each scenario.

5.3 Scenarios and User Profiles

Three scenarios covering the conversational structures implemented in the IMIX system were developed. We tried to design the scenarios as pleasant and humorous as possible in order to achieve an enjoyable interaction. Each scenario focuses on a specific task.

In the first scenario evaluators were asked to put a single question and analyze the corresponding conversation protocol. The scenario identifies the situation of a naive user with little medical expertise who uses IMIX to find out what means Repetitive Strain Injury (RSI).

In the second scenario the evaluators had to concentrate on the special case of follow-up questions using drawing and/or typing options. The user profile of this scenario corresponds to a subject with search engine expertise and interests in the medical domain. He uses IMIX to find information related to liver functions.

The third scenario addresses repair and meta-communication strategies when the answer to a question is not found. The user profile addresses an expert user who uses IMIX for entertaining purposes. He is seeking for information about the SARS virus.

The evaluators were asked to keep in mind the aim of the test: to look at the way the user is invited to interact with the system and NOT at the answer quality he/she might get back. They also had the possibility to repeat a scenario several times if they wish so.

5.4 Questionnaire Design

For each scenario a separate questionnaire was developed. The questionnaires were designed in accordance to Cognitive Walkthrough (CW) method[21]. However, the questions were adapted to fit the special case of verbal interactions and grouped in three units, each one corresponding to a separate conversation protocol category.

The purpose of the questionnaires is to detect affordances of the conversational sequences implemented in the protocol. The structure of the questions is similar for each unit. Evaluators have first to detect elements signalizing the current protocol. Then they have to anticipate users' re-action given a certain conversational sequence. Furthermore the evaluators have to determine whether the users' responses are acknowledged by the system and how they will perceive this feedback. Eventually each question unit ends with a question about potential violations of the conversation protocol. This last question is meant to catch issues that might have "escaped" the evaluator's observation.

5.5 Pilot Study

Before starting the experimental run a first pilot study with one expert evaluator was accomplished. From the pilot study three main observations could be gathered:

1) The relatively high difficulty degree of the question demanded the presence of experienced evaluators having some affinity with the CW method.
2) Typical CW questions like "Will the user notice the conversational starting signals(as the correct action available)?" are too general. Precise formulations similar to "Will the user understand this signal as an invitation to start a conversation?" seemed to be more appropriated even if the questionnaire size increases, e.g. for each signal one separate question.

Table 1. Classical CW versus Adapted CW

Classical CW	Adapted CW
1. Will the users be trying to produce whatever effect the action has?	1.0 What elements are used to signalize [*a certain action*]? 1.1 Does the word [..] suggest [*a certain action*]? 1.2 What kind of statements affords the question [..]?
2. Will users be able to notice that the correct action is available?	2.0 Will the users see there is [*a certain*] option ? 2.1 Will the users understand this signal as an invitation to [*do a certain action*]?
3. Once users find the correct action will they know that it is the right one?	3.0 Will the users know how to use [*a certain*] option?
4. After the action is taken, will users understand the feedback they get?	4.0 Will there be a feed-back to acknowledge the action performed by the users? 4.1 Will the users understand these kind of feed-back?
	5.0 Does the conversational protocol get in any way violated?

3) Since the answers to the CW questions are often not straightforward, requiring some deliberation time it seemed wise to record the testing session. In this way a considerable amount of time could be saved and no observation could get lost.

6 Results

The test was completed by five evaluators with design expertise recruited from our department. All evaluators except one were novice using the system. The results of their evaluation are summarized below following the conversational protocol categorization:

6.1 Conversation Initialization

The conversation initialization implemented in IMIX doesn't afford symmetrical response. The conversation's start is signalized by a textual welcome message, a short system presentation, a 'start' button and a talking head emerging from the background gazing and rising eyebrows. At this point the only afforded action is the pressing of the 'start' button to begin the "conversation". No other actions like greetings or salutation gestures in return are afforded, even though according to the conversation protocol a greeting affords another greeting. The occurrence of a signal (greeting message) combined with the lack of an adequate response (no greeting in return) indicates the presence of a false cognitive affordance.

Most of the evaluators agreed that this stage of the interaction has less resemblance to what is normally called a "conversation". The talking head appearance is not a very convincing invitation to talk even if its blinking eyes indicate a waiting behavior. One of the evaluator argued that the presence of speech, e.g. a welcome message read by the talking head would increase the users' feeling of being involved in a conversation.

An adjustment in the head mimic would be beneficial as well: a gazing behaviour combined with smiling is a more appropriate way to start a conversation.

The short system presentation was criticized as being too technical: especially less experienced users would have difficulties in understanding the meaning of having a "multimodal dialog" with the system.

Table 2. Affordances in conversation initialization

Affordance type	Affordance value	Action
Physical & Cognitive	Perceptible	Press 'start' button
Sensory & Physical	False	Click on highlighted words
Cognitive	False	Give symmetrical response

Also the text color should be uniform -one criticism addressed the presence of colored words in the text message, a fact that could mislead internet experienced users to click on it, as it is common on websites with embedded links. This would be a false sensory affordance.

6.2 Conversation Body

The conversation body includes single question-answer sequences, follow-up questions, meta-communication and repair strategies.

Single Question-Answer Sequences
The conversation body begins once the 'start' button is pressed. The users arrive on a new screen where they receive some brief instructions on how to interact with the system.

At this stage of the conversation the users have to choose between two input options: speech or typing. All evaluators agreed that input selection modalities seem to be afforded in a proper manner: the buttons are intuitively labeled and it was estimated that all user categories won't have difficulties to select an input option. A suggestion was made to use additionally explanatory icons like a pen for the typing option and a microphone for the speech option.

It was criticized the presence at the same level of two other buttons: one for the stop option and the other one for the new dialog. The 'stop' button should be placed in a corner -in order to be congruent with the typical design of closing buttons while the 'new dialog' button should be removed as being at the moment functionless and indicating a false physical affordance.

Another remark was to adapt the head position towards the typing input field while users are typing a question in order to increase the interactive feeling and to give a certain feedback. Due to the lack of adequate mimic reactions and synchronization with the current conversation stage the talking head gives the impression that it doesn't belong to the system.

By selecting the option "typing" an input field appears on the same window. The input field affords sentence-like questions as well as keywords. Most of the evaluators -excepting one- considered that the full sentence capability won't be easily perceived by more experienced users; they would probably associate the system's functionality with the one of a typical search engine and consequently would use keywords. The presence of a relatively extended input field is not a clear indication of the expected input and if sentence-like input is desired a short how-to-ask example should be provided. It can be concluded that the input field has hidden physical affordances.

The input field is introduced by the question "What would you like to ask or say?". The designer's intention was to let users know the system is able to

Table 3. Affordances in single question-answer sequences

Affordance type	Affordance value	Action
Cognitive	Perceptible	Chose input options
Physical	Perceptible	Type in the input field
Physical & Cognitive	False	Click 'new dialog' button (first dialog screen)
Cognitive	False	Perform whatever question
Physical	Hidden	Put sentence-like question
Cognitive	Hidden	Perform greeting
Cognitive	Hidden	Perform transition statement
Physical & Cognitive	Hidden	Press 'new dialog' button to type
Physical & Cognitive	Hidden	Press 'follow-up question' button to type

handle different types of statements like full-sentences questions, greetings or even transition formulations ("ok" or "thank you"). But being rather too open, the question suggests it can deal with any kind of statements which certainly is a false cognitive affordance.

On the other hand, most of the evaluators -excepting one- concluded that ironically, especially experienced users would not be aware of what exactly they can utter, e.g. greetings and transition statements affordances remain hidden, as nothing clearly indicates their possible usages.

We continue the analysis considering the case where a naive user will enter a greeting. The system will logically respond repeating the same question ("Hello! What would you like to ask or say?"), but won't indicate how to continue the dialog as the input field disappears. So far a direct answer is not afforded. The users need to press the button for either new-dialog or follow-up question in order to get to the input field to type in, a fact that complicates the conversational flow.

None of the labeled buttons suits semantically the actual conversational situation -the 'new dialog' button should be used in situations where a dialog session re-initialization is wanted, while the 'follow-up' button refers to situations where users are looking for more detailed information in the medical answer. Therefore, it can be concluded that both buttons afford in this conversational sequence a hidden action, namely to allow users to get back to the typing field.

Follow-up Questions

After receiving the first answer the users have the option to continue the information exchange on the same topic by selecting a 'follow-up question' button. The button is labeled with a text indicating that users can type or point at something in the answer. However, the pointing option is not intuitive and has not a specific usage indication. Besides, not only pointing but also drawing is supported, a fact that the label doesn't specify. All evaluators agreed on the fact that all user categories wouldn't know what the option does and how to use it. Moreover, it is not clear which advantages it has compared to the typing option. Therefore, we identify here a hidden physical affordance.

It is also not very clear the way a "drawn" follow-up question is entered for further processing. Since the 'ok' button located in the proximity of the input field can be also used for this purpose, the evaluators concluded the button has hidden physical and cognitive affordances.

Table 4. Affordances in follow-up questions

Affordance type	Affordance value	Action
Physical	Perceptible	Type in the input field
Physical	Hidden	Use the mouse to "drawn" a question
Physical & Cognitive	Hidden	Use the 'ok' button to enter a "drawn" question

There is a feedback to acknowledge the waiting pause and the users' query but the feedback is not specially meant for follow-up questions, fact that should not disturb the communication flow.

Meta-Communication and Repair Strategies
When the answer of a question is not found the system displays a message requesting rephrasing. The function of the rephrasing request should additionally help users to become more successfully in finding the desired information.

All evaluators found the request not supportive at all; according to their estimation even expert users would experience problems rephrasing their question.

After the rephrasing request users can choose between the follow-up question (in the form of typing or pointing) or the new dialog option in order to get cumbersomely back to the typing field. Both options were considered inadequate for this particular stage of the conversation. Just like in the follow-up paragraph these two buttons indicate the presence of hidden physical and cognitive affordances.

Table 5. Affordances in meta-communication

Affordance type	Affordance value	Action
Cognitive	Perceptible	Rephrase question
Physical & Cognitive	Hidden	Press 'new dialog' button to rephrase
Physical & Cognitive	Hidden	'Follow-up question' button to rephrase

Conversation Termination
The interaction can be interrupted by clicking the 'stop' button. A real conversational termination is not afforded. Users don't have the possibility to verbally express the intention to leave the conversation, as no typing field was designed at this stage of the conversation. They could click on the 'new dialog' button and type a farewell greeting as the system affords such statements. But this option seems rather a less logical as nobody will probably think to start a new dialog when in fact he/she wishes to stop it. Besides, even if the system replies logically to the farewell greeting it doesn't allow a verbal termination of the conversation.

There is no feedback to acknowledge the end of the conversation and users get the general impression of a system crash by clicking the 'stop' button. We certainly face a conversation protocol violation.

Table 6. Affordances in conversation termination

Affordance type	Affordance value	Action
Physical & Cognitive	Perceptible	Press the 'stop' button
Cognitive	False	No symmetrical response

7 Conclusions

Extrapolating the affordance definition given by Gaver we considered in this paper interactive information systems as artificial environments where verbal and graphical elements are artifacts leading users to perform certain actions. Therefore we proposed a design evaluation in which not only graphical but also verbal elements can be analyzed under the framework of the affordance concept.

The results of our experiment revealed several inefficient structures that could be identified analyzing affordances of conversational structures. The conversation initialization and termination implemented in IMIX do not afford a symmetrical response from the user, perturbing the natural dialog flow. Systems question formulations are too open, a fact that might generate false expectations or disorientation. The labeling of buttons should reflect the actions induced by the buttons. A question should automatically generate a response environment avoiding unnecessary pressing of additional buttons. The system reactions should be consistent at all appearance level, i.e. verbal, mimic, gestures.

The study of affordances also showed unnecessary functionalities that might be removed or adapted in order to become useful. For example, the presence of buttons leading to certain actions should be in accordance with the conversational sequence they are designed for: it makes no sense to start a "new dialog" when no other dialog had been started before. The affordance of certain conversational structures like greetings in the middle of an interaction shows a cooperative behavior. However, it is unlikely that someone would use greetings at that particular conversational stage. Special features like pointing or drawing on a virtual surface should be briefly introduced to users. It is rather unexpected that someone should use an unfamiliar option to ask questions when he/she has more natural choices like typing or speaking.

The affordance analysis also provided important observations about the system's easy-of-use. Users may not understand the system's description, as it seems to be too technical, may not be aware of its full sentence capabilities, may not know whether other transition statements are allowed, may experience difficulties using the pointing/drawing option or rephrasing their questions and may probably feel annoyed when they expect to be able to type a question and no input field is provided.

Last but not least the affordance analysis of verbal elements proved to be beneficial and confirmed our initial claim that many problems associated with natural language based interaction are originated from the lack of deeper understanding of communicative structure: most of false and hidden affordances identified were cognitive nature (6 pure cognitive and 6 physical-cognitive out of a total of 14).

We concluded that understanding affordance of verbal and graphical elements and being aware of their roles in conversational interaction design can help practitioners in diagnosing usability problems from early stage of development, since the affordance analysis using CW methods provides a useful and informational rich perspective for qualitative evaluations of prototypes without implying costly user studies.

Acknowledgments

The author is grateful to Boris van Schooten and Rieks op den Akker for interesting hints and discussions regarding this research, to Betsy van Dijk for useful comments during the pilot study, to Ronald Poppe, Olga Kulyk, Bart van Straalen and Dennis Reidsma for participating in this experiment and to Hendri Hondorp and Dennis Hops for helping with technical setup. Many thanks to Egon L. van den Broek for several valuable theoretical suggestions, to Blasimir Villa Rodriguez for practical discussions about concepts and careful proof-reading and to anonymous reviewers for helpful improvement suggestions.

References

1. Norman, D.A.: The Psychology of Everyday Things. Basic Book Inc., New York (1988)
2. Gibson, J.J.: The theory of affordances. In: Shaw, R., Bransford, J. (eds.) Perceiving, Acting, and Knowing, Erlbaum, Hillsdale NY (1977)
3. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
4. McGrenere, J., Ho, W.: Affordances: Clarifying and evolving a concept. In: Proc. of Graphic Interface, Montreal, Canada, pp. 179–186 (2000)
5. Oliver, M.: The problem with affordance. E-Learning 2(4), 402–413 (2005)
6. Norman, D.A.: Affordances, conventions and design. Interactions, 38–43 (1999)
7. Norman, D.A.: Memory and Attention: an introduction to human information processing. Wiley & Sons, London (1969)
8. Gaver, W.W.: Technology affordances. In: Proc. of CHI, New Orleans, U.S. (1991)
9. Hartson, H.R.: Cognitive, physical, sensory and functional affordances in interaction design. Behaviour & Information Technology 22(5), 315–338 (2003)
10. Vainio, L., Symes, E., Ellis, R., Tucker, M., Ottoboni, G.: On the relations between action planning, object identification, and motor representations of observed actions and objects. Cognition 108(2), 444–465 (2008)
11. Sheridan, J.G., Kortuem, G.: Affordance-based design of physical interfaces for ubiquitous environments. In: Youn, H.Y., Kim, M., Morikawa, H. (eds.) UCS 2006. LNCS, vol. 4239, pp. 183–199. Springer, Heidelberg (2006)
12. Luo, P., Eikman, E.A., Kealy, W., Qian, W.: Analysis of a mammography teaching program based on an affordance design model. Academic Radiology 13(12), 1542–1552 (2006)
13. Schegloff, E.A., Sacks, H.: Opening up closings. Semiotica 8(4), 289–327 (1973)
14. Bumann, H.: Lexikon der Sprachwissenschaft, 3rd edn. Kroener Verlag, Stuttgart (2002)
15. Cassell, J., Bickmore, T., Vilhjalmsson, H., Yan, H.: More than just a pretty face: Affordances of embodiment. In: Proceedings of 2000 International Conference on Intelligent User Interfaces, New Orleans (2000)
16. Holzinger, A.: Usability engineering for software developers. Communications of the ACM 48(1), 71–74 (2005)
17. Theune, M., van Schooten, B., op den Akker, R., Bosma, W., Hofs, D., Nijholt, A., Krahmer, E., van Hooijdonk, C., Marsi, E.: Questions, pictures, answers: introducing pictures in question-answering systems. In: Proc. of the 10th. International Symposium on Social Communication, Santiago de Cuba, Cuba (2007)

18. Mayer, R.E.: Multimedia learning. Cambridge, MA (2001)
19. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic media in computer science education; content complexity and learning performance: Is less more? Educational Technology & Society 11(1), 279–290 (2008)
20. op den Akker, R., Bunt, H., Keizer, S., van Schooten, B.: From question answering to spoken dialogue: Towards an information search assistant for interactive multimodal information extraction. In: Proc. of Interspeech, Lissbon, Portugal (2005)
21. Wharton, C., Rieman, J., Lewis, C., Polson, P.: The cognitive walkthrough method: A practitioner's guide. In: Nielsen, J., Mack, R. (eds.) Usability Inspection Methods, pp. 105–140. Wiley, New York (1994)

E-Learning: A Tool for Teachers with a Disability

Olga Revilla

Escuela Técnica Superior de Ingeniería Industrial
Universidad Politécnica de Cataluña. Spain
itakora@gmail.com

Abstract. Usability and accessibility in e-Learning has usually focused on the learners' experience and how to improve the learning efficiency. However, very little studies analyze the accessibility and usability issues from the teachers' point of view, especially when the teachers have one or more disabilities. Teachers with disabilities find it very difficult to manage in their off-line jobs. Therefore, it is expected that technology can make them easier to work as on-line teachers. In this paper, I will summarize which troubles Spanish teachers with disabilities at Elementary degree find out in their job and how the e-learning can improve their quality of job life.

Keywords. Usability, e-Learning, Accessibility, Disabled teachers.

1 Introduction

Usability and accessibility in e-Learning has usually focused on the learners' experience and how to improve the learning efficiency. However, very little studies analyze the accessibility and usability issues from the teachers' point of view, especially when the teachers have some disabilities. Teachers with disabilities find it very difficult to manage in their off-line jobs. Therefore, it is expected that technology can make them easier to work as on-line teachers.

I have done a research study to find out whether it is possible for people with disabilities to work as teachers and if so, if they experience problems when trying to carry out their tasks. From the problems outlined, I will discuss if going into the on-line world that is, carrying out her teaching duties through an e-Learning platform facilitates them their job. Thus, in this paper, first I will summarize which troubles Spanish teachers with disabilities at Elementary degree find out in their job and how the e-learning can improve their quality of job life. Next, I will go deeper into each need or requirement the teachers have and how to accomplish that requirement. And finally, I will propose my research on usability and accessibility of e-learning platforms since the elementary teachers' point of view.

2 People with a Disability and Job Market

Legislation at the EU countries tries to guarantee the incorporation of persons with a disability into the educational system and the job market. The following data are gathered from the Spanish Statistic Organism (INE), in August 2003 [1]:

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 237–246, 2008.

Table 1. People with and without disability according to their educational level

Educational level	With disability (%)	Wthout disability (%)
Total	2.339,2 (100%)	24.600,9 (100%)
No studies	175,9 (8%)	150,6 (1%)
Elementary Degree	1.197,2 (51%)	6.190,9 (25%)
Secondary Degree (1st grade)	518,0 (22%)	7.201,1 (29%)
Secondary Degree (2nd.grade)	244,0 (10%)	5.227,3 (21%)
University	204,1 (9%)	5.831,1 (24%)

Although these figures show that persons with a disability are almost completely integrated into the educational system, the next one will reveal that they find it difficult to enter into the job market.

Table 2. People with and without disability according to their job status

Job status	With disability (%)	Wthout disability (%)
Total	2.339,2 (100%)	24.600,9 (100%)
Not-working	1.551,4 (66%)	7.369,4 (30%)
Workers unemployed	120,9 (5%)	1.890,2 (8%)
Workers employed	666,9 (29%)	15.341,3 (62%)

Many people with a disability face the employers' distrust that finally ends in choosing other persons without any visible disability, despite the favorable economic conditions the government offers to companies that contracts people with disabilities. This is even clearer when they argue that 'how can a blind person take care of a class plenty of 3-age-children?" or "would you let your son or daughter behind the guardianship of a person with frequent attacks of epilepsy?" This theory is analyzed by [2] and [3], where several teachers with disabilities tell their difficulties to join the job market because of their disability and people prejudices.

Fortunately, disabled persons show everyday their capabilities, and recruiters are realizing it. For example, in Great Britain, for example, the Training and Development Agency for Schools (TDA), which have just lunched a campaign for recruiting disabled teachers because their "ability to change students' views of disabled people and the part they can play in society is particularly worthwhile" [4].

3 Problems of the Teachers with Disabilities

Although there are no official data of teachers with disabilities, we can deduce some problems they have and the requirements from the following facts.

3.1 Accessing to the Job

Most of the elementary and secondary schools in Spain are held by the local government, so teachers have to sit a competitive examination to access to their job. If they want, they can compete for a special quota: the only thing they have to do is to prove a committee that they are capable for the job of teacher. This point is critical to a lot of people because there are not clear rules about how to evaluate if a person is capable or not to teach (they say 'it depends on the person and his/her medical history'). Due to privacy rules, there are not public documents of the activities of these committees; the only case that is public is Mr. Juan Carlos Sainz, who claimed for politicians' help as he was denied to be qualified to be an elementary teacher [5]. Although he got his diploma on pedagogy, he was told that his disability (blindness) was not compatible with 'the control, surveillance, assistance and attention of children from 3 to 12 years without a full-time assistant'. This is over the reasonable adjustments the employer must do.

So here it is one of the first requirements we can ask for an e-learning platform: it must help teachers to respond to all of the needs of their pupils, not only educational tasks, but also the other ones a teacher is supposed to do without unreasonable adjustments, like the control, surveillance, assistance and attention of the students, no matter which age they are.

3.2 Mental Disorders

Following the main trade union in Spain studies [6], mental disorders like stress, depression and anxiety are very common in elementary teachers. Figures show that this pathologies are the 12,2% of the sick leaves.

The reasons are the transformation of social relationships, moral values changes, diversity of the pupils and the increasing number of problem students. Beside it must be remarked the transfer of responsibilities from the family to the teachers, the lack of social recognition of these professionals and the assumption of others responsibilities for lack of specialized personnel.

According to the problems identified above, e-learning platforms should help teachers in:

1. Managing different types of activities.
2. Personalizing the effort for each student.
3. Improving the communication among the others teachers.
4. Improve the communication with the administration.
5. Generate confidence in himself/herself as a good teacher.

Special effort is needed to support teachers with cognitive and neurological disabilities like seizure impairments.

3.3 Violence

Elementary degree teachers in Spain recognize to have trouble of violence in their school. According the trade union ANPE [7], the 65% of teachers have problems in developing their tasks. Teachers with disabilities are not an exception, as violent persons usually bother the weaker persons they find.

E-Learning benefits both teachers, violent and non-violent students, as they have no physical contact with each other. The inside problem of this approach is the isolation they must feel, which can be even more problematic.

In any case, Spanish punishment for the violent students is sending them off the school for some days. E-learning can make these students not to be behind of the rest of the students, and teachers can evaluate and teach them.

3.4 Educational Material Not Adapted

As teachers with disabilities are not 'the usual' in the educational primary level, almost everything is not accessible: books, posters, audiovisual material, etc. Depending on their disability, this would determine how much they could access to them or not. For example, blind persons will not use the same audiovisual material as other teachers, so it must be purchased. Although legislation says that it is forbidden to deny the purchase, special materials are not always ready when the teacher with a disability asks for them.

In turn, in the e-learning platform there must be accessible contents, not only for the students with disabilities, but also for (and made by) teachers disabilities.

3.5 Infrastructures Not Adapted

Due to legislation that forced the schools to be adapted [8] to students with physical disabilities, schools are usually also accessible for physically disabled teachers: stairs and ramps, toilets, doors, emergency exits... This is not always true for teachers with other disabilities, as they are usually taught in special schools (for the blindness, for the deafness, etc.). For example, it is very common in Spain to alert of the start or the end of the class time with an alarm sound: This is completely useless for a deaf teacher.

This concludes with the need of the e-learning platform to be accessible by itself, no matter which course is held on it. This means that e-Learning platforms should offer their functionality in an accessible way. In particular, those related to the administration facilities.

3.6 Mobility to the School

Many teachers do not live close to their schools, and often they have to drive or be driven to the school. Although there are grants for these issues, teachers must dedicate part of their salary in special transport.

If we translate this problem to the web, it happens that people with disabilities must buy special equipment to manage the computer, both input and output devices. The e-learning platform should be aware of this device independency.

4 Summary of Requirements

According to the previous facts, I have deduced several requirements the e-learning platform should afford to really usable for teachers, no matter if they have a disability or not. Some of these requirements are difficult to be achieved in face to face or off line scenarios for teachers with disabilities. However, translating them into the on-line world, e-Learning platforms provide a support for disabled teachers to carry on their tasks.

4.1 It Must Help Teachers to Guarantee the Control, Surveillance, Assistance and Attention of the Students

Works done by [9] have shown that it is possible to track all the activity of the students all the time, so it is possible for the teacher to know what are the students doing in the platform. In this way, all learners can be monitored and follow-up from the administration interface of the platform.

4.2 It Must Offer Different Types of Activities

We must be aware that e-learning platforms must be different for 3-years-old-children than adult education, as they learn different from each other [10]. So the platform must be opened to the inclusion of new modules or educational games, in order to accomplish such a wide audience. Evidently, teachers with and without disabilities must be a principal actor in deciding which one he or she would use into his or her pedagogical program.

The only way to make this possible is to have an open source platform, where teachers and developers could design and develop his or her own modules.

4.3 It Must Help in Personalizing the Effort for Each Student

Works done by [11] are concluding in using artificial intelligence methods based on the modeling of each user needs and preferences. This enables the teacher to survey a lot of different students, but it is the platform which guides them trough the best pedagogical way for each user.

4.4 It Has to Improve the Communication among the Others Teachers

E-learning platforms should not only be the interface between the teacher and the student, but also a place where different teachers coordinate their efforts, support their jobs and exchange experiences. Forums, blogs, wikis and other communication tools are required in order to decrease the isolation feeling the teacher can have in his or her job.

4.5 It Hast to Improve the Communication with the Administration

As we have seen before, many teachers complain that they are not valued by the administration, so new communication channels must be established: human resources departments are creating intranets with services for the teachers, like news, education, bulletin boards… The e-learning platform could and should be used also by the administrative personnel. This implies that different roles should be afforded by the platform: course administration, teacher, tutor, secretary, principal, etc.

4.6 It Must Generate Confidence in Himself/Herself as a Good Teacher

This could be the most difficult point to accomplish. Feeling good depends on each person, but I will use the definition of health according to the World Health

Organization [12], health is the state of complete physical, mental and social well-being. Although teaching is not a physical profession, many teachers at vocational schools often feel overburdened and emotionally exhausted; the main reason for early retirement of teachers at vocational schools is mental ill health [13]

Thus the e-learning platform should focus on ergonomics, usability and social relationships.

4.7 Possibility of Teaching Expelled Students by an Anonymous Teacher

This point can be achieved by recommending systems, as the one developed by [14]. In this way, the student is guided trough the platform activities under the surveillance of the tutor, but with no need of physical contact or communication. This feature can also be used for monitoring the rest of students, in a blended-learning program.

4.8 The E-Learning Platform Should Use Accessible Content

Spanish law for accessible web content [15] follows the rule UNE 139803:2004 [16], which is based on the WCAG 1.0 [17]. This law forces the web content generated and used by the administration must be (at least) Double-A. This implies HTML content that follow a group of guidelines to develop web based contents in an accessible way. As a result, anyone can claim to have accessible contents in an e-Learning platform, or at least, ask for the development of them following the WCAG guidelines.

Teachers, as content creators, should know and follow these guidelines, not only for disabled students that will follow the courses, but also for disabled teachers that would reuse it. Accessible authoring tools are also required to create accessible content.

4.9 It Has to Be Accessible Itself, No Matter Which Course Is Held

Spanish law [15] for accessible applications follows the rule UNE 139802:2003 [18] for software. This rule is similar to User Agent Accessibility Guidelines (UAAG). [19].

In this sense, very little research has been carried out about the platforms as applications [20], but an analysis done on an open source standard-based platform showed that the administration interface of e-Learning platforms is less accessible that the learner's part.

4.10 The Platform Should Be Aware of Device Independency

Spanish law [15] for accessible applications follows the rule UNE 139801:2003 [21] for hardware; this reflects on the e-learning platform requirements that it should be installed on accessible hardware; and to use it, we should use any web device, no matter of its features.

Beside Composite Capabilities/Preference Profiles [22] recommendations should be followed; IMS Accessibility for Learner Information Package [23] and ISO Personal Needs and Preferences [24], which allow the user register users' accessibility preferences, should be implemented in the e-learning platform.

5 Conclusions

Tele-working has become an opportunity to persons with disabilities [25]. Despite this, usability and accessibility in e-Learning platforms and content has usually

Table 3. Summary of requirements

Requirement (problem addressed)	Explanation
Help teachers to guarantee the control, surveillance, assistance and attention of the students	Continuous tracking of the learners' activity, so teachers can know what the students are doing in the platform
Generate confidence and well-being	According to the World Health Organization, health is the state of complete physical, mental and social well-being. Thus, the eLearning platform should follow ergonomics and usability guidelines.
No face-to-face teaching	Student can be guided trough the platform activities under the surveillance of the tutor, but with no need of physical contact or communication.
Automatic validation of contributions against WCAG checkpoints	W3C Web Content Accessibility Guidelines should be followed not only when the contents are created, but when learners and tutors contribute to the platform (e.g. posting forum messages, r adding comments in blogs or entries in wikis).
The administration space of the eLearning platform has to be accessible	User Agent Accessibility Guidelines should be followed in both the front-end and the back-end of the platform functionalities.
Offer different types of activities	The platform must provide different modules so teachers select the appropriate one to make students work a particular learning goal and considering the learners' needs and preferences.
Personalization the effort for each student	Artificial intelligence methods based on user modeling enables teachers in tutoring different students, as the platform guides them trough the individual best way.
The platform should be available on the Internet	In this way, teachers can access it from any place, and do not need to move to a specific building.
The platform should be device independent	eLearning platforms should be accessed from any type of device, which may have some assistive technology on it.
Improve the communication among the others teachers and the administration	eLearning platforms should be a place where teachers coordinate efforts, support and exchange experiences. Collaboration facilities are required to promote social relationships.

focused on the learners' experience and how to improve the learning efficiency. However, very little studies analyze the accessibility and usability issues from the teachers' point of view [26], especially when the teachers have some disabilities. In this paper, I have analyzed which problems teachers face and how to solve them when implementing or choosing an e-learning platform.

6 Future Work

This paper is the first part of a wider study on usability and accessibility. Next, I will perform the following experiment with the data gathered. I will work with dotLRN e-learning platform, as it is flexible, open and standard-based architecture [27], so it will allow me to configure it in different ways to test the requirements found in the first part of my research works.

Once I have configured different dotLRN platforms instances, teachers with disabilities will test it. I have defined the following steps:

1. I will define several tasks that the teachers should perform within a course.
2. I will group teachers depending on their disability and type of teaching they do. A control group will be made to guarantee the validity, integrity and independence of the results.
3. I will gather information from their activity at the platform, both automatically (logs, statistics…) and personally, interviewing the teachers about efficiency and user satisfaction.
4. Data will be analyzed, related, summarized and published in future works.

So, next step in my research is a survey amongst teachers with disabilities to gather more problems; then to evaluate these proposed requirements and check if they solve the problems found. I have contacted different trade unions and associations in order to get a wide sample of teachers with any kind of disabilities.

References

1. Spanish Institute for Statistics (INE): Persons with a disability and their relationship with job market. Based on the Active Population Quest (April-June 2002)
2. Anderson, R., Karp, J.M., Keller, C.E.: Enhancing Diversity. Educators with Disabilities. Gallaudet University Press (1998)
3. Training and Development Agency for Schools, http://www.tda.gov.uk/about/mediarelations/2008/240608.aspx
4. Wallace, J., Hoffman, S.M.: Troubling Bodies: Disabled Teachers in Educational Organizations. In: Proceeding of the CSSE Annual Conference (2007)
5. Socialist party accuses Education Department to discriminate a blind teacher aspirant. (El PSOE acusa a Educación de rechazar a un opositor por ser invidente.), http://www.elmundo.es/elmundo/2007/07/17/madrid/1184692887.html

6. Sánchez, L., Yepes, R.: Teachers job health conditions in Madrid (Las condiciones de salud laboral del colectivo de trabajadores de la enseñanza de la Comunidad de Madrid) FETE-UGT (2008)
7. ANPE: Violence at schools survey (2005)
8. Minimum accessibility requirements for buildings. (Real Decreto 556/89, de 19 de Mayo, por el que se arbitran medidas mínimas sobre accesibilidad en los edificios)
9. Santos, O.C., Boticario, J.G.: Recommendation Strategies for Promoting eLearning Performance Factors for All. In: Proceedings of the 6th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems. AAAI, Menlo Park (2008)
10. Knowles, M.S., Holton III, E.F., Swanson, R.A.: The Adult Learner. The Definitive Classic in Adult Education and Human Resource Development. Elsevier, Amsterdam (2005)
11. Santos, O.C., Boticario, J.G.: Supporting learners in an inclusive way with standard-based user modelling techniques. Learning Technology. IEEE Computer Society Technical Committee on Learning Technology 9(3 & 4) (2007)
12. WHO: Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June and entered into force on 7 April 1948 (1946)
13. Weber, A., Lederer, P.: Morbidity and early retirement among teachers at vocational schools. Versicherungsmedizin 58(1), 22–28 (2006)
14. Santos, O.C., Raffenne, E., Granado, J., Boticario, J.G.: Dynamic support in OpenACS/dotLRN: Technological infrastructure for providing dynamic recommendations for all in open and standard-based LMS. In: Proceedings of the International Conference and Workshops on Community based environments (2008)
15. Disabled persons rights for Technology, Products and Services related to the Information Society and Mass Media (Real Decreto 1494/2007, de 12 de noviembre, por el que se aprueba el Reglamento sobre las Condiciones Básicas para el Acceso de las Personas con Discapacidad a las Tecnologías, Productos y Servicios Relacionados con la Sociedad de la Información y Medios de Comunicación Social)
16. AENOR: UNE 139803. Computers for persons with disabilities. Web content requirements (Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad para contenidos en la Web) (2004)
17. W3C: WCAG. Web Content Accessibility Guidelines (1999)
18. AENOR: UNE 139802. Computers for persons with disabilities. Software requirements (Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Software) (2003)
19. W3C: UAAG. User Agent Accessibility Guidelines 1.0 (2002)
20. Revilla, O.: Can dotLRN be administered by all professors? In: Proceedings of OpenACS &.LRN Conference. International Conference and Workshops on Community Based Environments (2008)
21. AENOR: UNE 139801. Computers for persons with disabilities. Hardware requirements (Aplicaciones informáticas para personas con discapacidad. Requisitos de accesibilidad al ordenador. Hardware) (2003)
22. W3C: Composite Capabilities/Preference Profiles: Structure and Vocabularies 2.0 (2007)
23. IMS: IMS Learner Information Package Accessibility for LIP (2004)
24. ISO/IEC FDIS 24751-2. Information technology. Individualized adaptability and accessibility in e-learning, education and training. Part 2: "Access for all" personal needs and preferences for digital delivery (2008)

25. Viorreta, C.: How to incorporate teleworking in the enterprise: an option for disabled persons (Guía de incorporación del Teletrabajo en la empresa: Una opción para integrar personas con discapacidad) (2008)
26. Revilla, O.: Ten requirements for eLearning platforms to support disabled teachers. Learning Technology 10(3), 3–5 (2008)
27. Santos, O.C., Boticario, J.G., Raffenne, E., Pastor, R.: Why using dotLRN? UNED use cases. In: FLOSS (Free/Libre/Open Source Systems) International Conference (2007)

More Than Just a Game: Accessibility in Computer Games

Klaus Miesenberger[1], Roland Ossmann[1], Dominique Archambault[2], Gig Searle[3], and Andreas Holzinger[3]

[1] University of Linz, Institut Integriert Studieren
Altenbergerstraße 69, 4040 Linz, Austria
roland.ossmann@jku.at, klaus.miesenberger@jku.at
[2] Université Pierre et Marie Curie - Paris 6, Inova
9, quai Saint Bernard, Paris 5, France
dominique.archambault@upmc.fr
[3] Medical University Graz, A-8036 Graz, Austria
Institute for Medical Informatics, Statistics & Documentation (IMI)
Research Unit HCI4MED
gig.searle@meduni-graz.at,
andreas.holzinger@meduni-graz.at

Abstract. During the last decades, people with disabilities have gained access to Human-Computer Interfaces (HCI); with a resultant impact on their societal inclusion and participation possibilities, standard HCI must therefore be made with care to avoid a possible reduction in this accessibility. Games, considered as a field of research, could provide new interaction principles, which can be incorporated into the existing HCI Standards, thereby complimenting and expanding these standards positively. However, games also provide an interesting new potential for better access and for supporting people with disabilities. They can be used to acclimatize people, who have had little or no exposure to technology, to interaction with modern Information and Computer Technology (ICT). Some simulation games act as an interface between games playing and real life, where the end user, in the form of an avatar, can interact within modern communication systems. It is important to ensure that everyone has accessibility to this technology, regardless of abilities or age. This paper advocates pro-active "research in games accessibility" and provides some first considerations on establishing a) guidelines for accessible game development, b) Active Game Accessibility (AGA) development framework to support game developers and Assistive Technology (AT) providers and c) a collection of games or game scenario examples ("code pattern collection") as a reference for game and AT developers.

Keywords: Active games accessibility, guidelines, game-like interfaces, Assistive Technology.

1 Introduction and Motivation for Research

Why games accessibility? Are there no more important aspects of helping people with disabilities in the information society? So far, games accessibility has been

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 247–260, 2008.

recognized as an area of minor importance. Besides games to support special pedagogical and educational issues there seem to be far more "serious" problems for people with disabilities to focus on [1].

Moreover, mainstream games, although being a very attractive Information and Computer Technology (ICT) market, have not been considered as a subject for research and science, in particular for the Human-Computer Interaction (HCI) field, for a long time. Games have developed in parallel to the standard HCI market, leading to significantly different interface principles based on individual engines and hardware. This situation has been changing significantly during the last few years and games become more and more subject both to mainstream HCI and accessibility.

2 Background and Related Work

2.1 Learning with Games – Play as a Basic Stimulus for Learning

One of the fundamental problems of acquiring knowledge successfully, is how to maintain the motivation necessary to continue the learning process [2], [3]. While it is axiomatic that small children do not essentially differentiate between play and learning, play and work, fantasy and reality, this concept cannot be automatically applied to all age groups. However, the notion that playing can enhance our comprehension and inspire the creation of stimulating environments has intrigued classic educators, such as Dewey, Piaget and many others, for a long time. [4], [5].

The logical consequence to these considerations was to look at environments where play can also be used successfully to encourage learning among adults, facilitate training and provide mental and physical exercise for the elderly and disabled people.

As with the now famous case of the Short Message Service (SMS), which was originally intended to enable people with hearing difficulties to send messages via a mobile phone and is now used millions of times every day, the concept of Learning from the Extreme, states that any design, which takes peoples disabilities into consideration will be ergonomically sound for everyone.

It is our intention to implement this concept to improve the accessibility of games for learning to all social groups.

2.2 More Than Just a Game: A Science of Games

"The time has come to take computer games seriously, really seriously" [6] (USC GamePipe Laboratory, University of Southern California).

The mainstream commercial market for computer games and other multimedia products has shown an impressive growth over the last years. For instance, in 2002 costs for the development of a game could vary between 300,000 € for a game on wearable devices, to 30 millions for the biggest productions (involving nearly a hundred employees) [7], [8], [9]. Since 2002, games players' expectation of ever more impressive games has been a determining factor in the increase of development budgets and a more focused use of new technologies [10].

But academia and also R&D have started to focus on games to address unsolved "serious" problems such as problems in mathematics and science education [http://er.jsc.nasa.gov/seh/seh.html]. Game-based learning infrastructures and edutainment should help to find new and better didactical solutions. "Games stimulate certain areas within the brain that promote learning [11]." Games have entered areas such as job and security training and other areas where traditional, formal approaches to learning did not lead to corresponding results [12].

All this highlights the current trend of treating game interfaces as supplementary to standard interfaces. Game interfaces must be considered within HCI as of equal value with current standard interfaces.

Web 2.0 implementing game-like HCI concepts (e.g. avatar based interfaces, such as Second Life [13], [14] and emerging Non Classical Interfaces (e.g. virtual/augmented reality, embedded systems, pervasive computing) offer new possibilities for certain end user groups (see just some examples e.g. [15], [16], [17]) but also create challenges in retaining our successes, during the last decades, in enabling access to the mainstream HCI by people with disabilities and by elderly people.

Due to the maturity of elaborate game interfaces and the increasing power of standard computers, these concepts can be considered for standardization within HCI. As so called non-classical interfaces, game interfaces go beyond the traditional WIMP (Windows/Menus/Icons/Pointers) desktop and also more the modern SILK (Speech/Image/Language/Knowledge) metaphor.

We also start to move away from the desktop [18]:

- From the typical, rigid office environment to personal information clouds on a mobile computer;
- Multimodality: a wide range of interaction devices – not just the mouse and keyboard;
- From the desktop to a diverse set of interactive objects and embedded interfaces;
- From personal to inter-personal, group and social interaction;
- From low level to higher level interaction (from applications to services);

These are only a few trends and facts, which outline the considerations making it necessary **to establish** "a science of games" [6], with the goal of implementing games and game-like interfaces of general importance for a growing number of applications and as a general part in the design of Human-Computer HCI [19].

2.3 Using Games to Increase Acceptability

As newspapers all over the western world have been reporting for some time, the baby boomers are going into pension. Although, thanks to healthy diets and medical care, most elderly people in the west are still able to look after themselves, there is a growing number of elderly people, wishing to remain in their own homes, who will eventually be dependent on outside assistance. This, coupled with the declining number of Home Care Nurses, make it imperative that these people become acclimatized to the technical assistance and monitoring devices being offered.

Mobility decreases with advanced age. For many elderly people, this means an equivalent increase in the difficulties attached to social interaction. For those not living in a supervised environment (Residential Homes, Geriatric centers etc.), meeting friends and visiting social centers requires organization, which, in turn, can be facilitated by easy to use communication methods. Thanks to the expanding accessibility of flat rate computer and telephone connections, this need no longer be prohibitively expensive.

However, the elderly have, apart from acceptance difficulties, the same difficulties as many others, of all ages, with light physical and mental disabilities, which could eventually lead to various degrees of social isolation. Although there is no doubt that many mobile devices, such as PDAs and mobile phones, have not only made day to day living more comfortable, they can also play a major role in insuring the safety of the elderly, and the disabled, by the use of discrete monitoring (vital signs) for example with a *stress/heart monitor with warning device.* People, who would otherwise need constant supervision in a home or hospital, can remain autonomous far longer; however, they must be willing to use these devices. Increasing acceptability must be based, therefore, not only on a thorough knowledge of the requirements of the target groups but also on their misgivings. HCI can improve facilitation of usage must also aim at investigating ways to increase motivation and improve acceptance, by analyzing the requirements of the target group and using the results to make the design more user-friendly. For example, the same technology which has made miniature digital hearing aids possible has also made other technical devices, such as mobile phones, so much smaller and more compact that the use of these devices has become difficult for the elderly and people with motor disabilities. Designers and developers need to understand their needs, which need not necessarily be just bigger brighter virtual keyboards and larger script. Their motivation is different, their frustration level is lower and they may have to overcome previous, negative experience.

We are of the opinion that the design and development of mobile applications for the elderly and people with restricted abilities must support the users to overcome their fears and enable them to accept technological aids and mobile devices without reservations. The design must then reflect this acceptance and not be the cause of new biases [20].

In order to raise the acceptance of the elderly and encourage them to try new technologies, both for their own convenience and also to prepare them for the future, games, tailored to their tastes and frailties, could be a vital stepping stone. The range of tastes and abilities in playing games varies enormously, so a certain level of adaptability must exist.

2.4 More Than Just a Game: The Importance of Games Accessibility for People with Disabilities

People with disabilities are certainly one of the groups who benefit most from the ICT revolution. The potential of ICT is first of all based on

- flexibility and adaptability of the HCI to better address the needs of a diversity of end users in varying situations (multi media, multimodal).

- universal application of a limited number of HCI concepts (WIMP,/SILK, GUI) in different areas; these interfaces provide a stability for the end user by providing standardized, basic interaction principles, which stay the same, or at least remain similar, in changing technical and application scenarios.

This significantly reduces the cognitive overload [21] and makes the standard HCI – once made accessible to, and learned by, people with disabilities – a universal tool for inclusion. By combining this limited and stable number of interface objects and interface actions, an almost unlimited number of applications, systems and services is opened. Similar to the genome, where only 4 bases form the unlimited space for the evolution, a limited number of interaction principles (WIMP/SILK) leads to its almost unlimited potential of application and also to this revolutionary inclusive power for people with disabilities.

Over the last decades the emerging field of Assistive Technology (AT) has made considerable progress in connecting to, and expanding, current interfaces, thereby opening a door for people with disabilities to access all areas where ICT is used. Situations in their daily lives, at school as well as at work or at home, in mobility, etc. are enriched due to this interface to the standard HCI. Getting used to and developing skills in handling ICT/AT is a key factor for successful inclusion and independent living in the information society [22].

The main groups of people addressed by these accessibility issues are:

- People who cannot use the ordinary graphical interface, because they are totally blind or because they have a severe visual impairment (sight rated <0.05) [23];
- People who cannot use or have limited access to ordinary input devices, such as keyboard, mouse, joystick or game pad due to limited hand dexterity;
- People with cognitive problems who need support to better understand the course of events and to react properly (e.g. symbol, text, speech and easy to understand support);
- People with hearing problems or deafness not able to accommodate to sound based interaction modalities;
- People with problems in reacting to a strict time setting of the game out of various functional, cognitive and also psychological problems.

This revolutionary potential of AT to connect to standard ICT/HCI for people with disabilities emphasizes that every change to the standard HCI requires corresponding attention of the accessibility field to

- keep the achieved level of access, as well as
- exploiting the potential of new HCI possibilities for enhanced access for different groups of people with disabilities.

Considerations, effort and trends in “moving away or behind the desktop” [18] might, on the one hand, offer new challenging possibilities for increased accessibility, in particular for groups which often have not been considered extensively (e.g. people with cognitive problems). On the other hand, these trends could provoke significant accessibility problems (e.g. in avatar based interfaces for screen reader end users).

Games can be seen as an important learning enabler, in particular for groups with severe learning difficulties or those suffering from cognitive disabilities.

New approaches towards therapeutic and educational games to develop psycho motor and cognitive development for people with disabilities can be derived from the use of computer games [24].

Elearning and edutainment increasingly implement didactic games [12], [25], [26] and this should be addressed as an accessibility issue of growing importance.

Progressively, games tend to enrich formal and informal learning processes and addresses especially incidental learning to a high extent [27], [3].

Why should this not also offer new possibilities to address the problems of people with disabilities in education, job and societal inclusion? Using games, edutainment or game-like interfaces should be seen as enabler to overcome such obstacles.

As indicated, HCI is beginning to use concepts and methods derived from games. Games and game-like interfaces are recognized as a means to implement education, training, general Human Computer Interaction (HCI) and web applications with usability and effectiveness. This requires a pro-active focus on accessibility to keep pace with the general level of accessibility achieved over the last decades in standard HCI. When standard HCI changes, as several trends show, accessibility also has to change.

Treating computer games as being of minor importance or not “serious” enough for people with disabilities should also be seen as discriminating. Games are often good training to become accustomed to the standard HCI and in particular with AT, for children as well as for adults after accidents, diseases or the elderly, who fear they it is too difficult for them [20]. For people with disabilities, AT skills are of increased importance as outlined above. Neglecting games loses a promising way of addressing this issue.

Computer games have become an important part of child and youth culture, and most children in developed countries have considerable experience of such games. Games are used by a growing part of the population, especially among young adults (on average 25 years old, including 40% of women) but the proportion of players is also growing in other age groups of the population. The exclusion of people with disabilities tends to strengthen social exclusion and contradict the positive inclusive trend of general ICT/HCI.

It also can be expected that R&D in game accessibility will contribute to an increased usability of mobile game playing and the emergence of non-classical interfaces for mobile computing. As a recent study demonstrates [28], there is a large overlap between Web Content Accessibility Guidelines and Mobile Web Best Practices [W3C 06] (about 60%) with no inconsistency. This is explained by the similarities between the difficulties experienced by Mobile Web end users and the difficulties that people with disabilities meet. There are numerous examples which outline how accessibility supports usability in general [22].

3 Active Games Accessibility (AGA)

Although it is hard to predict when and how such trends might enter or alter standard HCI, it is worth analyzing them as they offer new challenging opportunities for specific groups of people with disabilities. Due to a growing awareness for accessibility and in accordance with legal requirements [22], software is becoming increasingly accessible to the aged and people with disabilities, this can be further facilitated by discussions based on the principles, guidelines and tools developed for games and game-like interfaces (e.g. simulation software, charts, virtual/augmented reality [15])

Designing games that work for players with disabilities is quite a challenge: The accessibility of games is a more complex problem than software or web accessibility in general. One fundamental difference is that games tend to expect certain behavior or skills, which makes adaptation to specific needs more challenging. Another reason for this difficulty, which may seem banal but must be emphasized due to its importance, is that accessible games must still be games [29]. It is not only about supporting special pedagogical aspects for people with disabilities, it is about taking part in a societal phenomenon of growing importance.

The basic idea of the AGA is to embed support for accessibility, possibilities of adaptability and interfaces to AT into mainstream games.

Working together with leading games and AT development companies and based on their technologies, it is planned to develop an accessibility framework that will provide:

- Game Accessibility Guidelines
- Active Games Accessibility framework
- AGA Code Box and Documentation

In the following pages, we will give a very short introduction to these activities which will be addressed in an international co-operation of accessibility and game experts.

3.1 Game Accessibility Guidelines

The W3C Web Accessibility Initiative (WAI) [W3C 99] is an excellent model and a good basis to work towards game accessibility guidelines and to outline how games can be made accessible to the broadest possible range of end end users, including those with disabilities and the aged. Computer game interfaces rely on the interaction of the end user, applying specific interface principles and AT.

There are two research projects in the area of guidelines for games accessibility, which are of particular interest to our research. Project one is from IGDA, the International Games Developer Association. Inside IGDA is the Games Accessibility Special Interest Group (GA-SIG), who has published a whitepaper about games accessibility, and one chapter is about rules and hints for game developers.

The other ongoing project is by the Norwegian IT company MediaLT. MediaLT has developed a set of guidelines, which were the basis for further development of out guidelines. Furthermore MediaLT is partner in our project.

With the guidelines from MediaLT, the rules and recommendations from GA-SIG and our own ideas, we developed our own guidelines and published them as a web page. The decision was made to create a web page making the GL accessible to everyone who wants to bring in new ideas or to help improve the existing GL.

These guidelines have five main categories:

- level/progression
- input
- graphics
- sound
- installation and settings

The guidelines have, beside the rules itself, a categorization in three classes of priorities:

- Priority 1 – Must have
  These are absolutely necessary for the listed group of gamers. Otherwise the game is not accessible for them.
- Priority 2 – Should have
  These are a big help for the listed group of gamers. The game is accessible without these extras but with them, the game is easier to learn or the fun factor is higher.
- Priority 3 – May have
  These are help features for the listed group of gamers. The game is still accessible without these extras.

Furthermore there are four groups of disabilities: visual, auditory, mobility and cognitive disabilities. These disabilities are allocated to priorities, e.g. one rule can have priority 1 for visually impaired people and priority 3 for aurally impaired people.

3.2 Active Games Accessibility Framework

An expert AGA should provide a developing framework supporting:

- Game developers to enable accessibility, by providing the information needed for designing assistive interfaces and
- Assistive Technology providers to create assistive interfaces, enabling end users with disabilities to play the games.

This should also include a so called AGA Toolbox, a set of applications, which help to check games for accessibility and to support developers in designing, checking and validating towards accessibility.

The work on the AGA framework can follow a well established model of accessibility work. First considerations will involve an extensive study of established models of software accessibility such as developed by Microsoft [30], Apple [31], Gnome [32] or KDE [33] etc. When moving away from the standard HCI, there is a need to extend these software accessibility frameworks to include game-like interface principles.

This framework will be developed based on a end user-driven design approach and by using working demonstrators of newly-developed and existing games. The large involvement of game and AT companies should guarantee a practice-oriented approach of this research. Special games for target groups as well as mainstream games will be considered.

The framework will consist of specific interfaces; the games should implement to be compatible with AGA, and protocols for exchanging messages and/or events. Specifications will be developed and interfaces for the main languages will be implemented. Plug-ins for game development platforms (for instance Microsoft XNA, etc.) will be developed in order to enable game developers using this platform to create accessible games easily.

It must guarantee that:

- the game and scenarios can be adapted to the abilities of the end user (knowledge level, motor skills, perceptual features, cognitive skills);
- the device/platform (input-output signals and modalities of imaging the game content) must support the AT of end users with disabilities;
- AT and disability specialists can adapt the interface for certain end users or end user groups;

The assistive interfaces will be of various kinds:

- Generic accessibility interfaces: These are applications that communicate with any game implementing the AGA framework in order to deliver the information from the game to alternative devices and/or to manipulate the game from alternative devices. These applications can be specific to one disability as well as including the needs of several kinds of disabilities. They will be useful to handle simple games, in which the interaction will correspond to one or several identified models. They might have limited possibilities for adding contents to a specific game scripts.
- Specific game adaptations: These are applications that will be developed specifically for making one game accessible to one or several handicaps. It will correspond to the case of a game for which all interaction has to be adapted.

This categorization is not strict. Indeed, this framework will also support the possibility of developing an adaptation for a family of games, for instance in case of using a particular game engine (such as Quake).

In accordance with this, the involvement of AT experts and developers will also allow the adoption and further development of AT functionalities and interfaces, so that end users with disabilities can successfully play games in an accessibly, enjoyable and in some aspects also competitive manner. Game accessibility is therefore also treated as a challenge for AT.

3.3 AGA Code Box and Documentation

The AGA Code Box and Documentation will provide a set of examples of best practice and design/software patterns giving developers guidance on how to deal with certain games or game scenarios (e.g. shooting, car driving, climbing, moving around, etc ...) plus an extensive documentation including tutorials and references.

AGA will design a number of prototypes, demonstrators, games and all necessary material to support the implementation of the AGA framework in mainstream games (tutorials, guidelines, best practices, and code box with open-source code examples). The AGA framework will enable a large variety of interactive software to become accessible. This includes, of course, video games.

A professional design and documentation should allow reusability and better learning by examples. In particular, this should show how it is possible to make such games accessible as:

- action games: e.g. drive a vehicle in a race game, shooting
- simulation: e.g. flight/driving simulator
- strategic games: e.g. chess
- scrabble
- etc.

Disabilities should be treated as certain *contexts of use*. It should be stressed that we are not considering the accessibility in terms of disability but in terms of a set of alternative capabilities for interacting with the game, which would be equally valid for the elder player and help them acquire the confidence necessary to interact with new technology without anxiety. These alternative abilities can only be employed if a game supports accessibility and AT, and they should be supported within the framework

On the other hand, functional difficulties can also apply to people who are not qualified as being part of the target group. For instance, one could play a game with a mobile device while traveling back home after work. In a crowded bus, this player would not have access to the sound of the game, because of the noise in the close environment. Then, while walking back home after taking of the bus, the player would not have access to the screen, but to the audio via a headset.

4 Game-Like Interfaces

Games are more than *just games*. Games, as outlined, are seen as important new possibilities to enrich the standard HCI and the game-like interfaces, in particular, to enhance learning and teaching strategies. One of the most famous games making its way out of gaming into considerations for mainstream application is Second Life [13]. Second life is perhaps the most popular virtual world platform in use today, with an emphasis on social interaction and with enormous pedagogical potentials [34], [35].

Although Second Life utilizes a game based interface, it cannot be considered as a game due to the definition that a game usually includes a target to be reached, an enemy to defeat, or a lap time to beat. It is a simulation of real or virtual life situations, where a character (called avatar) represents the player in the game. Style, behavior and movements of the avatar can be defined. Most important are the communication with other "people" and the interaction with defined virtual objects. As the game became more popular, companies discovered Second Life as a good place for their own activities, such as training and advertisement.

Companies, governments and other organizations established their space within the game. This trend culminates with special sessions at conferences and job interviews in Second Life.

Avatar based interfaces, derived from game-playing, are first of all pure graphical interfaces. According to the established accessibility standards, there is a need for alternative text for such objects and according meta-information on their behavior. There is a need to provide direct access as an alternative to a hand-eye co-coordinated interaction. These are two important aspects derived from the W3C/WAI set of criteria, which would require corresponding interfaces in Second Life. First tests have demonstrated that such interfaces can be designed and, according to content authoring guidelines, might lead to accessible Second Life scenarios [36].

Despite the fact that the future of Second Life can not be predicted, it offers an interesting test bed for next generation interfaces, which go beyond the standard desktop.

In the AGA framework, it is used to test how the virtualization of real life situations (e.g. opening a public office in second life) could be put in place in an accessible manner.

Such activities are good examples of the fact that games accessibility is not just dealing with leisure time activities, it has extended to serious topics of life. If a governmental body opens a virtual office in Second Life, it is at least subject to web accessibility regulations and demands that games accessibility is addressed seriously.

5 Conclusion and Future Outlook

We have outlined that game accessibility is of growing importance since principles and techniques deriving from game interfaces should enter the standard HCI. In addition, games offer new possibilities to support people with disabilities. Due to this, game accessibility, which so far was not considered under the accessibility work, should be proactively addressed to smooth the potential transition into new interface paradigms.

We have presented some considerations on game accessibility work, comprising of work on a) guidelines, b) AGA supporting game and AT developers and c) code patterns, documentation and examples of accessible games or game scenarios. The work so far demonstrated the challenges and difficulties in games accessibility but also underlined its potential for accessibility and increased usability. More efforts are needed and increased research and development efforts is necessary, in order to reach the goal of game accessibility.

Further research also needs to be made as to the type of games most likely to appeal to older people, educational, creative, mind jogging or practical – i.e.: cooking [37] and which adaptations are necessary to encourage them to experiment and play in order to familiarize themselves with the wide palate of technology on offer.

By taking their physical and mental difficulties into account, we can increase accessibility for everyone and improve acceptability among a wider range of end users.

Acknowledgements

A part of this work (Medical University of Graz) has been co-funded by the European Commission under the sixth framework programme in the IST research priority, contract number 045056 (EMERGE, www.emerge-project.eu).

References

1. Velleman, E., van Tol, R., Huiberts, S., Verwey, H.: 3D shooting games, multimodal games, sound games and more working examples of the future of games for the blind. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 257–263. Springer, Heidelberg (2004)
2. Holzinger, A.: Computer-aided Mathematics Instruction with Mathematica 3.0. Mathematica in Education and Research 6(4), 37–40 (1997)
3. Holzinger, A., Pichler, A., Maurer, H.: Multi Media e-Learning Software TRIANGLE Case-Study: Experimental Results and Lessons Learned. Journal of Universal Science and Technology of Learning 0(0), 61–92 (2006), http://www.justl.org/justl_0_0/multi_media_elearning_software/justl_0_0_0061_0092_holzinger.pdf
4. Dewey, J.: Democracy and Education. An introduction to the philosophy of education (Reprint 1997). Free Press, Rockland (1916)
5. Piaget, J.: Play, dreams, and imitation in childhood. Norton, New York (1951)
6. Zyda, M.: Creating a science of games. Communications of the ACM 50(7), 26–29 (2007)
7. Natkin, S., Prado, C.L., Gal, V., Vega, L.: Quelques aspects de l'économie du jeu vidéo, Actes des journées d'étude: Internet, jeu et socialisation (2002)
8. Fries, F.: Propositions pour développer l'industrie du jeu vidéo, Rapport a l'attention de M. Francis Mern Ministre de l'Économie, des Finances et de l'Industrie, et de Mme Nicole Fontaine, Ministre déléguée a l'Industrie (2003)
9. Natkin, S.: Video games and interactive media: A glimpse at new digital entertainment. AK Peters (2006)
10. Ossmann, R., Archambault, D., Miesenberger, K.: Accessibility issues in game-like interfaces. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 601–604. Springer, Heidelberg (2008)
11. Mayo, M.J.: Games for science and engineering education. Communications of the ACM 50(7), 30–35 (2007)
12. Chatham, R.E.: Games for training. Communications of the ACM 50(7), 36–43 (2007)
13. LindenLab: Second Life (last access: 2008-09-17), http://www.secondlife.com
14. Eliens, A., Feldberg, F., Konijn, E., Compter, E.: Mashups in Second Life @ VU. Eurosis, Ghent (2007)
15. Nischelwitzer, A., Lenz, F.-J., Searle, G., Holzinger, A.: Some Aspects of the Development of Low-Cost Augmented Reality Learning Environments as examples for Future Interfaces in Technology Enhanced Learning. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 728–737. Springer, Heidelberg (2007)
16. Christian, J., Krieger, H., Holzinger, A., Behringer, R.: Virtual and mixed reality interface for e-training: examples of applications in ultralight / light sport aircraft maintenance. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 520–529. Springer, Heidelberg (2007)

17. Behringer, R., Christian, J., Holzinger, A., Wilkinson, S.: Some Usability Issues of Augmented and Mixed Reality for e-Health Applications in the Medical Domain HCI and Usability for Medicine and Health Care. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 255–266. Springer, Heidelberg (2007)
18. Moran, T.P., Zhai, S.: Beyond the Desktop Metaphor in Seven Dimensions. In: Kaptelinin, V., Czerwinski, M. (eds.) Designing Integrated Digital Work Environments: Beyond the Desktop Metaphor. MIT Press, Cambridge (2007)
19. Kellogg, W.A., Ellis, J., Thomas, J.C.: Towards Supple Enterprises: Learning from N64's Super Mario 64, Wii Bowling, and a Corporate Second Life. In: CHI 2007 Workshop Supple Interfaces: Designing and evaluating for richer human connections and experiences (2007), http://www.sics.se/~kia/supple/papers/Kelloggetal.pdf
20. Holzinger, A., Searle, G., Nischelwitzer, A.: On some Aspects of Improving Mobile Applications for the Elderly. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 923–932. Springer, Heidelberg (2007)
21. Workman, M., Lesser, M.F., Kim, J.: An exploratory study of cognitive load in diagnosing patient conditions. International Journal for Quality in Health Care 19, 127–133 (2007)
22. Miesenberger, K.: Best Practice in Design for All. In: Stephanidis, C. (ed.) The Universal Access Handbook. Lawrence Erlbaum, Mahwah (2008)
23. Archambault, D., Buaud, A., Lerebourg, S., Olivier, D.: Adapting mainstream multimedia games for severely visually impaired children. Eurosis, Ghent (2004)
24. Hildén, A., Svensson, H.: Can All Young Disabled Children Play at the Computer. In: Miesenberger, K., Klaus, J., Zagler, W.L. (eds.) ICCHP 2002. LNCS, vol. 2398. Springer, Heidelberg (2002)
25. Ebner, M., Holzinger, A.: Successful Implementation of User-Centered Game Based Learning in Higher Education – an Example from Civil Engineering. Computers & Education 49(3), 873–890 (2007)
26. Ebner, M., Holzinger, A., Zechner, J.: Spielend lernen - auch auf der Hochschule möglich? In: Jandl, M., Pauschenwein, J. (eds.) E-Learning Multimediales Lernen, Nausner&Nausner, pp. 88–96 (2004)
27. Holzinger, A., Pichler, A., Almer, W., Maurer, H.: TRIANGLE: A Multi-Media test-bed for examining incidental learning, motivation and the Tamagotchi-Effect within a Game-Show like Computer Based Learning Module Educational Multimedia, Hypermedia and Telecommunication 2001, Association for the Advancement of Computing in Education, Charlottesville, VA, pp. 766–771 (2001)
28. Hazaël-Massieux, D.: The Mobile Web: for everyone and everywhere? Accessible online services, a benefit for all, First European e-Accessibility Forum (2007)
29. Archambault, D., Olivier, D., Svensson, H.: Computer Games that work for visually impaired children. In: Stephanidis, C. (ed.) Proceedings of HCI International 2005 Conference (2005)
30. Microsoft: iAccessible Interface (last access: 2008-09-17), http://msdn2.microsoft.com/enus/library/accessibility.iaccessible.aspx
31. Apple: Accessibility (last access: 2008-09-17), http://www.apple.com/accessibility
32. Gnome: The Project GNOME. GNOME ACCESSIBILITY (last access: 2008-09-17), http://developer.gnome.org/projects/gap
33. KDE: KDE Accessibility (last access: 2008-09-17), http://accessibility.kde.org

34. Boulos, M.N.K., Hetherington, L., Wheeler, S.: Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. Health Information and Libraries Journal 24(4), 233–245 (2007)
35. Tim, R., Billy, H.: The use of Second Life for distance education. Journal of Computing Sciences in Colleges 23(6), 110–116 (2008)
36. Malcher, W.: Game accessibility: General problems and approaches and a particularly look at Second Life. Unpublished Bachelor thesis at the Institute Integriert Studieren, Linz (2008)
37. Terrenghi, L., Kranz, M., Holleis, P., Schmidt, A.: A cube to learn: a tangible user interface for the design of a learning appliance. Personal and Ubiquitous Computing 10(2-3), 153–158 (2006)

Visualizations at First Sight: Do Insights Require Training?

Michael Smuc[1], Eva Mayr[1], Tim Lammarsch[2], Alessio Bertone[2], Wolfgang Aigner[2], Hanna Risku[1], and Silvia Miksch[2]

[1] Research Center KnowComm, Department for Knowledge and Communication Management, Danube University Krems
[2] Department of Information and Knowledge Engineering, Danube University Krems
Dr. Karl Dorrek Str. 30, 3500 Krems, Austria
{michael.smuc,eva.mayr,tim.lammarsch,alessio.bertone, wolfgang.aigner,hanna.risku,silvia.miksch}@donau-uni.ac.at

Abstract. Understanding novel visualizations can be a challenge even for experienced users. During iterative usability engineering phases in the DisCō project, visualizations of time-oriented data are explored by domain experts and non-experts. The aim of our study is to analyze the generation of knowledge and understanding by means of visualizations without previous user training. Focusing on applicability in various business domains for personnel planning and time scheduling, we tested mockups of visualizations with a method based on user-reported insights. Results show almost identical behavior of domain experts and non-experts when generating insights into the data from scratch. In the course of working with a visualization, an interchange of insights into the visualization and insights into the data was found.

Keywords: Usability engineering, insight method, Visual Analytics, time-oriented data, participatory usability design.

1 Introduction

User studies for the evaluation of complex visualizations with non-experts often start with intense training sessions (e.g., [1][2][3][4]). In this context training focuses not only on the usage of tools but also on the acquisition of expert and domain knowledge [5][6]. For various scientific domains, such as bioinformatics [7], meteorology [8], and mathematics [9], the demand to integrate huge amounts of expert knowledge to obtain deeper insight and understanding of data visualizations seems evident. But this kind of knowledge-based training is not necessarily needed to this extent for other domains. In the business domains of personnel planning and workforce scheduling, complexity arises mainly from the interference of irregular and highly entangled time structures and huge multivariate datasets.

Although most of the central variables sound familiar from experience in daily work life and are comprehensible for most of us (e.g., volume of sales, occupancy rates), there are many differences between experts and novices in the interpretation of data. Experienced personnel planning consultants are time-structure-experts with deep

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 261–280, 2008.

knowledge and cognitive scripts for data checking and aggregation, a strong background about conventions, policies, and procedures about customer's business and elaborated mental models about economy. All these attributes are necessary for rich data interpretation, problem solving, and the generation of forecasting models. If structures of time are suitably put into pictures, many valuable insights into data can be gathered during explorative analysis.

One decisive point for personnel planning consultants is that results and data insights have to be communicated in most cases to semi- or non-experts for time structures. Usually time for additional explanations about the inner logic of the visualizations is limited during presentations. Although all of us are confronted with various time structures in daily life, such as business days, weeks, or months, the question raises, how fast and accurate can novel visualizations of time-oriented data be comprehended? How intuitive are these new visualizations for non-experts as well as for experts of "temporal analytics" [10]? In this paper we want to address these questions.

This paper is structured as follows: At first we want to give some insights into the challenges of depicting time-oriented data. Because the users' understanding is a crucial factor in the field of data visualization, we conducted a case study that will be described in the subsequent sections.

2 Visualizing Time – A Challenge for Visual Analytics

During the last decade capabilities to both generate and collect data and information have seen an explosive growth. Advances in scientific and business data collection (e.g., from remote sensors or from retail and production devices) as well as in data storage technology enabled us to store high volumes of data and information. However, they clearly overwhelm the traditional manual methods of explorative data analysis. The generation of new methods and tools for more complex explorative data analyses are the subject of the emerging field of "Visual Analytics" [11]. Its basic idea is the integration of the outstanding capabilities of humans in terms of visual information exploration and the enormous processing power of computers to form a powerful knowledge discovery environment.

Exploring trends, patterns, and relationships are particularly important tasks when dealing with time-oriented data and information. Time-oriented data are highly complex in their structure: there are different granularities (e.g., minutes, hours), different forms of divisions (e.g., 60 minutes per hour but 24 hours per day), different calendar systems (e.g., Gregorian or Business calendars), and irregularities like leap seconds or leap years. Moreover, time contains natural, but also social cycles and reoccurrences; for example seasons or holidays.

Within the DisCō[1] project we aim to develop novel Visual Analytics methods to visually as well as computationally analyse multivariate, time-oriented data. The novel visualizations are developed as a plugin for TIS[2] (described in [12][13]). In the

[1] From Latin "discō" (inf. discere), which means "I learn".

[2] [TIS] (Time Intelligence Solutions ®) is a flexible Business Intelligence Software developed by XIMES GmbH. It focuses on time related data with the aim to support personnel planning, controlling, and forecasting. Thereby, TIS facilitates the fast development of company specific solutions and the exchange of methods and ideas in this field.

DisCō project, the development of visualizations is accompanied by usability studies to support the human reasoning process and ensure intuitive human-computer-interaction.

3 Usability Studies

The usability study design is characterised by an early focus on users and their tasks. The different stages of usability evaluation can be described in three phases, namely task and user analysis, iterative process of User Driven Design, and Usability Testing (cp. Table 1).

Table 1. Usability evaluation in three phases

Phase	Methods	Questions & Topics
1. Task and User Analysis	10 qualitative interviews (users and software developers)	o Who are the potential users of the system (e.g. educational background)? What kind of expectations do they have? o What are the typical goals and information needs of users? o What context do users work in (infrastructure, tools, …)?
2. Iterative Process of User Driven Design	o Mockup testing, think-aloud, interviews o Usability inspection and heuristic evaluation o Focus groups in order to provide additional information about preferences	o Discussion and optimization of visualizations with potential users, analysis of insight patterns o Identification of guideline violations; in-depth testing o Participatory development of interaction design
3. Usability Testing	Summative testing procedures	Validate usability design

The present study relates to the iterative design process (phase 2), where the first mockups of visualizations were tested. In order to present the context of this study, we want to give a brief overview of some results of the already completed user and task analysis (see table 1, phase 1). Further, some specific peculiarities will be discussed with regard to existing findings in literature.

Who are the users of the visualization tools? The participants described their work as business consultants, human resources planners, and controllers. Their tasks are to plan and forecast personnel requirements and to evaluate organisational interventions in this field. In addition, they analyse time-related personnel data for different groups or locations in conjunction with metadata and various management ratios. The users described the application scenarios of data analysis in different industries, like transport and logistics, service industries, retail, health care, and the public sector.

Users reported that they were often confronted with ill-defined problems; although customers are aware of having a problem, they are frequently unable to describe or even to name it. Hence, data analysis has to be mainly exploratory at first. Similar to the workflow in exploratory data analysis in scientific domains (see [14][15][16] and

[8]), the workflow in the field of human resources planning lacks linearity. It can rather be described as an iterative or circular process. During their data analysis process, users reported several cases where it was necessary to go back one or more steps and change, normalise, or filter data. On the data level, frequently reported cases are inconsistencies in data like missing values or implausible time shifts and missing consistency with other variables. These validation processes are supported by an extensive toolset and are reported as one of the most time consuming parts of the analysis. On the user level, new insights and deeper understanding of data and their relations, adapted frames of understanding [15][17], and newly developed hypotheses (derived from previous results) often generate the necessity for recalculation.

Testing statistical hypotheses is not of high priority in these domains. Due to narrow time boundaries, results usually have to be gathered within minutes or hours, data analyses that span over weeks or months rarely occur. Visualizations are not seen as a "by-product" of the analysis (as stated by [16]), but as a tool supporting analysis and presentation of results. Users favour visualizations which can be presented to their management board or CEOs with little additional effort of editing or with automatic support for creating new visualizations to illustrate their insights and findings. Therefore, one of the requirements for the software application is intuitive visualization of highly complex multivariate, time-oriented data – so that it can be understood without extensive training lessons.

User requirements, as described above, have to be satisfied. The aim of participatory design is to meet these requirements. To test the support of the human reasoning process, various novel methods have been developed over the last years. Classical benchmarking metrics, like efficiency and efficacy or time and error rate, turned out to be of limited use for of the improvement of information visualizations as they are typically task based and used in a highly standardised experimental setting. Therefore, tasks have to be compact and predefined. Definitive, unambiguous, and distinctive answers are forced through the experimental setting and time constraints leave little room for deeper elaboration of the findings [18]. Due to the exploratory nature of knowledge discovery in Visual Analytics, new paradigms [5][19] for testing "beyond time and errors" [20][21] were promoted to fill these explanatory gaps. One of these metrics is the qualitative and quantitative measurement of user reported insights [7], which will be described in the following sections.

3.1 Defining Insights

A review of literature on insights in visualizations from Psychology and Visual Analytics showed that no commonly accepted definition exists [1][7][22][23]. Hartmann [22] sees the problem of the term insight in using "specialized meanings for [an] ordinary word" (p.242). While in Cognitive Psychology insights are sometimes defined as a sudden realization of a problem's solution, as an aha-experience [23], in Visual Analytics insights are often defined as discoveries about data [7]. We are not able to solve this ongoing discussion, especially as the granularity of insights is unclear: is a new insight gained when a pattern is detected in visualizations, when scripts for data analysis are identified, or when a mental model about the data is completed? Therefore, we define insight only for the purpose of this study as *the generation of new knowledge by individuals out of visualization for data analysis*. We focus on the finest granularity of insight, a single observation of the visualization.

4 Insight Study

This study addresses the question of whether prior knowledge about a specific visualization is necessary to generate insights. Most studies implemented extensive training [1][3][4][7] prior to the generation of insights or invited participants who were already familiar with this visualization [7][8][9]. By selecting this procedure, researchers implicitly assumed that some knowledge about a specific visualization is necessary. But is this the case?

To shed some light on these questions we conducted a case study. For this purpose we carried out and analyzed interviews about two visualization-mockups.

4.1 Methods and Materials

To provide a better understanding of the applied visualizations, we would like at first to give a brief introduction into Cycleplot and Multiscale visualization techniques. Afterwards, the participants of the case study and the analysis procedures will be described in detail.

4.1.1 Cycleplots

One of the more prominent effects of the structure of time is the fact that special patterns emerge from granularities, like day of week or hour of day. For example, values from one Monday are closer to values from the next Monday than values from the next Friday. The next Monday also has a next Monday that is similar. Values from a Tuesday are of course similar to values from the next Tuesday. A pattern emerges that is seven days long. Patterns like this are called cycles. Another phenomenon that is often found in time-oriented datasets is a trend. A trend is a long-term effect that overlays short term patterns or cycles. If a cycle has very obvious fluctuations, it can be hard to detect a trend that is overlaid by it. Fig. 1 shows a dashed line depicting a cycle and a dotted line depicting a trend. The solid line shows what happens if this cycle and this trend overlap.

Simple attempts on visualizing time-oriented data aggregate to one granularity and show a certain range of data in a single curve. Improvements upon this visualization split the data and show several curves. E.g., seven curves are shown for the days of week, but only one week is shown in total and each curve shows one day straight from the beginning to the end. These visualizations only show a very general trend. However, it is possible that different trends exist for the several parts of a cycle. E.g., Mondays are getting quieter while Sundays are getting busier. This fact is hidden by the visualizations we have described so far.

Cycleplots have been introduced by Cleveland [24][25] as a means to detect multiple trends in substructures of time. When generating a Cycleplot, the user has to choose a granularity that is used to group the data into bins. The important point is that this granularity should not be the coarsest one but the one that contains the most prominent cycle. Inside each bin, the whole selected range is plotted, but filtered for the value of the granularity that corresponds to the bin. Finer granularities than the one used for the grouping are usually aggregated to an average value. Fig. 2 shows an example. The number of simultaneous assignments of a police department has been plotted over the course of a month. These assignments mostly result from emergency calls, so it is difficult to plan them in advance. Experts in shift planning have the task

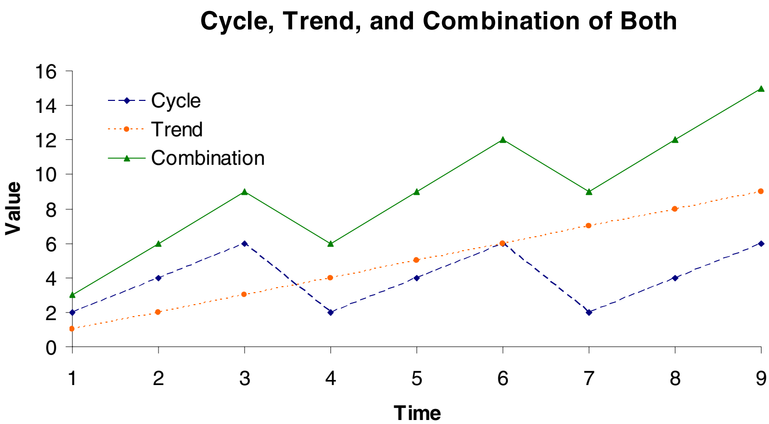


Fig. 1. Example for cycles and trends. The cycle is plotted as a dashed line, the trend is plotted as a dotted line, their sum is plotted as a solid line.

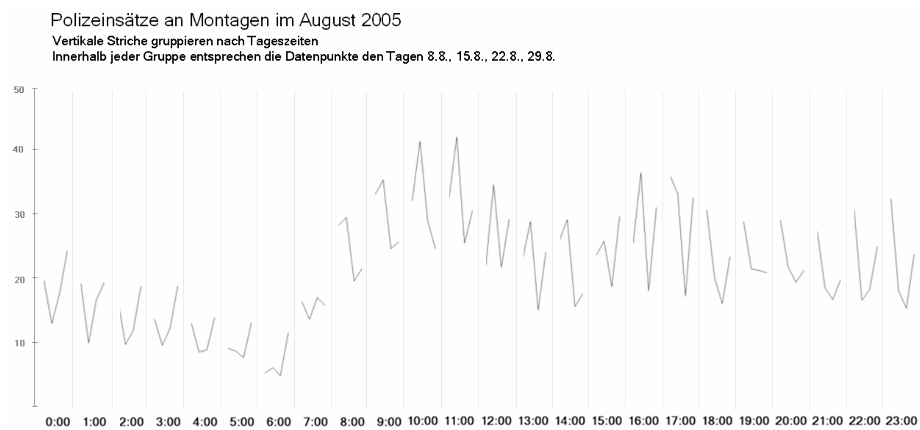


Fig. 2. Cycleplot example, used for mockup testing (text in German). The horizontal axis shows several bins grouped by time of day, each containing the Mondays over the course of one month. The vertical axis shows number of simultaneous police assignments (title translated "Police assignments on Mondays in August 2005").

to find out how many assignments are most likely to take place. In this example, the data has been filtered to show only assignments on Mondays. It is aggregated to use hours as the smallest granularity. Every hour of the day is shown in a separate bin. Note that the course of the line inside a bin does not reflect the minutes of that hour, but instead shows the average value of that hour of the day for several different Mondays. Trends are visible, but they are different for each hour. This visualization is well-suited for classification done by the user.

The Cycleplot performs very well in the detection of trends and some other aspects. However, for a data set the size of our example, it needs aggregation or filtering. When showing all available data points at a lower level than days, the optical clutter becomes too strong. We will now present a visualization that can show more details.

4.1.2 Multiscale Visualizations

To deal with a huge amount of time-oriented data to be displayed, several interaction mechanism are usually employed: *Scrolling* is used to show only a part of the dataset. There is no loss of detail by employing scrolling, but it is difficult to overlook the data or compare values that are temporally apart from each other. *Filtering* (e.g., showing only Mondays) is very similar to scrolling. *Aggregating* (e.g., calculating the mean over a certain amount of time) is effective for gaining overview, but it hides a great deal of information.

To reduce the need for scrolling or aggregation, it is necessary to show as much data as possible in a limited space. This can be done by encoding the position on the time axis using two dimensions. In these visualizations, it is necessary to find a different way to encode the value. One way is using different colors for different values. The position within a two-dimensional grid is determined by one time granularity for the position along the horizontal axis (e.g., day of week) and the value of another time granularity for the position on the vertical axis (e.g., week of month). A well-known example for this arrangement is a calendar sheet, with the day of the week on the horizontal axis and the week of month on the vertical axis. While a calendar plots the number of the day of month in different colors inside each square, a matrix-based visualization usually fills the whole block with a color that represents a data value. If the size of such a block is shrinked down to a single pixel, we also refer to them pixel-based visualizations.

Multiscale visualizations have been developed by Shimabukuro et al. [26] in a work mainly focused on the combination of time-oriented and space-oriented data where overview and detail are combined in one visualization. Although the work of Shimabukuro et al. has a strong focus on space-oriented data, the Multiscale visualization is solely for time-oriented data. A range of time (e.g., a month) is divided in a matrix-based visualization (e.g., a calendar sheet). In a parallel view, a second version of the visualization is shown. In that view, each block is further divided (e.g., in hours and minutes) using again a two-dimensional matrix. Therefore, each larger block contains another matrix-based visualization. The visualization makes strong use of the structure of time, allowing the combination of four granularities and their data content. Unfortunately, the original paper from Shimabukuro et al. does not stress the granularity aspect further as it is focused in a different direction. For us, it is important to stress that any combination of granularities can be used in any order.

The example in Fig. 3 shows the same dataset of police assignments as the Cycle-plot example (described above). However, the finer granularity consists of 5-minute-blocks. The fact that even at this time raster, a whole month can be displayed (without filtering for Mondays) shows how large datasets can be displayed in full detail using the Multiscale visualization (this visualization consists of 7803 separate data points). We have used a visual mapping of data value to color (hue) on the coarse level and a visual mapping of data value to gray scale on the fine level.

We have increased the flexibility of the Multiscale visualization by allowing the mapping of any granularity on any of the four nested axes. We can combine both views within one visualization by mapping the average value at the coarse level to

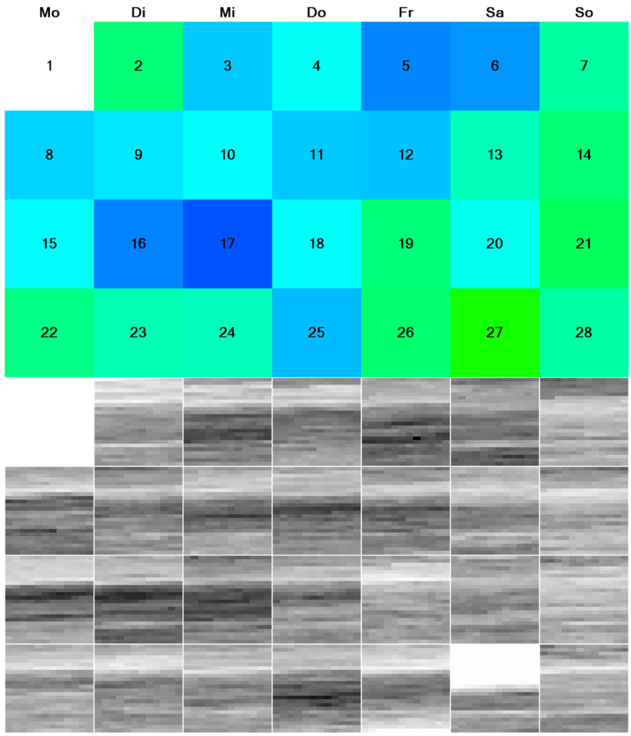


Fig. 3. Multiscale visualization example. Each block shows one day. Inside each block of the finer scale display (bottom), rows represent hours and each pixel in a row a 5-minute-interval.

hue and the average value of the fine level to gray scale (see Fig. 4). This variant consumes less space and combines overview and detail more integratedly, freeing the user from the need of switching the focus back and forth.

Both examples reveal the following phenomenon: Tuesday 16th and Wednesday the 17th have a hue that shows very high activity of the police. While the hue of Monday 15th suggests that there were less assignments on that day, a look at the value, showing the finer granularity, reveals that the high number of assignments already started on Monday morning. The quiet night from Sunday to Monday hides that fact if the data is aggregated to days, as in our initial example. The Multiscale visualization can reveal the true circumstances.

It is possible to make comparisons along any of the granularities when reading a Multiscale visualization. Inside each block, we can read along an hour or compare hours to each other. Between blocks, we can try and detect trends over the course of the week, or you can compare a day of the week to the same day one week later. In the example, Sundays are very similar too each other, but there are very different kinds of Saturdays.

These two novel visualization techniques – the Cycleplot and Multiscale visualization – were tested using the procedure described in the following section.

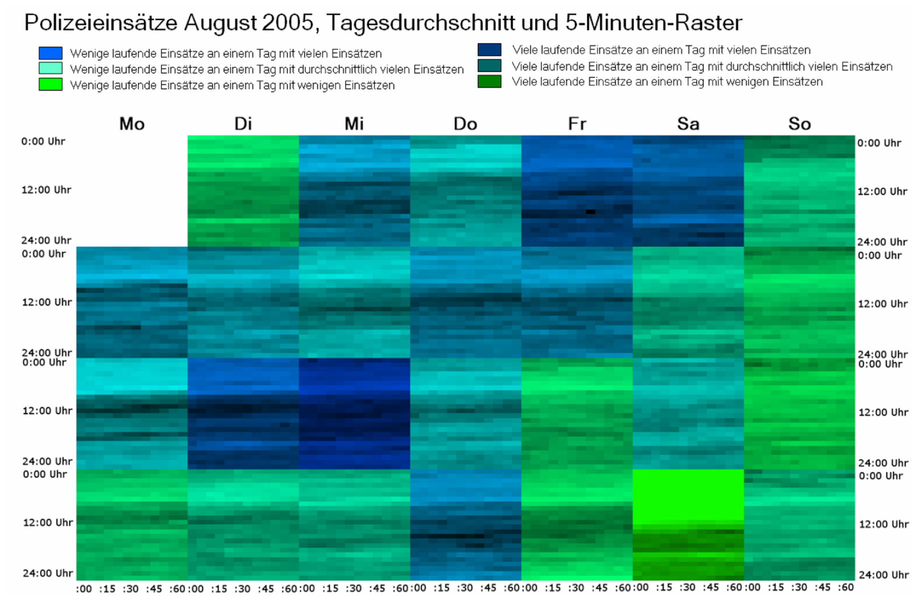


Fig. 4. Multiscale visualization example, used for mockup testing (text in German). Each block shows one day. Inside each block, rows represent hours and each pixel in a row a 5-minute-interval (title translated: "Police assignments in August 2005, daily mean and 5-minute-raster").

4.1.3 Procedure

To test the visualizations (Fig. 2 & 4), participants were asked about their insights into the visualizations using the think-aloud method. The two visualizations were presented as printed out mockups, without trainings and further explanations. When a participant was not able to generate any more insights into a visualization after repeated requests, a standardized short explanation of the visualization (cp. 4.1.1 and 4.1.2) was provided. After this explanation, we asked the participant again to generate insights and suggest further improvements. To reduce the interviewer's workload, audio protocols were recorded.

4.1.4 Analyses

Prior to the analyses we transcribed the audio protocols from all interviews. Then, we segmented the transcripts into smallest meaningful units (see [27]) and created a time-stamp for each unit from the audio protocols.

As first step in the analysis of the interviews we coded each transcript segment as an insight or not an insight: we coded an insight when new knowledge was generated by the participant. We then grouped the insights according to the knowledge generated: Segments referring to the same knowledge were coded as one insight with the time-stamp of the first reference.

For the categorization of insights we followed the procedure of North and Saraiya [7][18]. We coded insights with respect to the following characteristics:

- *Observation*: Occurrence of insights for each visualization
- *Time*: When was the insight generated (e.g. before or after explanation)?

- *Correctness*: Was the insight correct?
- *Category*: Which types of insights were observed?

A first analysis of the insights showed two main categories of insights: Insights into the visualization and insights into the data the visualization displays. At first sight, this categorization seems to be different to existing insight studies, but according to North [18] insight "capture[s] the intuitive notion of a visualization's purpose" (p. 6). This definition compasses both categories of insights used in our study.

Visualization insights can be further divided into insights what the visualization displays ("how-insights"), insights how the visualization has to be read ("meta-insights"), and insights how the visualization can be improved ("improvement-insights"; see Table 2 for examples of insights in each category).

In addition, we coded data insights according to the purpose of the two visualizations: Cycleplot enables the differentiation of trends and cycles. Therefore, we coded data insights into the Cycleplot as either trend or cycle. Multiscale displays both overview and detail in one integrated view. Therefore, we coded data insights into the Multiscale as either overview or detail (see Table 2).

Additionally, we coded whether prior knowledge was integrated by the participant for generating the insight and what knowledge was used (see Table 2 for an example). To visualize the uptake of prior knowledge or insights we built an uptake graph [28].

Table 2. Examples for insight categories

Visualization Insights	How-insight	"The more green the less assignments, the more blue the more assignments."
	Meta-insight	"Okay, first I'm looking at the days, if I can detect any patterns."
	Improvement-insight	"It would be good to be able to filter out one day."
Data Insights	Cycleplot: Cycle	"Starting in the morning it rises to a peak around 10, 11 am. Then it calms down at noon with a second peak around 4, 5 pm. and falls down again."
	Cycleplot: Trend	"The first Monday is high, descending on the second, and rising again on the third and forth."
	Multiscale: Overview	"Sundays are rather low, on average."
	Multiscale: Detail	"Especially at noon it's higher than before or after noon. It's always darkest then."
Integration of Prior Knowledge		"It decreases until 6 in the morning, to a minimum. I assume this is due to […], to my knowledge, change of shift."

4.1.5 Case Description

The recordings of three participants were analyzed for our study. Two of them were experts in time-scheduling and personnel planning/controlling (abbreviated as P.Y. or participant Y., P.X or participant X. in the following sections), with an experience of 5-10 years as consultants in various business domains (service industries, retail, production, health care and public sector). We interviewed also one expert (P.Z. or participant Z.) in data visualization with focus on dynamic organizational mapping and Social Network Analysis. He has 3 years experience in research and development in the academic field and consulting.

The cases were selected carefully on the basis of previous interviews in the phase of task and user analysis. The selected participants were those who showed the most demonstrative behavior when analyzing data. We are aware that three cases will not suffice to produce generalizable results. Still, according to Holzinger [29], at least three participants are necessary for think-aloud-analysis. Also, it has to be pointed out that these results are preliminary, since three other interviews are currently under transcription (see discussion). So far we have not found any indications in the additional interviews that argue against the results described below.

4.2 Results

Observation. On average, each participant generated 34.67 data insights overall. Comparing data insights for each visualization, the Multiscale (M = 19.67) produced more data insights than the Cycleplot (M = 15.00). Still, if we relate the number of data points for each visualization to the number of insights produced, relatively few insights were produced using the Multiscale. In contrast, the amount of insights into the Cycleplot was very high using this relative point of view.

Time. Participants had insights into the visualizations straightforward from the beginning; the first insights started between 13 and 79 seconds after instruction. The time course of insights for both visualizations per participant can be seen in Fig. 5.

Participant Z. generated most insights into both visualizations, but many of them after the explanations. If the amount of insights before the explanation is compared to the amount of insights afterwards, it can be seen that the expert participants X. and Y. generated most insights before the explanation (64 and 90 %), whereas participant Z. generated most insights after the explanation (62 %). Further analyses were conducted to see whether there are qualitative differences between participant Z.'s insights on the one hand and X. and Y.'s insights on the other hand.

Correctness. Overall, almost no incorrect insights were observed. It was audible sometimes that a wrong conception of the visualization was corrected, but the final insight was correct. For example, "I missed that: It's not one week, but four weeks."

Category. First, data and visualization insights for each visualization technique were differentiated. A similar amount of data and visualization insights was observed for each technique (cf. Table 3).

Most visualization insights focused on how the visualization can be improved (M = 5.33 for the Cycleplot; M = 5.67 for the Multiscale). The Multiscale needed more insights how the visualization works (M = 4.33) than the Cycleplot (M = 1.33). On the contrary, more meta-insights were gained into the Cycleplot (M = 2.67) than

Table 3. Data and visualization insights into the two visualizations (left two columns) and time of first insight (right)

	Multiscale Insights		Cycleplot Insights		First Insight	
	Visualization	Data	Visualization	Data	Visualization	Data
P.X.	9	9	5	5	81 sec.	203 sec.
P.Y.	7	9	9	6	15 sec.	53 sec.
P.Z.	18	13	13	8	74 sec.	112 sec.

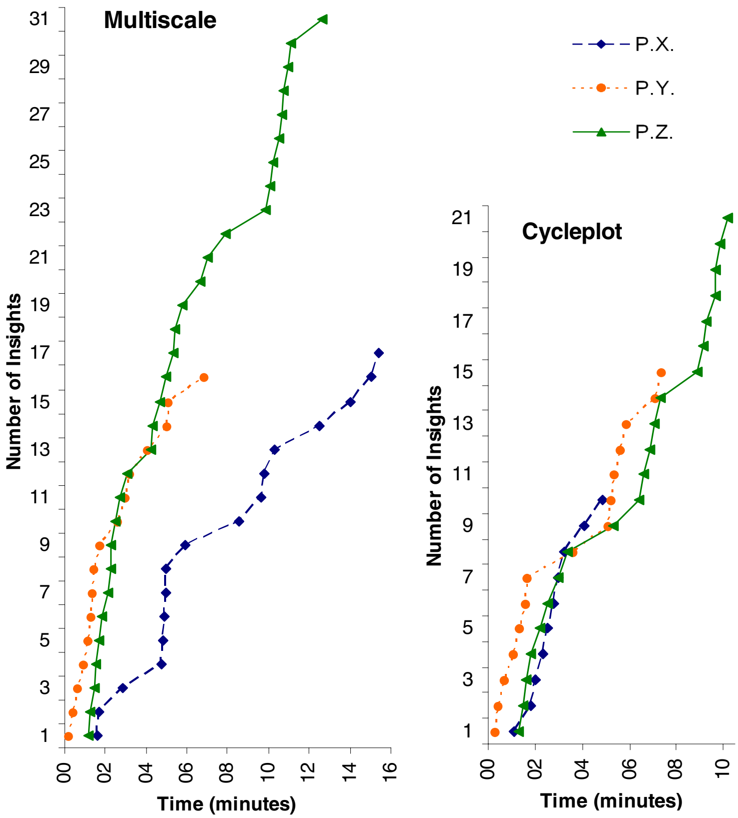


Fig. 5. Sequence of insights for Multiscale (left) and Cycleplot (right) per participant over time (P.X. = participant X.; P.Y. = participant Y.; P.Z. = participant Z.)

into the Multiscale (M = 1.33). Comparing the three participants, X. and Z. generated more improvement insights, whereas Y. focused more on how- and meta-insights.

Regarding data insights into the Cycleplot, only participant Z. generated data insights with respect to categories, cycles and trends. Participant X. generated only insights with respect to cycles, whereas participant Y. generated only insights with respect to trends and once remarked that “a daily cycle can also be seen”.

Data insights into the Multiscale were more evenly distributed: Every participant gained insights on an overview as well as on a detail level. Again, participant X. focused on one level: He generated eight overview-insights and only one detail-insight.

Integration of prior knowledge or prior insights. The most important observation – though intuitively logical – is that insights into the visualization always precede insights into data. On average, the first insight into the visualization was after 57 seconds, into the data after 122 seconds (cp. Table 3, right column). As can be seen in the following uptake graphs (see Fig. 6 and Fig. 7), many data insights build upon a preceding insight, how the visualization works.

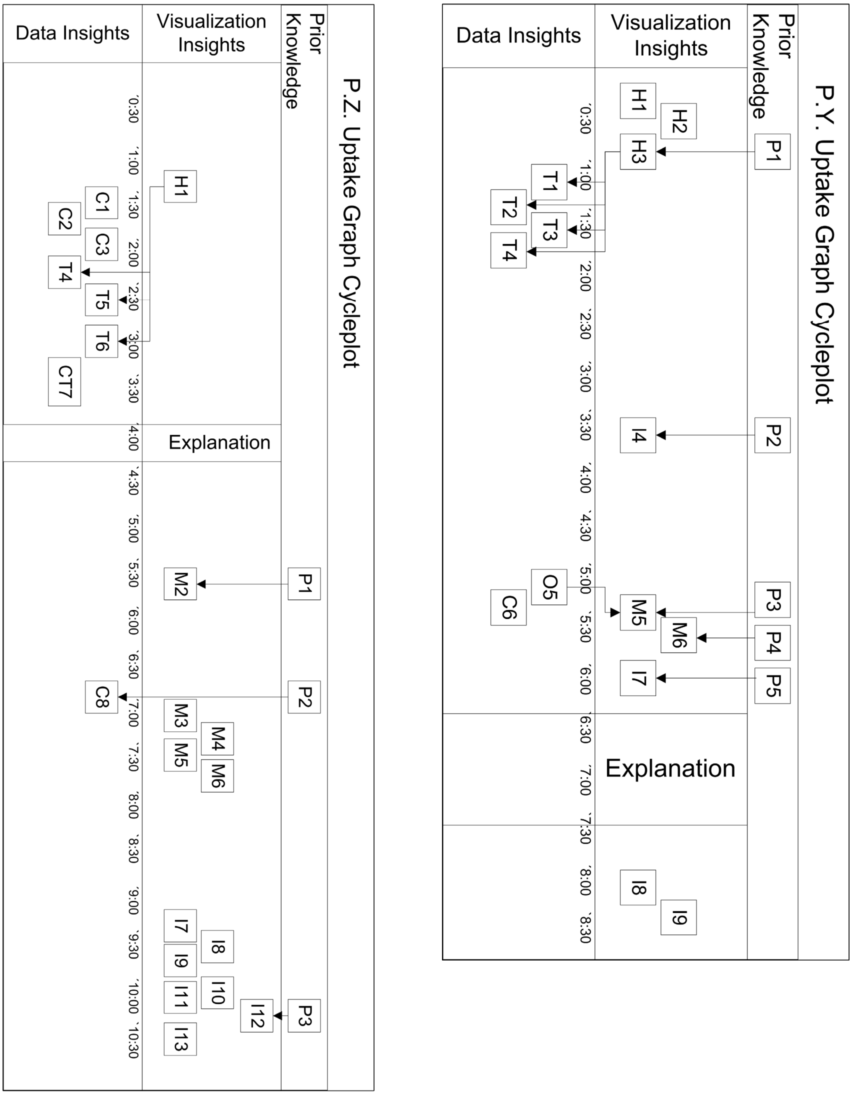


Fig. 6. Uptake graph for participants Z. (left) and Z. (right) for the Cycleplot.

This is very clear for participants Y. and Z., but participant X. seemingly did not verbalize every visualization insight during think-aloud procedure – many of his insights require other insights that were not explicated and therefore have to be inferred by the researcher. We therefore focus in the following paragraphs on participants Y. and Z., who reported their insights more accurately and completely. We are going to analyze their uptake of prior insights and prior knowledge in the Cycleplot first and then present their insights into the Multiscale.

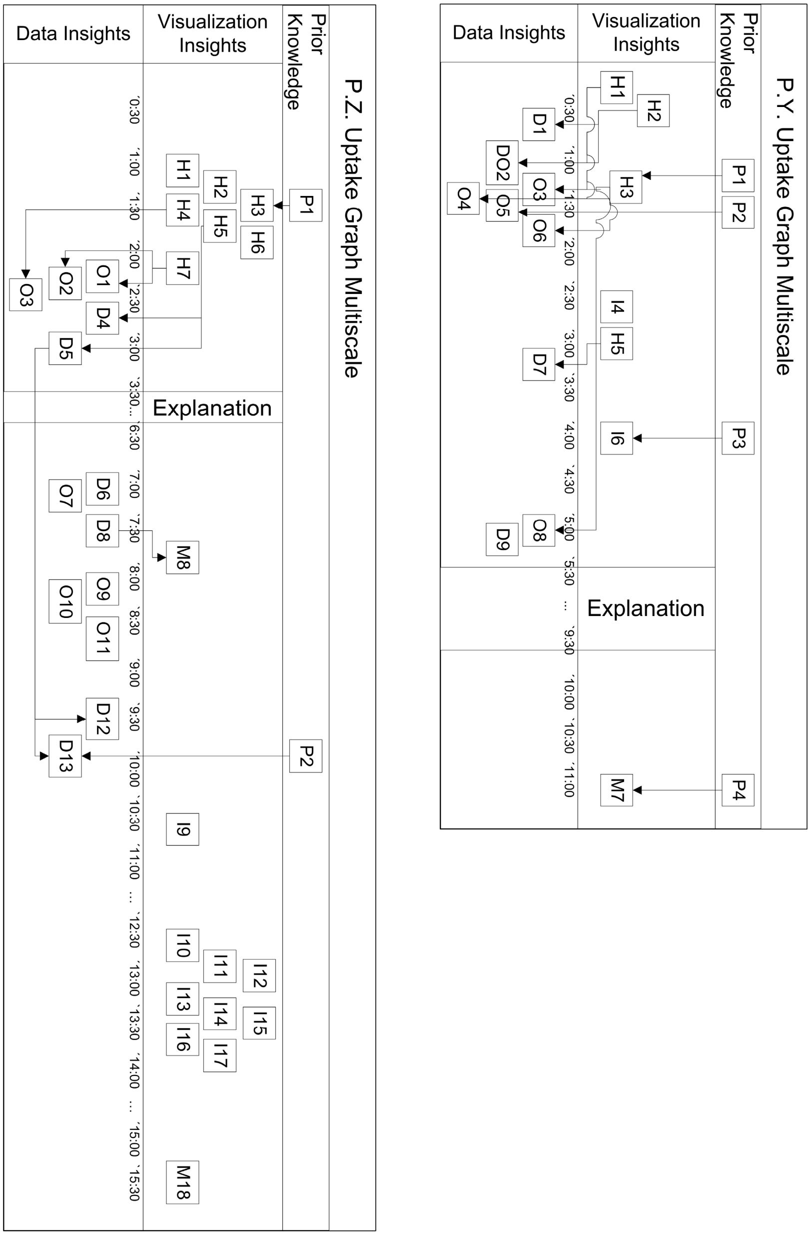


Fig. 7. Uptake graph for participants Y. (upper) and Z. (lower) for the Multiscale

Fig. 6 shows the uptake graph of both participants for the Cycleplot (see Fig. 2): In uptake graphs, dependencies of events (here insights) on two or more different levels can be visualized in chronological order. The right column displays the integration of prior knowledge P. Visualization insights are viewed as H (how-insight), M (meta-insight), and I (improvement-insights) in the middle column. The left column shows data insights on cycles C, trends T, and other O. Visualization and data insights are numbered according to the time point when they were gained.

P.Y. is the only participant who fully verbalized all visualization insights, how the Cycleplot works (H1 to H3). However, in his analysis of data, he had mainly insights into trends (T1 to T4, building on H3), but did not use the insights H1 and H2 for interpretation of the data.

Maybe the integration of prior knowledge (P1) and the identification of the practical value of these insights let him focus only on trend insights, neglecting cycle insights. In addition to the trend insights (T1 to T4) he extracted a further data insight that led to an insight of a further use of this visualization, namely a combined view of means and standard deviations (M5) that can also be used to identify outliers (M6). All other visualization insights (4 to 9) focus on improvement possibilities (I4, I7 to I9). After the explanation, P.Y. did not generate any further data insights.

What can be seen from P.Y.'s uptake graph is a high use of prior knowledge for the interpretation of the visualization: P.Y.'s cognitive scripts and knowledge about time data influenced to a high degree how he extracted information from this visualization. In contrast to P.X., this was especially true for improvement insights into the visualization.

Comparing all participants' insights into the Cycleplot, P.Z. generated most insights – into the data and into the visualization. Still, he seemingly did not verbalize all visualization insights, as the first data insights identifying cycles (C1 to C3) do not build upon a visualization insight. In contrast the trend insights (T4 to T6) build upon an explicated insight how the visualization works (H1). All other visualization insights were either meta-insights (M2 to M6) or insights, how the visualization can be improved (I7 to I13). As was the case for the two experts, also P.Z. did not generate many data insights after the explanation. The only data insight he generated afterwards (C8) interestingly builds upon a data insight P.Z. had into the Multiscale visualization (both built upon the same data set!). This integration of prior insights into the data shows that P.Z. built a mental model on the data set across the two visualizations.

Fig. 7 shows the uptake graph of both participants for the Multiscale (see Fig. 4): it is built analogous to the one for the Cycleplot. The data insights in the first column are coded as overview O or detail D.

Participant Y.'s uptake of prior knowledge and prior insights during the exploration of the Multiscale is highly complex: nearly all his data insights build upon prior visualization insights. Very interesting is his data insight O5 where he uses prior knowledge to interpret his insights into the data. As for the Cycleplot, he does not produce any data insights after the explanation.

Participant Z's uptake graph can be segmented in four very clear exploration phases: During the first phase, he gains insights how the visualization works (H1 to H7). In the second phase he uses this knowledge to interpret the data, first on an overview (O1 to O3), then on a detail level (D4 to D5). P.Z. did not fully understand the visualization, he did not decode the 5-minutes-raster. It was explained to him, but did not lead to any data insights. Still, after the explanation he generated further data insights (D6 to D13), suggested improvements for the visualization (I9 to I17), and reflected it on a meta-level (M18). An interesting insight was D8, where he reflected his analysis behavior on a meta-level (M8).

Comparing both participants, Y.'s uptake is much more complex and his insights more often build upon prior knowledge. The time sequence of Z.'s uptakes is more clearly structured in phases. Looking across both visualizations, “individual styles” of exploration can be identified: P.Y. uses his prior knowledge to interpret both visualizations and reports his insights in full. P.Z. has similar exploration phases in both graphs and analyses the visualizations in a more step-wise fashion. What can be observed for both participants is that the Multiscale requires more how-insights into the visualization before data insights are gained than the Cycleplot.

4.3 Discussion

Overall, our analyses show that participants generated insights into visualizations from scratch, without training and even without expert-knowledge in the domain of time-oriented data analysis. To gain insight into the data, they first had to gain insight into the visualization on which their data insights could build upon. However, they did not need a full understanding of the visualization prior to data insights; they only had to gain some relevant visualization insights first. One of our participants did not obtain a full understanding of the Cycleplot visualization at all, but he still gained insights into the data. In the course of interpreting a visualization, insights into the visualization and insights into the data interchange and partly build upon each other.

The data insights gained at first sight of a visualization were not – as one might expect – superficial, but most of them were rather complex: Henry and Fekete [30] suggest to categorize data extractions in dependence of the needed degree of interpretation as highly complex (e.g., correlations, patterns, trends), medium complex (e.g., groups, outliers), or little complex (e.g., single values). Following this categorization for the data insights into the Cycleplot, 79 % are highly complex and 11 % medium complex. The data insights into the Multiscale are slightly less complex, but still 42 % are highly complex and 52 % are medium complex. So overall, the complexity of the data insights is rather high, even though participants had only mockups for interpretation.

The high quality of the insights is also reflected by their correctness: No incorrect insights were observed. An open question is whether no incorrect mental models about the data existed, whether these models were corrected too fast to be verbalized, or whether these models existed but were not verbalized.

Even though all participants generated a high amount of insights from scratch, the insights of the experts highly depended on their prior knowledge. Still, it seems that their analysis scripts were not fully activated: An important step of data analysis, the plausibility checks and validation of the data set, was not conducted: half a day missing in the month view was not mentioned by any participant.

The prototypical progress of insights over time for all participants can be divided in three phases: first, they generated (at least one) insight, how the visualization works. Second, they generated data insights that often built upon these first visualization insights. In the end, they suggested improvements for the visualization that would let them generate further data insights. The last phase might be artificial, as participants were instructed to suggest improvements and additional interactive features. However, some of these insights were generated before this instruction, suggesting that phase three also exists without instruction during mockup testing.

Participants were instructed to generate insights into two visualizations, the Multiscale and the Cycleplot. Slightly more insights were gained into the Multiscale. Interestingly, all participants had an audible aha-experience once during the interview, when they figured out that the Multiscale visualization contains a monthly calendar view and not only a week. Therefore, the calendar-metaphor seems to suit this visualization well and leads to fast understanding when recognized. As intended, insights into the Multiscale were both on the level of details and of the overview.

In contrast, for the Cycleplot not all participants generated insights on both intended levels, namely cycles and trends: only one participant generated both kinds of insight, one participant only gained cycle insights, and one participant decisively focused only on trend insights. These results suggest that this visualization is more difficult to decode than the Multiscale. This might be due to the fact that no everyday metaphor, like the calendar, exists for the Cycleplot. Therefore, more effort is needed to create an "ad-hoc-metaphor on the fly". These results show a need for improvement of the visualization or of the caption so that everyday metaphors can be applied, trends and cycles are highlighted alike, and both insights are supported.

During the testing, participants generated many suggestions for improvements: the most important suggestions were derived from problems encountered with interpreting the mockup of the visualization. Additional improvements were suggested, which are mostly related to participants' analysis scripts and prior knowledge of interactive feature.

- Individual selection of coloring (Multiscale): One participant had problems to differentiate between the green and blue colors used.
- Addition of parallel curves / metadata (Cycleplot): All participants asked for the possibility to display other curves parallel to the Cycleplot to identify correlations.
- Selection of data displayed (Multiscale, Cycleplot): Changing the initial selection, but also adding and filtering of data.
- Details onClick (Multiscale, Cycleplot): Being able to see the real value onMouseover or onClick.
- Additional statistics (Multiscale, Cycleplot): Displaying mean values for days, month, etc.; for the Multiscale it was also suggested to display the "mean color" (e.g. for weekdays) or the mean daily course.
- Zooming, scrolling, annotation (Multiscale, Cycleplot)

5 Conclusions

Many other studies [1][4][7][19] have used insight-methodology to measure the productivity related to visualizations. In contrast, in our study, insight evaluation techniques were used to analyze and improve two visualization mockups and to gain deeper knowledge about the workflow of exploratory analyses.

To stay close to the typical process of analysing time-oriented data, we asked experts in personnel planning to participate in this study. As we are confronted with time in our everyday life, we also asked a non-expert to participate in this study to compare novice and expert users. Overall, we did see little differences between experts and non-experts: both were able to generate insights into the visualization and the data from scratch in similar time and quality. However, the experts more often used their prior knowledge to interpret data. On the other hand, it seems that existing cognitive scripts also hindered alternative analysis processes. Both experts were unable to extract different data insights from the Cycleplot (cycles and trends), whereas our novice user was able to do so. Preliminary results from other interviews with novices currently under transcription show similar capabilities of non-experts. Other explanations for this difference could be a higher motivation of the novice user or that our experts – although familiar with time-oriented data – are lacking background knowledge and expertise in analysis of data from the police sector. This is also reflected by rare interpretations of gained data insights. Further studies within this project will account for this fact by a more task-oriented design, which will allow each user to use data of his/her own domain.

Another limitation of our study is the open design of the instruction: participants were not given any problems to solve or tasks to complete. Rather, they were asked to tell whatever they see in the data. This lead to a very explorative analysis behaviour, might have impeded the activation of existing analysis scripts (like checking for missing data), and provided a wide space for interindividual differences. Therefore, results should be generalized only tentatively beyond our cases.

Even though we found many differences between participants, all participants were able to produce insights from scratch. This indicates that no instruction is needed, but that these visualizations can be understood with only little effort at first sight. Therefore, these two visualizations are also suited to visually present results to an auditorium unfamiliar with them, as they are rather self-explaining or easy to explain. Our analyses further show that no complete understanding of visualizations is needed to obtain insights, but already partial understanding is sufficient.

Despite these limitations, the method used appeared to be helpful for our purpose, as this study focuses on formative rather than summative testing of visualizations. The insights gained by our participants can thereby be used to further improve and develop visualizations of time-oriented data. Insight studies, especially when focusing on insights into both the visualization and the data, turned out to be a promising methodology for participatory design of visualizations.

Acknowledgments. The DisCō project (project number: 813388) is supported by the program “FIT-IT Visual Computing” of the Federal Ministry of Transport, Innovation and Technology, Austria. Thanks to our project partner XIMES GmbH for valuable feedback and help in finding participants for our study.

References

1. Rester, M., Pohl, M., Wiltner, S., Hinum, K., Miksch, S., Popow, C., Ohmann, S.: Evaluating an InfoVis Technique Using Insight Reports. In: Proc. 11th International Conference Information Visualization (IV 2007), pp. 693–700. IEEE Computer Society, Washington (2007)
2. Perer, A., Shneiderman, B.: Systematic Yet Flexible Discovery: Guiding Domain Experts through Exploratory Data Analysis. In: Intelligent User Interfaces, pp. 109–118. ACM, New York (2008)
3. Robinson, A.C., Chen, J., Lengerich, E.J., Meyer, H.G., MacEachren, A.M.: Combining Usability Techniques to Design Geovisualization Tools for Epidemiology. CaGIS 32(4), 243–255 (2005)

4. Saraiya, P.B.: Insight-Based Studies for Pathway and Microarray Visualization Tools. Doctoral Thesis. Virginia Polytechnic Institute and State University (2006)
5. Shneiderman, B., Plaisant, C.: Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. In: BELIV 2006. ACM, New York (2006)
6. Andrews, K.: Evaluating Information Visualisations. In: BELIV 2006. ACM, New York (2006)
7. Sarayaia, P.B., North, C., Duca, K.: An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. IEEE Trans. Vis. Comput. Graph. 11(4), 443–456 (2005)
8. Trafton, J.G., Kirschenbaum, S.S., Tsui, T.L., Miyamoto, R.T., Ballas, J.A., Raymond, P.D.: Turning Pictures into Numbers: Extracting and Generating Information from Complex Visualizations. Int. J. Hum-Comput. St. 53(5), 827–850 (2000)
9. Zazkis, R., Dubinsky, E., Dautermann, J.: Coordinating Visual and Analytic Strategies: A Study of Students Understanding of the Group D4. J. Res. Math. Educ. 27, 435–437 (1996)
10. Kuchar, O.A., Hoeft, T.J., Havre, S., Perrine, K.A.: Graphically Speaking - Isn't It About Time? IEEE Comput. Graph. 26(1), 80–83 (2006)
11. Thomas, J.J., Cook, K.A.: Illuminating the Path: The Research and Development Agenda for Visual Analytics. IEEE CS Press, Los Alamitos (2005), http://nvac.pnl.gov/agenda.stm
12. Gaertner, J.: Evidenzbasierte Personal-Bedarfsbestimmung - Unterstuetzung durch Time Intelligence Solutions [TIS] (2008), http://www.gor-hcm.de/htdocs/Folien/2008/Gaertner.pdf
13. Gaertner, J.: Time Intelligence - Bedarfsanalysen, Forecasting, Arbeitszeitcontrolling. In: Fank, M., Scherf, B. (eds.) Handbuch Personaleinsatzplanung, Datakontext, Frechen, pp. 173–185 (2005)
14. Shneiderman, B.: Inventing Discovery Tools: Combining Information Visualization with Data Mining. Information Visualization 1(1), 5–12 (2002)
15. Pirolli, P., Card, S.: The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis. In: Proc. of 2005 Int. Conf. on Intelligence Analysis (2005), http://vadl.cc.gatech.edu/documents/2__card-sensemaking.pdf
16. Springmeyer, R.R., Blattner, M.M., Max, N.L.: A Characterization of the Scientific Data Analysis Process. In: Visualization 1992, pp. 235–242. IEEE Press, New York (1992)
17. Russell, D.M., Stefik, M.J., Pirolli, P., Card, S.K.: The Cost Structure of Sensemaking. In: Conf. on Human Factors in Computing Systems, pp. 269–276. ACM, New York (1993)
18. North, C.: Toward Measuring Visualization Insight. IEEE CG&A 26(3), 6–9 (2006)
19. Plaisant, C., Fekete, J., Grinstein, G.: Promoting Insight-Based Evaluation of Visualizations: From Contest to Benchmark Repository. IEEE Trans. Vis. Comput. Graph. 14, 120–134 (2008)
20. Bertini, E., Plaisant, C., Santucci, G.: BELIV'06: Beyond Time and Errors; Novel Evaluation Methods for Information Visualization. Interactions 14(3), 59–60 (2007)
21. Ellis, G., Dix, A.: An Explorative Analysis of User Evaluation Studies in Information Visualisation. In: BELIV 2006. ACM, New York (2006)
22. Hartmann, G.W.: The concept and criteria of insight. Psychol. Rev. 38, 242–252 (1931)
23. Jones, G.: Testing Two Cognitive Theories of Insight. J. Exp. Psychol. Learn. Mem. Cogn. 29, 1017–1027 (2003)
24. Cleveland, W.S.: The Elements of Graphing Data. Hobart Press, Summit (1994)
25. Cleveland, W.S.: Visualizing Data. Hobart Press, Summit (1994)

26. Shimabukuro, M., Flores, E., Oliveira, M.D., Levkowitz, H.: Coordinated Views to Assist Exploration of Spatio-Temporal Data: A Case Study. In: 2nd Int. Conf. on Coordinated and Multiple Views in Exploratory Visualization, pp. 107–117. IEEE Press, New York (2004)
27. Ratwani, R.M., Trafton, J.G., Boehm-Davis, D.A.: Thinking Graphically: Extracting Local and Global Information. In: CogSci 2003. Erlbaum, Boston (2003)
28. Suthers, D., Dwyer, N., Vatrapu, R., Medina, R.: An Abstract Transcript Notation for Analyzing Interactional Construction of Meaning in Online Learning. System Sciences. In: HICSS 2007, p. 4c. Computer Society Press, Hawaii (2007)
29. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
30. Henry, N., Fekete, J.: Evaluating Visual Table Data Understanding. In: BELIV 2006. ACM, New York (2006)

PowerPoint Multimedia Presentations in Computer Science Education: What Do Users Need?

Elke I. Reuss, Beat Signer, and Moira C. Norrie

Institute for Information Systems, ETH Zurich, CH-8092 Zurich
{reuss,signer,norrie}@inf.ethz.ch

Abstract. PowerPoint is one of the most frequently used tools to present multimedia for educational purposes. Nevertheless, little is known about the users' needs when using PowerPoint during lecturing. Our study focused on the presenter's needs in this context of use. We interviewed nine university lecturers from the computer science department by means of a questionnaire. The results show that users require features that are not yet realised with PowerPoint. For example, the control and use of other media should be adequately integrated. Or the navigation within the slide collection should be improved to better meet the users' needs. Based on our findings, we outline required system features and suggest solutions in form of a prototype.

Keywords: Multimedia Presentations in Education, User Centred Design, Interactive Paper.

1 Introduction

There are no independent studies about the spread of PowerPoint, but Microsoft estimates 30 million PowerPoint presentations are made every day, and the tool has about ninety-five per cent of the presentation-software market [1]. Hence, PowerPoint is probably one of the most commonly used tools in university education. In this application, not only slides are presented to students, but also other media such as video sequences or animations.

There are plenty of studies investigating the impact of PowerPoint on education quality [2,3]. An informative overview is given by Craig and Amernic [4]. Some studies also investigate content complexity and learning performance [5]. But only very few studies deal with the lecturers' needs when using PowerPoint for educational purposes. Golub [6] found in informal discussions with colleagues that preparing slides with formulae for a mathematical course was too time consuming, and hence users tend to use other media such as the whiteboard. To overcome this drawback, Golub suggested a computer-based system to produce handwritten transparencies. Another pen-based solution is proposed by Anderson et al. [7,8]. Their system allows the user to handwrite over computer-projected slides. Tufte [9] criticises that PowerPoint enforces a linear progression through a content that is arranged in a complex hierarchy that is often confusing. Based on this critic, Holman et al. [10] tried to

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 281–298, 2008.

improve absorbtion and recall of presentation content by using a mind map based visualisation concept. Also Zoomable User Interfaces [11] aims to simplify access to required slides. However, these criticism about PowerPoint and the suggested solutions are not based on the results of user studies. Asking users about their needs is a vital prerequisite when aiming to redesign a system to improve its usability.

Therefore, our investigation focused on the use of PowerPoint during lectures or tutorials for educational purpose. Does the tool really meet the users' needs when interacting with the system during lectures? In particular, our study aimed to answer the following questions:

(a) What is the content and organisation of the presentation?
(b) What media are used during the lecture/tutorial?
(c) How are users interacting with PowerPoint during their presentation?
(d) Do users have suggestions for improvement of the tool?

In the next two sections, we describe the study design and results. In section 4 we formulate implications for system features that should better reflect the users' needs and hence improve the interaction during presentations. Section 5 presents our prototype that implements the suggested features, and section 6 discusses our findings. Concluding remarks and an outlook to future work are given in the last section.

2 Study Design

With the help of a questionnaire we interviewed nine university lecturers that regularly use PowerPoint during their lectures and/or tutorials. Five assistants and four professors of the computer science department of of ETH Zurich and the University Zurich volunteered to participate in our study. An interview took on average 37 minutes. The questionnaire comprised several multiple choice and open end questions that dealt with the users way of interacting with PowerPoint during presentations. Participants filled in the questionnaire and gave additional information orally. The interviewer took notes, asked further questions if necessary and reported about these results later on in a detailed text. Each participant was asked to provide us with a representative example of one of their presentations for student education.

The participants' ages ranged from 25 to 62 years with a median of 32 years, one participant was female and eight male. All those interviewed have used PowerPoint for between 7 and 15 years, and the majority (n=8) either use version 2003 or 2007. Six persons stated that they use PowerPoint for 90% or more of their lectures/tutorials, and three participants use it between 50 to 70% of their lectures/tutorials.

The questionnaire comprised some general questions about the content and structure of the presentation and what devices are used when presenting with PowerPoint. We also asked participants what other media they use during lectures, what functionality of PowerPoint they usually use, and whether they prefer to stand at the lecturn or nearby their audience during the presentation. Another important issue was to find out whether and how the interaction with PowerPoint could be improved from the user's point of view.

3 Results

To learn about the users' possibilities when using PowerPoint, we asked participants what hardware they use during their lecture or tutorial. Eight participants stated that they use a laptop, and one a Tablet PC. Additionally, most of the users apply a pointing device, i.e. a laser pointer or a stick (n=8) and a remote control (n=6). Although three participants had a pen-based input facility, they used it rather infrequently, as they did not like its haptic or the required posture when using it.

3.1 Content and Organisation of Presentation

Participants stated that their slides mostly contain textual information (ranging from 20-90%, with a median of 55%), but also graphics/pictures (ranging between 8 to 60, median=40) and formulae (ranging between 0 to 30, median=5, see Figure 1).

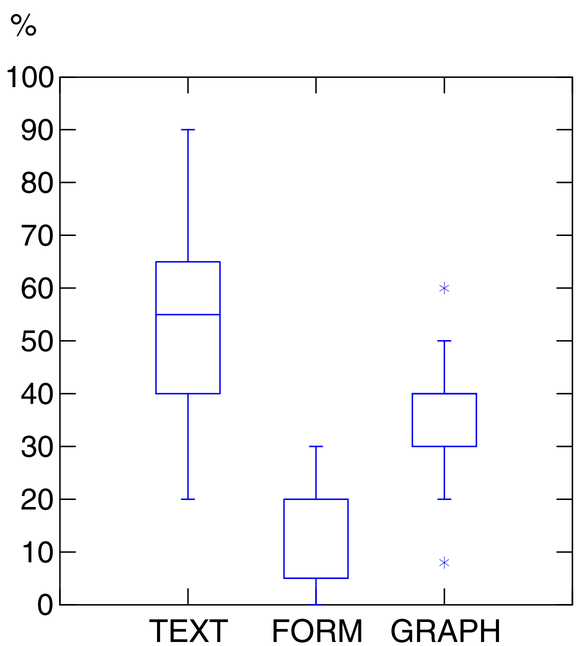


Fig. 1. Percentage of information types on the slides

The evaluation of the provided example presentations showed that slides containing only text have between 2 to 14 lines and a minimum at 18 point font was used (see Figure 2). Text in graphics was smaller, but at least it was a 12 point font. Furthermore, the majority of the evaluated presentations contained one or more animations.

When preparing the presentation, all participants sorted the slides according to the order in which they intended to present them. Almost all interviewed stated that they

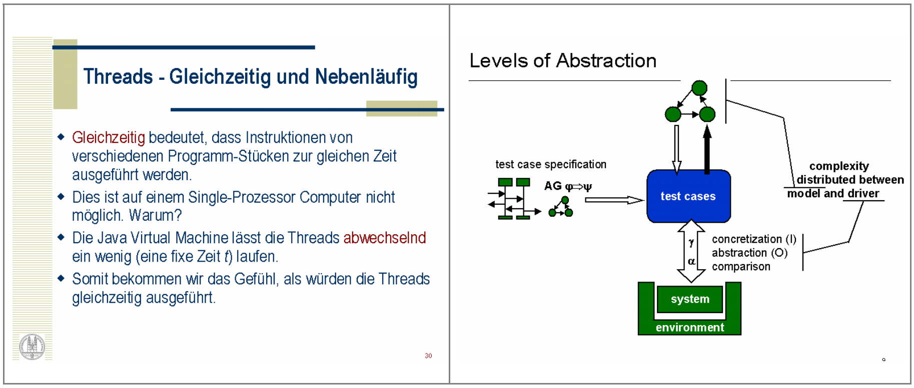


Fig. 2. Two example slides from lecturers

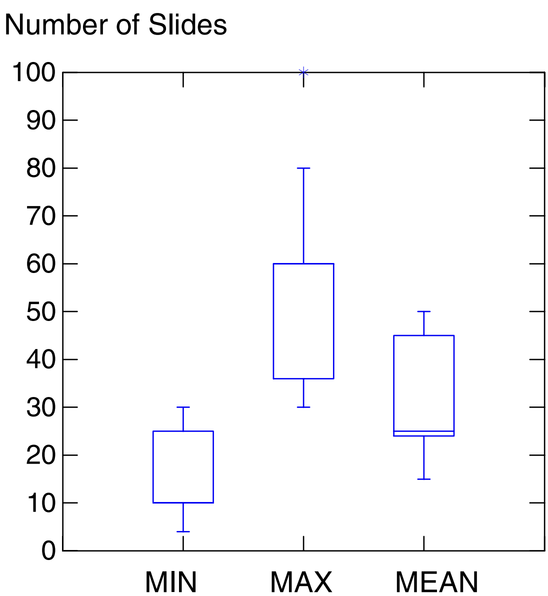


Fig. 3. Number of slides of the presentation

use extra slides (n=8) either filed at the end of the presentation (n=6) or within the presentation as a hidden slide (n=2). Another strategy used was to store extra slides in a separate file (n=2). A lecture takes between one and three hours, ususally made up of 45 minutes blocks.

According to the interviewed participants, their presentations for a lecture or tutorial block contain in a minimum of between 4 to 30 slides (median is 10, see Figure 3), and at most between 30 to 100 slides (with a median of 60). On average, the participants' presentations consist of 15-50 slides (with a median equal to 25, and an average of about 33 slides).

All participants provide students with a copy of their presentation, seven of them deliver all the slides (hardcopy or file) before the lecture/tutorial takes place, and four of them after the presentation took place. Eight participants stated that they also make minor changes on the presentation shortly before the lecture or tutorial starts. Only one of those interviewed changed the order of the slides, the majority (n=7) edited, deleted or added slides.

3.2 Presenting Information

We asked participants whether they prefer to stand at the lecturn or nearby the audience during the presentation. Eight interviewed prefer to stand nearby the audience, and all like to move around during the presentation. All interviewed also frequently look at the presented slides to remember keywords, to verify slide progression or to point at an item with the pointing device.

When presenting with PowerPoint, participants use additional media. All interviewed write on white- or blackboards, seven participants also work with an overhead projector and six persons present video sequences during their lectures/tutorials (see Figure 4). The whiteboards and blackboards are mainly used to make rather brief notes, e.g. giving ad hoc examples or derive something (e.g. code or formulae). The overhead projector is also used for such ad hoc notes, but six of the seven users stated that they also use it to present prepared transparencies in parallel to the PowerPoint slides. These participants emphasized the need to present information in parallel, e.g. they presented a list with abbreviations or an important diagram in parallel to the slides.

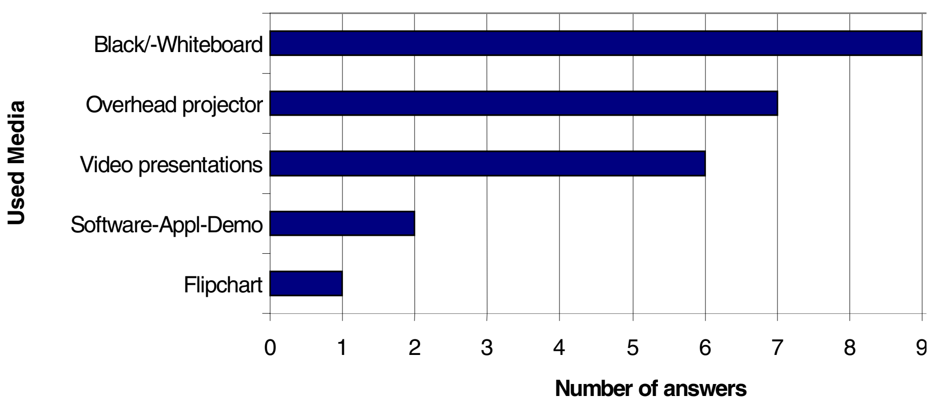


Fig. 4. Use of other media during lecturing with Power Point

Only one of the interviewees provided students with a copy of his notes taken on the whiteboard after the tutorial. All other participants stated that students should also make notes themselves during the presentation.

From the six users that present video sequences, five would like to control it using functions provided by PowerPoint such as starting or stopping the video, or to replay a part. But all six users stated that they do not trust the current PowerPoint video control features, and therefore do not use it.

All interviewed (n=9) frequently look at the screen or at the computer display. This was done to check which slide is currently displayed, to have a closer look at the content of the current or next slide (e.g. to recall keywords or to observe slide progression), and to highlight or annotate information on the presented slide. As mentioned before, eight participants use a pointing device to highlight information. Two participants stated that the present way of highlighting with the laser pointer to be unsatisfying. On the one hand, in larger lecture rooms, it is difficult to detect the beam. On the other hand, if students did not permanently observe the screen, they might miss when something was pointed at with the laser pointer.

3.3 Use of Annotations

Only two participants stated that they annotate slides during their presentation. One of them annotates about 1% of his slides using the menu to choose the pen. The other participant annotates 50-75% of his slides by writing on the whiteboard that is used as a screen to display the slides. This participant believes that no electronic tool can ever reach the haptic and feeling of pen & paper or whiteboard. He even stated to have another handwriting when using no electronic tools. Another interviewee changed about 5% of the slides during her presentation by using the keyboard.

The majority of participants did not annotate their slides argueing that it was uncomfortable to use the digital pen (n=2) or that there was no need for them to annotate the slides (n=3). One participant stated to have difficulties to annotate without having a pen-based input facility. However, supposing that the handling of PowerPoint would be as easy as using pen and paper, seven from the nine participants would like to add annotations, marks or sketches during their presentation. Participants would mainly use such a feature to encircle, underline or mark information for accentuation. They also would like to add sketches or write down information on a blank sheet to substitute the use of a black- or whiteboard.

3.4 Navigation and Orientation

When going from one slide to the next or preceding slide, all users used the “next or previous slide” functionality by pressing a key or using the presenter tool to switch. This was the most common way of navigating as the participants sorted their slides in the order in which they were going to present them.

To access a specific slide that was not the next or preceding one, five users also used the “next or previous slide” functionality, but only if the required slide was within the range of about two to ten slides. If this was not the case, all users navigated using the thumbnails in the presenter mode or the slide sorter view. All users stated that they know their presentation quite well, as they prepared it themselves. One participant even knew the slide number of important slides and entered the slide number with the keyboard to access them.

For three participants, their concept to navigate to a required slide was fully satisfying. One of the participants put it in these words: "In the slide sorter view, I look for the specific layout of the needed slide or other slides with graphics nearby to orientate myself with the help of such landmarks. I also use the slide title that helps me to orientate myself ".

Another three users also felt comfortable with their concept, but they had suggestions for improvement of PowerPoint's navigation features. Three interviewed stated that they are uncomfortable with their concept of accessing a specific slide. Those participants who were not satisfied argued as follows. "It may be hard to navigate to the most important slides and to do so in a directed order. It would help me to have some visual grouping by topic/concept in the presentation."

Another participant said "it can be confusing to the audience when I escape the presentation to search a specific slide. It is also hard for me to concentrate and to get back to the flow afterwards. Additionally it is difficult for me to come back to the slide, where I interrupted the flow (i.e. branched to the specific slide). It would be useful, if the tool supported me to easily get back to this slide."

A third intervieweed stated that "the used title approach, i.e. searching a specific slide with the help of its title, often requires fine-tuning and using the thumbnails, the information sometimes is presented too small. The tool should provide me with a feature that enlarges those thumbnails I want to have a closer look at in order to decide whether to present them. This should help to choose the right slide when looking for a specific slide. Additionally, the audience should not notice this search process and hence enlarging some thumbnails should be independent from presented info. It would also be helpful, if a larger preview of the next slide to be displayed would be offered by the tool."

From those six users that said they are comfortable with their navigation concept, three had suggestions to improve access: "A quick search functionality would be useful." "I find needed slides easily within the slide sorter view, but I would prefer if the audience does not see that I changed to the slide sorter." "The presenter view is ok for me, though a better overview would be nice and also a configurable presenter view."

All in all, six participants had suggestions to improve access to specific slides.

3.5 Further Suggestions for Improvement

Eight interviewed had further suggestions or remarks to improve the interaction with PowerPoint. Their answers can be categorized and summed up as follows.

A) Improvement of the visual representation of the slide collection. It should be easy to identify the presentation's structure/hierarchy, to get a quick overview and to detect important slides. (n=4)
B) A quick search by keywords should be provided (n=1)
C) Improvement of the video control. (n=1)
D) The system should provide a flexible way at organising information displayed on the presenter view screen. (n=1).

4 Implications for System Design

Based on the presented results, we outline implications for the system design in the form of features formulated in a general way, i.e., not describing a solution. In the next section, we describe possible solutions and prototypes.

Table 1 lists the suggested features that are briefly described in the following. Feature F01 should provide a simple functionality to highlight information on the slide or to make annotations on it. The user and the audience should be able to easily detect the highlighted item. If information should be highlighted permanently, the user should be able to easily switch to the annotation mode. The handling should be as easy and intuitive as with pen and paper, as this is an important prerequisite for the user's acceptance.

As all interviewed use a white- or blackboard to write down information, the system should also provide feature F02. In this scenario, the annotation feature described above is used together with a blank sheet, that should be quickly accessed when needed. We are aware that most of these interviewed use the white-/ blackboard or the overhead projector to present information in parallel to the presented slides. To meet this user requirement, a second beamer would be required in the lecture room. But at the two institutions where the study was performed, lecture rooms are usually equipped with only one beamer. Apart from this constraint, it would be no problem to design the feature such that it provides a parallel presentation of information.

Table 1. Suggested system features

Feature ID	Brief feature description
F01	Highlight / annotate something on the slide
F02	Use blank sheet
F03	Use video control
F04	Use system mobile
F05	Orientate efficiently within slide collection
F06	See content of presented and nearby slides

Most of the participants present video sequences during their lecture. Hence, feature F03 requires a reliable functionality to control the video, i.e. to start, stop or replay it, or to access a specific scene.

The mobility of the system is also an important feature. Eight from the nine interviewed prefer to stand nearby the audience and not at the lecturn, and all like to move during their presentation.

As outlined in section 3.4, an efficient way of navigating the slide collection is also of great concern to users. System feature F05 should help the user to easily go through all slides of the collection including extra slides, and to efficiently access a specific slide. It also should provide the user with facilities to identify the presentation's structure and to detect relevant slides. A suitable overview which also includes visual landmarks should be supported. As the study showed, the presented information comprises between 10 to 80% graphics or formulae (with a median of 45%) that call for a characteristic layout and hence should provide the user with such cues. Furthermore, this feature should help users to easily get back to the slide from which they have branched.

Based on the results of section 3.4, we suggest feature F06 that is also a prerequisite for features F01 and F05. Users should see the content of the presented slide, for example to highlight or annotate something, or to recall keywords. As the results show, users also require to control the progression of the presentation, which also should be provided by this feature.

5 Paperpoint Prototype

The PaperPoint prototype for giving interactive paper-based PowerPoint presentations described in this section is based on the interactive paper platform (iPaper) [12, 13] that we have developed over the past few years. The iPaper framework uses an input device based on Digital Pen and Paper technology offered by Anoto [14] to enable links between paper and digital information/ services. After a special Anoto pattern consisting of tiny and almost invisible dots has been printed on regular paper documents, the digital pen can track its absolute position within these paper documents.

Our iPaper platform allows active areas in the form of arbitrary shapes to be defined on paper documents. As soon as the digital pen is used within such an active paper region, additional digital information is provided or a digital service is invoked. In the case of the PaperPoint prototype different PowerPoint functionality is accessed either directly though the Component Object Model (COM) or indirectly though additional application logic implemented in Java. More information about the PowerPoint integration and the description of an earlier PaperPoint prototype can be found in [15].

Based on the implications for the system design presented in section 4, we have implemented a new PaperPoint prototype paying attention to the different features. In this section, we present our initial solutions to fulfil these requirements and discuss potential alternatives where appropriate. Please note that this is an initial PaperPoint prototype and the usability of the new functionality is going to be evaluated in detailed lab studies. The PaperPoint presentation prototype consists of the two different types of interactive paper documents shown in Figures 5 and 6. The interactive handouts presented in Figure 5 contain 12 slides on each page as well as some additional control buttons to interact with a specific PowerPoint presentation. In addition, we provide interactive note sheets (see Figure 6) for creating and visualising digital notes at presentation time. In the remainder of this section, we will describe the detailed functionality of these two document types. Note that the preparation of a new PaperPoint presentation does not require any additional steps. After a user has created a regular PowerPoint presentation, it can be turned into an interactive PaperPoint presentation by printing it using our special interactive paper printer driver. The printer driver will add the required Anoto pattern and, at the same time register the printout with the corresponding PaperPoint application.

While the PaperPoint presentation tool has to run on a laptop or some other computing device, the digital pen is connected over a wireless Bluetooth connection. This implies that the presenter can freely move around during their presentation as long as

Fig. 5. PaperPoint interactive handouts

Fig. 6. PaperPoint interactive note sheets

Fig. 7. Structure of the presentation

they stay in the area covered by the Bluetooth connection (in our case at least 15 meters). The weight of the system, which has a direct impact on the system's mobility, thereby has been reduced to a minimum. The presenter only has to carry the digital pen and a few paper handouts with them. We found that many PaperPoint users attach the paper documents to a clipboard, to get a proper writing surface. However, even by using an additional clipboard the overall weight of the PaperPoint presentation tool is still only a fraction of any Tablet PC based solution for mobile presentations. Therefore, we think that our PaperPoint prototype pays attention to our system feature F04 for mobility introduced in the previous section.

A first decision we had to make in terms of the design was the number of slides that we should put on a single handout page of our prototype. With each additional handout page, the navigation within the slide set becomes more difficult since the user may have to turn a lot of pages to get access to the corresponding printed slide. However, there is a trade-off between the number of slides we can put on a single page and their readability. We decided to use A4 sized handout pages (in landscape mode) and to put 12 slides on a single handout page as shown in Figure 5.

Based on our findings from the interviews that the mean number of slides for a lecture/tutorial (45 minutes) shows a median of 25 slides, about 50% percent of the presentations probably can be handled by two interactive paper handout pages (24 slides). By touching one of the printed slides with the digital pen the PaperPoint presentation will immediately switch to the selected slide. A presenter can easily orientate themselves in a collection of slides and has fast access to any slide within the presentation by pointing to any printed slide on the two handout pages with the digital pen. The PaperPoint handout navigation allows a presenter to efficiently navigate within their collection of slides, as required by feature F05, without using the PowerPoint slide sorter functionality.

The PaperPoint handouts do not only provide a good overview over a large set of slides but also provide contextual information about the previous and next slides. As required by our system design feature F06, the presenter can always see at least one succeeding slide, except if they are at the last slide of a handout page. The visibility of succeeding slides on the paper handouts may help the presenter to orientate themselves within their presentation.

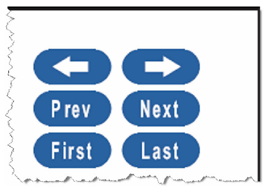


Fig. 8. Navigational controls

Furthermore, we plan to provide additional information about the structure/hierarchy of a presentation as requested by some users. Since our interactive paper printer driver enables us to add any supplemental information to the printed handouts, we suggest to use different coloured slide backgrounds to group slides and highlight a presentations' structure as shown in Figure 7. By applying this colour coding for grouping slides, the user is well aware about the structure of their presentation which might help them to navigate within their collection of slides. While it is easy to add the differently coloured slides backgrounds, we are not yet sure what will be the easiest way to add the additional structural metadata during the slide authoring process. Since PowerPoint does not offer any slide grouping functionality, we will have to find an alternative way how this structural information can be entered. One possibility would be to add empty slides as markers for the boundaries between different slide groups when authoring a presentation in PowerPoint. These empty slides would show the interactive paper printer driver when the background colour has to be changed, but they would be rendered on the printed PaperPoint handouts.

The first set of navigational controls are the Prev and Next buttons. They can be used to access the next or previous slides in a similar way to pressing PgUp and PgDn on the computer keyboard while running PowerPoint in presentation mode. To easily access a presentation's first or last slide we further provide the First and Last buttons. As mentioned earlier, an advantage of the PaperPoint presentation tool is that users can easily present their slides in a nonlinear order. For example, if there is time pressure they can easily skip some slides or if there is a question they can switch to a specific slide for clarification. To further support nonlinear navigation, we provide two additional buttons '←' and '→'. The two buttons work similar to the forward/backward buttons available in any web browser by providing access to the next or previous slide in the order, in which actually they have been presented as opposed to the order in which they occur in the prepared presentation. The forward button ('→') works slightly different than the forward button available in web browsers. If there is a "next slide" in the navigational buffer, this slide will be shown. However, if there is no "next slide" in the buffer the forward buttons works like the "Next" button and returns the succeeding slide in a linear order.

While the '←' and '→' buttons help to nonlinearly navigate forwards and backwards in the digital presentation, the user may have a problem to do the same on the paper documents. If there are only one or two paper handouts the user should be able to synchronise the paper and digital versions and find the corresponding slide on the paper handouts after using the'←' or '→' buttons. However, in the case that the number of slides and paper handouts increases, it might get more difficult to find the corresponding slide on the paper handouts. For this reason, we plan to add an additional "Where am I" button which will help the user to find their position on the paper handouts by providing acoustic or visual feedback. This feedback could for example consist of the handout number quickly showing up on the presentation screen after the user selected the "Where am I" button. The user can then turn to the corresponding paper handout page and should be able to find the current slide. Note that since we do not use the handout print functionality offered by PowerPoint, but rather assemble the

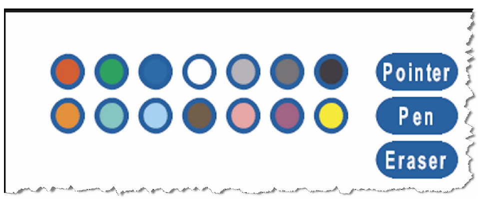


Fig. 9. Pointer/annotation controls

slide handouts within our printer driver, we are much more flexible in the different slide handouts that can be created with PaperPoint (e.g. arbitrary number of slides on a single page). Therefore, in the future, we might offer the user other options such as printing all slides as thumbnails on a single page which can then be used as an additional navigational tool.

As shown in section 3.2, only two participants of the study currently annotate their slides. However, most of the participants further stated that they do not annotate their slides because it is not easily supported in existing systems. The participants using laser pointers further mentioned that they would like to have an alternative more persistent way of highlighting specific parts of a slide. The following functionality for feature F01 therefore pays attention to highlighting and annotating parts of a slide. The pointer controls that have to be used for these tasks are located in the upper left corner of the paper handouts and are shown in detail in Figure 9. The user has the possibility to choose, whether the highlighting or the annotation mode is used by default.

By touching the Pen button, the user can switch to the annotation mode where parts of the slide can be annotated or permanently highlighted. After optionally choosing a colour, the user can draw on any paper slide and the annotations will be shown in the digital presentation in real-time. All annotations are persistently stored within the digital copy for later access. Furthermore, any annotations can be replayed at a later stage based on the timestamp information stored together with the position data provided by the digital pen. If a user wants to delete parts of their annotation, they can do that by first selecting the Eraser paper button and then crossing out the corresponding stroke on the paper handout. Note that while the annotations will be removed from the digital copy, persistent traces will remain on the paper handouts.

In addition to the persistent highlighting/annotation of content, we provide a simple pointer to temporarily highlight parts of a slide. After selecting the Pointer button, the user can highlight specific slide regions. We currently use the standard pointer functionality offered by PowerPoint. However, since some of the participants mentioned that they would like to have a slightly more persistent way of highlighting slide regions (not possible with a laser pointer) we plan to extend the pointer functionality by adding some kind of afterglow effect. This means that the pointer would leave a trace that fades out after a specific amount of time.

In addition to the annotation of existing content, the user can easily create new content at presentation time by using the interactive note sheets shown earlier in Figure 6. By using the blank sheets, the presenter can not only take notes on the paper documents but again make the information accessible in real-time to their audience. As soon as the presenter start to draw on one of these empty sheets, the digital presentation will switch to a digital blackboard and show all information that is written/drawn on the paper note sheet. The user can switch between multiple digital blackboard instances by simply drawing on different paper note pages. As soon as they continue to add additional information to a paper note, the digital blackboard with all previous information that has been written on this specific instance will be shown. By providing the paper note pages we achieve a convenient integration of the requested blackboard/whiteboard functionality (feature F02).

Note that our system supports multiple pens at the same time. This means that the PaperPoint presentation tool can also be used in more open meeting situations. In this situation each user gets a copy of the paper note pages and a digital pen. The users can then collaboratively brainstorm by drawing on their individual paper copies. Each individuals contribution will be immediately integrated into the single digital blackboard instance. By choosing different pen colours user can furthermore provide information about who contributed what to the digital sketches. Even without using different pen colours, we can later still distinguish the individual contributions based on the individual logging information from the different pens. Of course the collaborative sketches can be printed out at any stage so that the participants of a meeting can take the integrated paper notes with them after a meeting.

Finally, our PaperPoint prototype provides some functionality for controlling embedded videos as proposed by feature 03. A detailed version of the video controls is shown in Figure 10. The user can play, pause or stop a video embedded in a PowerPoint presentation by using the Play, Pause and Stop paper buttons. The current implementation can only deal with a single video at a given time. However, for the final version of the prototype we plan to keep track of the status of all videos in a presentation. This means that a user could pause a video on a given slide, switch to another video embedded in another slide and later come back to the video paused earlier to continue the replay by pointing to the Play button.

In addition we plan to provide a "paper slider" component which allows users to navigate within a video based on timing information. The slider provides access to the video in a time neutral format (from 0 to 100 percent) and after pointing somewhere within the slider component the video will jump to the corresponding location. We are still investigating how the slider component can be implemented based on the functionality offered by PowerPoint. Of course this time-based control assumes that the

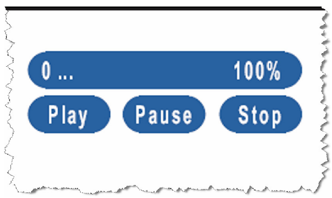


Fig. 10. Video controls

presenter knows about the temporal ordering of a video's components. Since we have full control over the printing process, a potential future extension might analyse the video during the printing process. We could then identify the video's key frames based on existing approaches and print thumbnails for the key frames within a video slide. These thumbnails would represent visual bookmarks for efficiently accessing different parts of a video.

6 Discussion

Nine university lecturers from Computer Science have been interviewed to investigate their needs when interacting with PowerPoint for educational purpose. Although this seems to be a rather small number of participants, according to Nielsen and Holtzblatt et al. it is a sufficient number of interviewees for usability engineering [16,17]. The method how to perform interviews is described in [17].

Users frequently highlight items during their presentation, and they also would like to annotate information. These findings are confirmed by a study of Anderson et al. [8]. In this study, three lecturers of computer science courses used a Tablet PC system with highlighting and annotation facilities. It showed that users in fact annotate prepared or blank slides, and they also use annotations to highlight information. But the system was installed at the lecturn and hence mobile use was not possible. As our study shows, users like to stand nearby their audience and move a lot which requires a lightweight mobile system as proposed with PaperPoint.

The intervieweed also stated that the handling of an annotation feature should be as simple as using pen and paper. Existing prototypes [6,8] suggest to use a Tablet PC to provide annotations. With PaperPoint, pen and paper is used to either annotate slides or to write on a blank slide.

Together with the white-/blackboard, users use more media during their presentations such as video, which is confirmed by several studies [4,18]. We consider that all required media should be controlled by one system, and hence suggest to additionally integrate a video control that can be easily used when needed. If the lecture rooms are equipped with multiple beamers, a framework to manage multiple displays as described in [19] could be used.

When presentations with dynamic multimedia such as animations or video presentations are used, one should consider Holzinger et al. findings [5] that especially dynamic media must be attuned to learners previous knowledge. Information, animations, and elements, which are not necessary to comprehend a concept should be avoided.

Birnholtz found by means of observation of one lecturer in a computer science course that slides were changed on average every 3 minutes [18] which results in 15 presented slides in one lecture. Another observation study in computer science with three participants found an average number of about 21 presented slides per 45 minutes [8], with the lecturers heavily annotating slides. Our participants stated to have on average about 33 slides in a presentation for a 45 minutes lecture. Hence, the different results are not really comparable due to several facts. First of all, the results from Birnholtz and Anderson et al. report on the presented slides but not about the number of slides contained in the presentation. Second, the two studies did not report on the content of the presented slides. The slides of our participants show large text fonts and between 2 to 14 text lines, i.e., not too much information per slide which might explain the difference to some extent. Furthermore, our result is based on the users‘ estimation.

We were interested in the number of slides of the presentation as this is the amount to be managed when navigating. During lectures, users usually access slides according to the prepared order, and hence PaperPoint provides an interaction handout presenting this order. When a specific slide was needed, 6 out of 9 participants had suggestions to improve PowerPoint's navigation. Our participants consider visual landmarks to be helpful. Such marks are either pictures, graphics or formulae on the slides. Furthermore, users would like to see the presentations's structure to more efficiently detect relevant slides within the collection.

Visual landmarks are also suggested by Good and Bederson [11]. They suggest to use a zoomable presentation and planned to evaluate their approach. The Fly tool [10] uses a mind map based slide layout to provide users with structural information. Holman et al. evaluated their Fly system using a presentation with five slides. They focused on the users' retention of content compared with the linear presentation of PowerPoint, but found no significant difference between the two concepts. It is not evident, if the difference would be more clearly when more slides, e.g. 25 or more, were used.

PaperPoint suggests the combination of both, visual landmarks and a layout that visualizes the presentation's structure. We think, this should improve the users‘ navigation and provide an efficient access to needed content.

7 Conclusion and Future Work

We investigated the users' needs when interacting with PowerPoint during lecturing. Based on our findings, we suggest a mobile system that improves and enhances PowerPoint. In particular, our solution provides users with the ability to manage multimedia and control the system as easily as when working with pen and paper. In a next step, we aim to perform user studies to evaluate our prototype.

Acknowledgments. We would like to sincerely thank all professors and assistants who volunteered for this study. We would further like to thank Nadir Weibel and Claudio Hatz for implementing parts of the PaperPoint application.

References

1. Parker, I.: Absolute PowerPoint: Can a software package edit your thoughts? The New Yorker (2001) (accessed March 6, 2008), `http://nbc-links.com/miscellaneous/AbsolutePPT.pdf`
2. Bartsch, R.A., Cobern, K.M.: Effectiveness of PowerPoint presentations in lectures. Computers & Education 41(1), 77–86 (2003)
3. Susskind, J.E.: Limits of PowerPoint's Power: Enhancing students' self-efficacy and attitudes but not their behavior. Computers & Education 50(4), 1228–1239 (2008)
4. Craig, R., Amernic, J.: PowerPoint Presentation Technology and the Dynamics of Teaching. Innov. High Educ. 31, 147–160 (2006)
5. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
6. Golub, E.: Handwritten Slides on a TabletPC in a Discrete Mathematics Course. In: Proc. of the 35th SIGCSE Technical Symposium on Computer Science Education, pp. 51–55 (2004)
7. Anderson, R., Simon, B., Wolfman, S.A., VanDeGrift, T., Yasuhara, K.: Experiences with a tablet PC based lecture presentation system in computer science courses. In: Proc. of the 35th SIGCSE Technical Symposium on Computer Science Education, pp. 56–60 (2004)
8. Anderson, R.J., Hoyer, C., Wolfman, S.A., Anderson, R.: A study of digital ink in lecture presentation. In: Proc. of the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria, pp. 567–574 (2004)
9. Tufte, E.: The Cognitive Style of Power Point. Graphics Press (2003)
10. Holman, D., et al.: Fly: An Organic Presentation Tool. In: Work in Progress Contribution to CHI 2006, pp. 863–868 (2006)
11. Good, L., Bederson, B.: Zoomable User Interfaces as a Medium for Slide Show Presentations. Information Visualization 1(1), 35–49 (2002)
12. Norrie, M.C., Signer, B., Weibel, N.: General Framework for the Rapid Development of Interactive Paper Applications. In: CoPADD 2006, 1st International Workshop on Collaborating over Paper and Digital Documents, Banff, Canada, pp. 9–12 (November 2006)

13. Signer, B.: Fundamental Concepts for Interactive Paper and Cross-Media Information Spaces. PhD thesis, ETH Zurich, Dissertation ETH No. 16218 (May 2006)
14. Anoto, A.B.: http://www.anoto.com
15. Signer, B., Norrie, M.C.: PaperPoint: A Paper-Based Presentation and Interactive Paper Prototyping Tool. In: Proceedings of TEI 2007, First International Conference on Tangible and Embedded Interaction, Baton Rouge, USA, pp. 57–64 (February 2007)
16. Nielsen, J.: Usability Engineering. Academic Press, Chestnut Hill (1993)
17. Holtzblatt, K., Burns Wendell, J., Wood, S.: Rapid Contextual Design. A How-To Guide to Key Techniques for User-Centered Design. Morgan Kaufmann, San Francisco (2005)
18. Birnholtz, J.: Back to School: Design Principles for Improving Webcast Interactivity From Face-to-Face Classroom Observation, DIS, 311–319 (2006)
19. Liu, Q., Kimber, D., Zhao, F., Huang, J.: Framework for effective use of multiple displays. In: Proceedings of SPIE Int. Symposium ITCom on Multimedia Systems and Applications VIII (2005)

Analysis of Ontology Visualization Techniques for Modular Curricula

Simone Kriglstein

University of Vienna, Department of Knowledge and Business Engineering,
Rathausstrasse 19/9, A-1010 Vienna, Austria
simone.kriglstein@univie.ac.at

Abstract. It is often a challenge for students and lecturers to see the dependences between courses and modules in curricula that exhibit complex interrelationships. Therefore, a visual representation of the structure and the relations between courses and modules helps to transparently expose curricular structures. Visualization via ontologies is a good alternative to show the structure and their interrelationship. Furthermore, ontologies give users the possibilities to understand, exchange, analyze or share knowledge of a specific domain. However, it is a big challenge for the visualization to represent ontology in a way that is useable for users according to their needs. Based on the fact that more research is needed in regard to ontology visualization with respect to users' needs, we present an example how an ontology based on a curriculum can be composed and analyze three different visualization techniques to find out useful requirements.

Keywords: Information Visualization, Ontologies, Structure of Curricula, Visualization of Knowledge, Treemap, Node Link Approach.

1 Introduction

The curricula of the University of Vienna were newly rolled out during the winter term 2006 based on the requirements of the Bologna Process [1]. The new curricula (bachelor and master curricula in business, medicine, media, scientific computing and bioinformatics) are defined by a clear modular structure [2]. Each module has a set of courses and several modules exhibit complex interrelationships. Generally, universities particularly provide their curricula only in textural from, e.g. as lists of course name and descriptions, which makes it difficult to overlook the training offer (see [3]). Additionally, if the curricula are particularly described in words, it can be difficult for students to see the dependences between modules as well as courses at first glance. Furthermore, for lecturers it is also a challenge to get a good overview about the relevant relationships between the courses. It is difficult to see which courses are a prerequisite of their courses. This additional information is necessary for lecturers to organize and align the content of their courses optimally in consideration of students' needs. The challenge to handle students with different prior knowledge is also noted by Rubin and Misra in their paper [4] in which they describe their experiences in developing a computer security curriculum. Another interesting aspect which influences a curricula program, is that, especially in the field of computer science, which is grown in the last years, new subfields are emerged as well as the exiting subfields extend their knowledge. Therefore Kay et al. [5] note among

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 299–312, 2008.

other things that it is necessary that curricula provide a configurable program. This program should contain core courses as well as optional courses to provide flexibility. The drawback is that the number optional courses can rise as well as their dependencies according to other courses.

Based of these challenges it is meaningful to use additional graphical representations to produce abstract information structures for human users [6]. Therefore, a visual representation can help to grasp the structure and inter-object relationships in an easy way and makes the structure and dependencies between the courses more transparent. In other words, visualization helps humans to see the unseen and provides for example what students are doing, what they have learned, where they may be going [7]. Since curricula can be very large and complex, visualizations provide dynamic elements (e.g. interaction between users and visualization) in addition to static elements (e.g. text) for a deeper understanding of the structure. Holzinger et al. present a study about the learning performance about static versus dynamic media in [8]. They found out that the learning performance is better if using dynamic media depending on the complexity of the learning content. These results show that dynamic visualizations can reduce the cognitive load and can correspond to the mental model. Additionally, according to Aigner et al. [9], visualizations have the advantages to facilitate new insights as well as enable knowledge discovery and help human in their decisions. Another significant aspect is that visualizations are particularly independent of users experience and education according to Hartmann and Bauer [10]. Therefore, lecturers or students who want to enroll can use the visualization of curricula independent of its prior knowledge.

Several research works in this field show a considerable increase of these kinds of visualizations to satisfy demand. For example, Sommaruga and Catenazzi present in their paper [3] bachelor curriculum visualization in 3D. The virtual environment contains simple geometric modules and each department has its own area, which is divided in six categories (the bachelor curricula are organized in six semesters). Each module of the curricula is presented as block and the size of this block depends one different attributes. For example, the height of the module shows the number of credits. A preliminary evaluation shows that unfamiliar users get lost and some information is not easily readable. However, the learning time is very low and the interface design is intuitive. Another interesting work is described by Kabicher et al. [2]. They present an e-learning project called "Active Curriculum for Computer Science (ActiveCC)" where a web application supports the coordination of modules and their courses for lecturers. In addition to the textural description of the curricula, a graph-based visualization is used to represent the curriculum structure and module dependencies. Modules are visualized as rectangles and the colors of the rectangles depend on their uppermodule. The relationships between modules are visualized as arrows. For example, Fig. 1 shows the submodules of the compulsory module and the modules with the color green are the submodules of the module "Information Technology".

Another good possibility to visualize curricula is ontologies. They represent the structure as well as the relationship between courses clearly and allow one to share, exchange, reuse, analyze and extend curricula. Another advantage that ontologies are becoming popular in various communities is that they provide a common understanding of the domain which can be communicated between people and applications [11].

Generally, an ontology is usually used in philosophy and from this point of view, an ontology is the science of what is, of the kinds and structures of the objects, properties

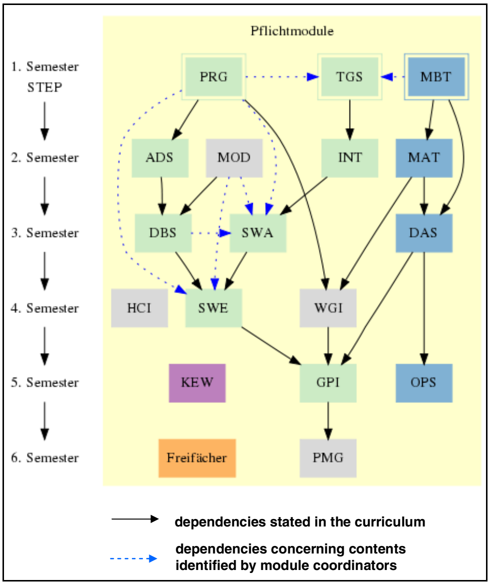


Fig. 1. The graph shows the compulsory module of the bachelor curricula of the University of Vienna and its dependencies (Source: [2])

and relations in every area of reality [12]. In the context of knowledge management, there are different definitions of ontologies. For example according to Fensel [11] an ontology is an explicit conceptualization for a specific domain. Swartout et al. wrote in [13] that an ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.

Ontologies contain classes, which present the relevant concepts, and instances, which stand for the real objects of the class. Furthermore, class hierarchy and interrelationships present the structure of the ontology. The concept depends on the given context represented by the ontology. In our case, we use an ontology to visualize the bachelor curriculum of computer science in business administration which is useful as a foundation for further curricula visualizations. This visualization allows users to better understand the dependences between courses and modules and gives an overview about the structure, which is very essential for the orientation. Furthermore, it is useful to combine the detail view and the general view, as discussed in [14]. Therefore, it is important and necessary that the design of the visualization is useable for the users with respect to their needs.

Several visualization approaches as well as techniques exist to represent ontologies. Katifori et al. present in [15] a survey about different visualization approaches. Several of these approaches as well as techniques are used in other contexts, but the authors point out that the presented approaches and techniques can be adapted for ontology representations. Therefore, for our research we want to find out advantages and disadvantages in regard to their usability of different visualization approaches. We analyze three representative visualization techniques: OntoViz [16] (node-link approach), Treemap [17] (space-filling approach) and Jambalaya [18] (zoomable approach). For a better comparison between the different techniques (as proposed in [19]) we use the same data set

This paper is structured as follows. First we start in Section 2 with the description of the ontology which represents the computer science bachelor curriculum. In Section 3 we analyze three different visualization techniques, which are applied the afore defined ontology. Finally, in Section 4 we discuss the requirements which ontology visualizations should fulfill in regard to users' needs.

2 Structure of the Ontology for Computer Science Curriculum

This section presents the structure of the ontology, which describes the bachelor curriculum of computer science in business administration. The curriculum is modeled as a lightweight ontology, because in our case we need particularly a hierarchic structure, which reflects the modular structure of the curriculum. To make the structure more understandable, it is important that our ontology contains other relationship types to show the different dependencies inside the curriculum.

The basic framework of the ontology contains classes to describe the modules, and instances of the classes to characterize the courses and different types of relationships between courses. Furthermore, each module can contain other modules or a set of courses (see Fig. 2 for an example). In addition to the modules, two further classes are defined to assign for every course their course type and in which semester the course is held.

Beside the *subClassOf* relationships, other relationships are defined to show the dependences between each course and interrelations between courses, course types and semester (Fig. 3 shows an example in regard to the different interrelationships).

In addition, annotation relationships are used to give further information about the semester periods per week and the ECTS points between course types and courses.

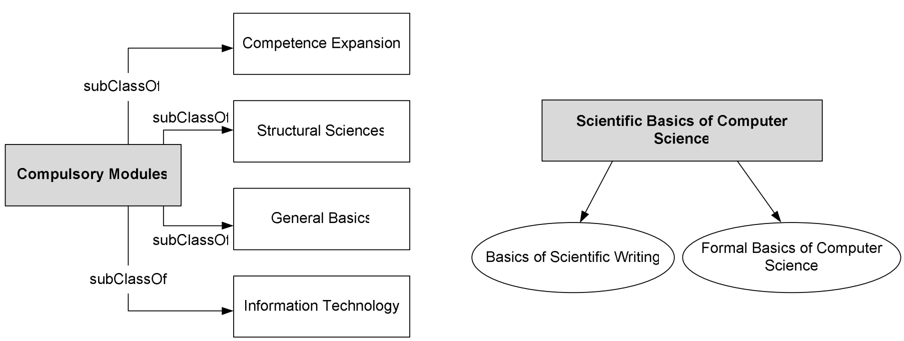


Fig. 2. The left diagram shows the module Compulsory Modules and its submodules Competence Expansion, General Basics, Information Technology and Structural Sciences. The right diagram shows the module Scientific Basics of Computer Science, which contains the courses Basics of Scientific Writing and Formal Basics of Computer Science.

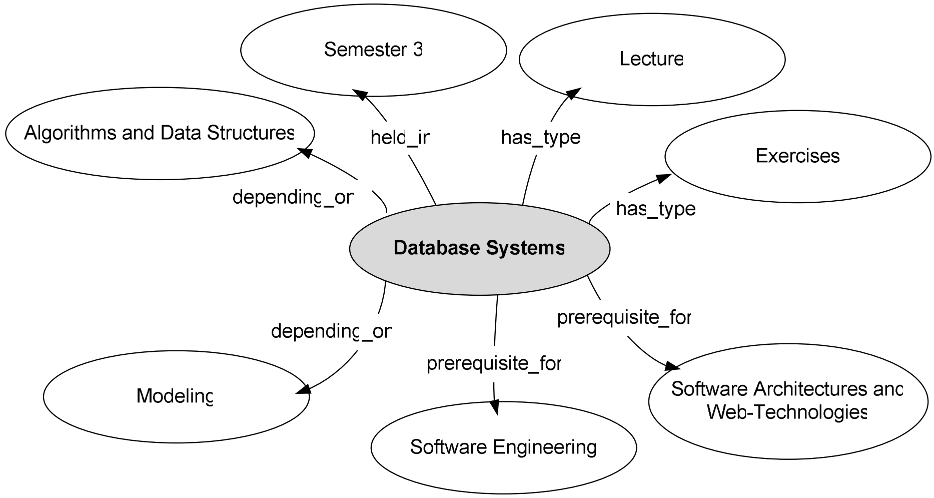


Fig. 3. The representation shows the relationships of the course *Database Systems*. It shows that the course is split in lecture and exercises and is held in the third semester. Furthermore, it is prerequisite for the courses *Software Engineering* and *Software Architectures and Web-Technologies* and depends on *Algorithms and Data Structures* and *Modeling*.

3 Visualization Techniques

We analyzed and compared three visualization techniques according to the ontology of the curriculum, which we defined in Section 2. The techniques are representatives of different visualization approaches – OntoViz for node link approach, Treemap for space filling approach and Jambalaya for zoomable approach (especially nested graphs).

3.1 OntoViz

OntoViz is a plug-in for Protégé and uses the node link approach to visualize ontologies. This approach presents classes and instances as nodes and the relationships between them as links. We selected this visualization technique because node link visualizations show the relationships between the nodes in a clear manner and generally give a good overview of the hierarchy.

Fig. 4 shows the visualization of the whole curriculum ontology, which is described in Section 2. Modules and courses are shown as rectangles and additionally OntoViz uses different color-coding (modules are black and courses are red) for a better differentiation. The *subClassOf* relationships are colored in black and the other relationships are blue, thereby OntoViz does not make further differentiations between the other defined connections. Therefore, the annotation relationships as well as the relationship, which shows the dependencies between the courses, have the same color.

Furthermore, the users can zoom into areas of interest via right clicking on the graph visualization. In addition, they can select single classes or instances via a configuration panel to visualize specific parts of the ontology (see Fig. 5 for an example).

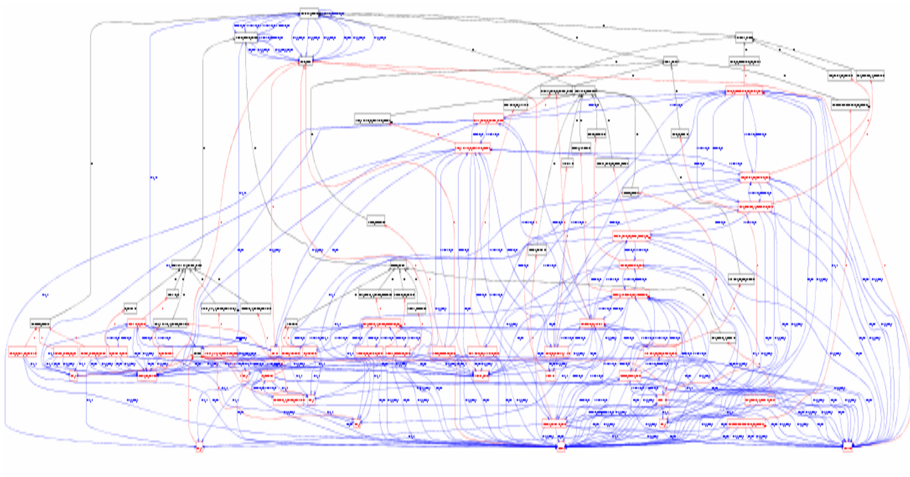


Fig. 4. The representation shows the whole ontology of the bachelor curriculum of computer science in business administration which is visualized in OntoViz

Advantages. The node link representation directly shows the relationships between courses or modules. The direction of the connection is shown via arrows. The hierarchical structure of the curriculum ontology is visible due to the top-to-bottom alignment of the modules.

Another advantage is that the node link approach is familiar for users and it is an intuitive representation to show the connections between modules, their submodules and their courses.

The differences between modules and courses or between *subClassOf* relationships and the other relationship types are color-coded and therefore users can scan the ontology very fast. Furthermore, to assure that the contrast between them is high enough, the colors are selected in a way that the values of the color are far away to each other in the color spectrum. This helps users to effectively distinguish between the different types.

The additional possibility to select specific classes or instances, gives user the choice which parts of the ontology is more interesting for them. Therefore, they can hide parts, which are not important at the moment, to use the whole available screen space to visualize the selected subgraph.

Disadvantages. OntoViz is inefficient in regard to screen space, especially if larger graphs are visualized and therefore it happens that the graph is overcrowded.

It is necessary to zoom out to represent the whole ontology visualization, but then it is not readable any more (Fig. 4 shows an example). Furthermore, the focus on a specific module, course or relationships is lost during zooming. Alternatively, the users can navigate inside the visualization via scrolling but then only a part of the ontology is visible. It could happen that the users get lost in the graph if only a part of the graph is visible and it is difficult to see the whole picture with its relations.

If the links between the modules or courses are long, it is difficult to follow them and to distinguish between the different relationships.

Furthermore, the labels of the relationships are not clearly attributed to the connections. For example, in Fig. 5 the label *prerequisite_for* could stand for the connection between the courses *Software Architectures and Web-Technologies* and *Introduction to Programming* or for the connection between the courses *Introduction to Programming* and *Software Engineering*.

The visualization technique only distinguishes between the *subClassOf* relationships and other relationship types, but it is also essential for an effective scan to make further differentiations. For example, if the user wants to know how the courses depend on other courses, it is necessary for an effective search that these relationships are distinguishable from the other relationship types.

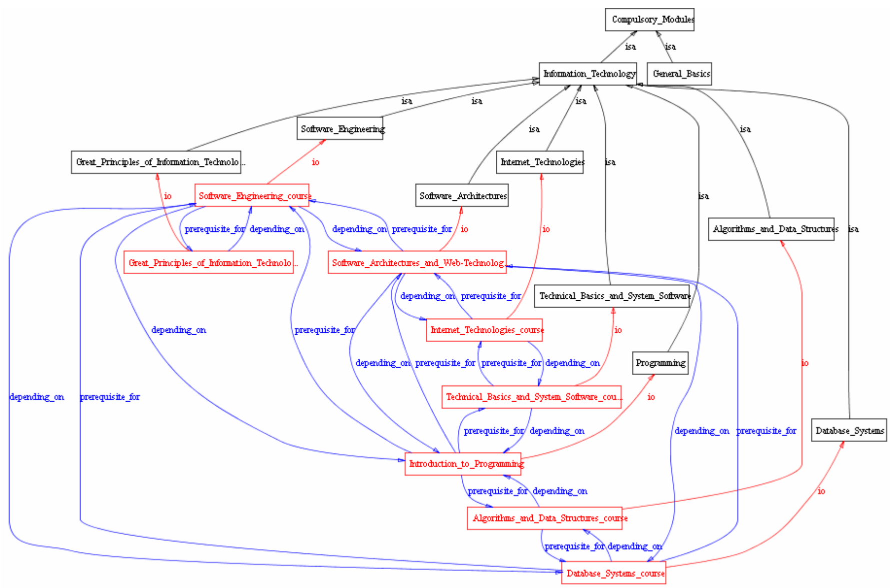


Fig. 5. Visualization shows the subgraph for the module *Information Technology* with its submodules and its courses. The *subClassOf* relationships and the relationships to show the dependencies between the courses are visible.

3.2 Treemap

Treemap is an alternative to the traditional tree visualization to represent the hierarchical structure. According to Shneiderman [17] the layout of Treemaps is based on a space-filling approach to use the whole screen space via subdivision in regard to nodes and their children. Every node is represented as rectangle and depends on its children and the rectangle is subdivided into smaller rectangles via subdivision methods (see e.g. [20]).

We used Treemap 4.1 for our analysis and converted the ontology of curriculum to this format. This result is shown in Fig. 6. The reason to choose this visualization technique was that Treemaps use the available screen space in an effective way and they are suitable for hierarchical structures. Additionally, Treemap 4.1 allows attributes to each node for annotation purposes.

It was important to compare the original structure of the ontology to get a meaningful analysis of these techniques. Therefore, submodules are incorporated in their modules and courses are leaf nodes. The classes, which describe the course types and semester information, are represented as attributes for every course. Furthermore, only the *subClassOf* relationships are directly visualized in Treemap 4.1. The relationships, which show the dependencies between course and the annotation relationships for semester periods per week and ECTS points, are also described as attributes.

The course types are color-coded, for example, the color dark blue shows the courses, which are exercises. The size of the rectangles of the courses depends on in which semester the course is held. Tooltips show basic information (course name, module name of the course, course type and semester) and zooming with the mouse gives the user the possibility to jump to the courses of interest.

Advantages. One advantage of this visualization technique is that the whole structure of the curriculum is visible at first glance and no available screen space is lost. The users can rapidly jump between the modules and courses without losing the orientation.

Furthermore, Treemap 4.1 clearly shows which modules contain many or few submodules and therefore users gain a fast overview about the internal structure of the modules. For example, the Treemap 4.1 visualization of the curriculum in Fig. 6 shows that the module *Competence Expansion* has no further submodules in contrast to the module *Information Technology* which contains eight submodules.

Additionally color-coded rectangles and different rectangle sizes between the courses help users to scan very fast the course type or semester of their interests. For example (see Fig. 6) the user can see that the submodules of the module *Information Technology* contains course which is held in semester 1 to 5 and that the course type semester (colored light green) is only found in the submodule *Great Principles of Information Technology*.

The additional information, which can be added via attributes, gives the visualization the possibility to specify the modules and courses in more detail.

Disadvantages. In our example, we only visualize the bachelor curriculum of computer science in business administration. However, the bachelor curriculum contains further modules of the other specializations like medicine, media, scientific computing or bioinformatics. This has a negative effect on the overview, because the number of rectangles increases in order to show all modules, submodules and courses and therefore no space remains for internal nodes. Therefore, further navigational techniques are necessary to focus on the modules or courses of interest in more detail via semantic zooming.

Furthermore, if more modules and courses are visualized, it is more difficult to find specific courses, especially when the users are not familiar with the modules of the courses or the general structure of curriculum is unknown.

It is difficult to distinguish between semesters according to the number of courses if this information is only visible based on the rectangle sizes. Moreover, the size for each value of the attribute is not consistent in every module. The rectangle sizes can vary inside every module and therefore the users have to find out the correct values for the rectangle sizes for each module.

Only *subClassOf* relationships are represented visually, but in many cases additional types of relationships are important to visualize as direct connections between elements. For example, the dependencies between the courses are only shown as attribute and therefore the users have to search separately for every course, which is listed in the attribute field, to get more information (e.g. in which module the course

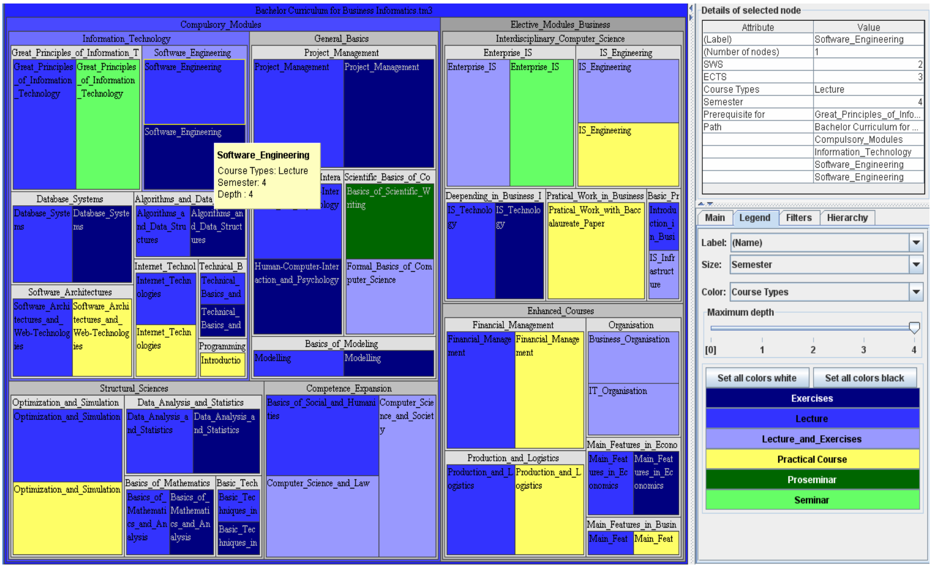


Fig. 6. The representation shows the bachelor curriculum of computer science in business administration which was generated with Treemap 4.1. The right side of the window shows a list of the different attributes and a configuration panel. In this case, the settings for the visual cues are shown. The left side shows the Treemap visualization and the course *Software Engineering* is selected. Additionally, the tooltip of this course is visible.

is or which course type the course has). However, human short-term memory is limited and it is difficult for users to keep much different information at the same time.

3.3 Jambalaya

Jambalaya like OntoViz is a plug-in for Protégé and visualizes the ontology as nested graphs. Generally, nested graphs are suited for hierarchical structures. The nodes of a nested graph are classes or instances and classes can contain other classes or instances (Fig. 7 shows the curriculum ontology as nested graph). For a clear differentiation, classes and instances have different colors.

It is similar Treemap because both approaches nest the submodules or courses in their modules. However, the nested graphs of Jambalaya do not use the available screen space as effective as the space filling approaches. In contrast to Treemaps, the relationships between the courses are directly visible as arcs and every relationship type has its own color.

Jambalaya allows user to zoom into the child nodes to show more detailed information. When the user starts Jambalaya, he/she only sees the classes of the top level. In our example, the classes *Compulsory_Modules*, *Elective_Modules_Business*, *Course_Types* and *Semester* are firstly visible. After the user clicks on a class, the next level of this class as well as the corresponding relationships are visible. For example if the user clicks on the class *Compulsory_Modules*, he/she sees the classes *Competence_Expansion*, *General_Basics*, *Information_Technology* and *Struc-*

tural_Sciences. This is repeated until the user is on the last level, which usually shows the instances. Fig. 7 shows an example when all levels are open.

Advantages. One advantage is that hierarchical relationships between modules and their submodules are clearly visible because of their nested structure. Therefore, the users can distinguish, according to their structure, which modules have more as well as fewer submodules. Furthermore, they can get a picture of modules, which are more extensive according to the number of courses. Moreover, the other relationship types are visible as arcs and so the interdependent connections are obvious at first glance with less cognitive effort.

Based on the fact that users can immerse oneself in the visualization to get more detail information, Jambalaya follows the main principle "*Overview first, zoom and filter, then details on demand*" [21] for visual design which is defined by Shneiderman. The user can decide how much information he/she wants to see without loosing the general overview.

Different colors for classes and instances give the user the possibility to scan fast between modules and their courses. For example, in Fig. 7 the courses, which have the color purple, are easy to distinguish from the modules, which have the color yellow. Another advantage is that the different relationship types have their own color and therefore the user can directly see which kind of connections both objects have.

Like Treemaps, the whole structure of the ontology is visible and the user can rapidly jump between the different modules without losing the complete overview.

Disadvantages. If the user opens many classes with many connections to other classes or instances, it is very difficult to distinguish between the specific relationships and furthermore they constraint the view. Fig. 7 shows that the differentiation

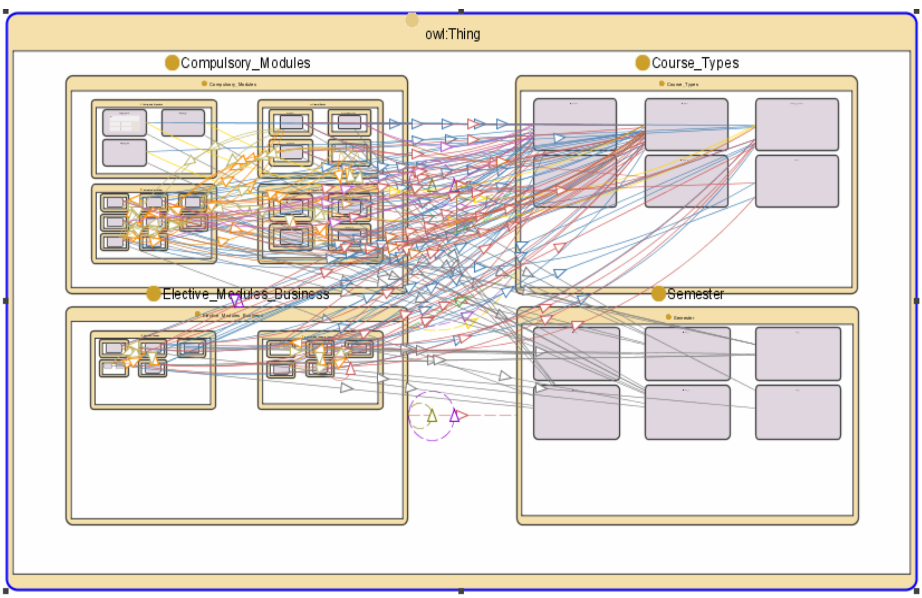


Fig. 7. This visualization shows the bachelor curriculum of computer science in business administration ontology as nested graph. In this example, all levels are open to see additionally the whole structure and relationships between modules and courses.

between relationship types is not possible or rather very difficult. Moreover, the visible connections constrain further interaction in the visualization, because if the connection overlaps classes, it is very difficult to select the class without clicking the connection before. For example, in Fig. 7 it is not possible to select courses from the submodule *Structural_Sciences*, which is in the bottom right corner in the module *Compulsory_Modules*.

The labels of classes or instances are often difficult to read, because they are sometimes too small, overlap with each other or are not visible. If the users want to know the names of the modules or courses, they have two possibilities. The first possibility is that the user reads the names via tooltip, which occurs when he/she moves the mouse cursor over the modules or courses, and the second possibility is via zooming.

Furthermore, the visualization technique is not adapted well to users who are not familiar with the curriculum structure or for users who do not know exactly which module contains the desired course. They have to select every module to jump to the next level and have to repeat this until they have found the specific course.

Another drawback is that the visualization technique does not use the available screen space in an effective way. For example in Fig. 7 the module *Elective_Modules_Business* contains two submodules, however the space inside the module is not utilized fully according to the sizes of its submodules.

4 Ontology Visualization Requirements

In this section, we discuss ontology visualization requirements with focus on users' needs based on the analysis of the three visualization techniques from Section 3.

4.1 General View of the Structure of the Ontology

A general view of the whole structure is particularly necessary for ontologies if they have multiple subtrees and levels. They provide a better orientation and conduce to understand the structure of the ontology. Furthermore, a clear visualization of the whole structure helps users to find classes as well as instances more easily to support a quick scan with less cognitive effort. An overview allows to compare subtrees effectively and the users can see further information, for example the number of subclasses or the levels of a class.

4.2 Visibility of Classes and Instances

Besides the general view, it is important that users can see very fast and clearly every class, because they are the classification of the ontology. Such a view helps the user to find specific instances as well as to understand the whole structure of the ontology. Instances present the real objects in the ontology, for example courses or course types. Therefore, they frequently play an important role for the user and a clear visibility of instances is essential to find them quickly.

Furthermore, it is necessary that classes and instances are visually distinguishable from each other and the names of the classes and instances are readable to support an effective search. For example, the use of different colors for classes and instances is an effective way for an easy differentiation. A high contrast between the colors further aids the visual distinction.

4.3 Visibility of Relationships

Connections between classes and courses describe their relations and a visual representation of interconnections makes it more comprehensible for the user. Additionally to the *subClassOf* relationships, which show the hierarchical connections, other relationship types can be defined. Therefore, it is essential to make a visually distinguishable classification between the relationship types, so that the user can directly see the different interconnections in the visualization. Also in this case, the use of different colors for different relationship types enables an easier distinction according to their classification.

Table 1. Comparison between OntoViz, Treemap 4.1 and Jambalaya based on the ontology visualization requirements. + if the requirement is complied, **-** if the requirement is not complied and ~ if the requirement is partially fulfilled

Requirement	OntoViz	Treemap 4.1	Jambalaya
Visibility of structure			
General view	+	+	+
Detail view	+	+	+
Effective combination	~	~	~
Visibility of classes			
Readability	~	+	~
Color-coded	+	-	+
Visibility of instances			
Readability	~	+	~
Color-coded	+	-	+
Visibility of relationships			
Readability	~	~	~
subClassOf relationships	+	+	+
Differentiation between types	~	-	~
Color-coded	~	-	+

4.4 Result of the Comparison

Table 1 shows the result of the comparison between the three visualization techniques according to the ontology requirements. The classification and allocation is based on the analysis.

OntoViz, Treemap 4.1 and Jambalaya show the general as well as the detail view clearly. However all techniques have drawbacks in regard to the combination of these both views.

Zooming out in OntoViz can affect the readability of classes and instances and in Jambalaya, users have to use tooltips or zooming to read the label names of classes and instances, especially for the deeper levels.

OntoViz makes only a clear visible differentiation between *subClassOf* relationships and other relationship types via color-coding. In Treemap 4.1 only the hierarchical connections are visible, the other relations has to be defined as attributes. Jambalaya shows *subClassOf* relationships as well as other relationship types and uses different colors to make the other defined relationship types distinguishable. However, if the ontology contains too many interrelationships, the visualization appears overcrowded and it is difficult to differentiate between single relationships.

5 Conclusions

Often curricula are only described textually, but to end users like students or lecturers who want to see dependences between courses or modules, a visual representation is essential.

An ontology is a good possibility to show the structure of curricula and their dependences clearly and allow – among other things – to reuse and to extend the curricula easily. The last mentioned attributes are especially necessary for computer science curricula, because computer science is grown in the last years in regard to new subfields or the existing subfields are extended their knowledge. For our analysis, we defined an ontology for the bachelor curriculum of computer science in business administration which is useful as a starting point for further curricula visualizations.

Generally, there are several different approaches to visualize ontologies and therefore we compared three different visualization techniques (OntoViz for node link approach, Treemap for space filling approach and Jambalaya for zoomable approach) to find out which requirements are useful for ontology visualizations in regard to users' needs. The comparison shows that the most requirements are complied by Jambalaya. The result of OntoViz is similar to Jambalaya. However, Jambalaya clearly shows different relationships via color-coding. Treemap 4.1 satisfies many requirements but the significant difference to Jambalaya and OntoViz is in the visibility of relationships. Jambalaya and OntoViz have the advantage that they additionally show other defined relationship types. Nevertheless, if the ontology has many interrelationships, the visualization appears overcrowded and it is not easy to differentiate between single relationships.

The comparison shows that there is no specific technique that is appropriate for all requirements. Therefore, it is important to choose or combine different visualization techniques depending on users' tasks. The results of this research will be an inspiration to propose visualization methods aimed at meeting users' needs and tasks.

5.1 Outlook on Further Work

Furthermore, an extensive evaluation with user studies is planned to find out more about the need of this kind of visualizations. Pohl et al. present in their paper [22] the significance about students' attitudes in e-learning. This is also a relevant topic according to curricula visualization and therefore further research about students' attitudes as well as lecturers' attitudes is necessary. We intend to use survey and interviews to find out more in this area.

Acknowledgments. I gratefully acknowledge the critical feedback of Günter Wallner from the University of Applied Arts Vienna and Renate Motschnig from the Research Lab for Educational Technologies at the University of Vienna.

References

1. Commission, E.: The Bologna Process (last access: 2008-06-03), `http://ec.europa.eu/education/policies/educ/bologna/bologna_en.html`
2. Kabicher, S., Derntl, M., Motschnig-Pitrik, R.: Towards an Active Curriculum for Computer Science. In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications, Vienna, Austria (2008)

3. Sommaruga, L., Catenazzi, N.: Curriculum visualization in 3D. In: Proceedings of the Twelfth international Conference on 3D Web Technology, Perugia, Italy (2007)
4. Rubin, B.S., Misra, B.S.: Creating a Computer Security Curriculum in a Software Engineering Program. In: Proceedings of the International Conference on Software Engineering (2007)
5. Kay, D.G., Hoek, A.v.d., Richardson, D.J.: Informatics: a focus on computer science in context. In: Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education, St. Louis, Missouri, USA (2005)
6. Geroimenko, V., Chen, C.: Visualizing the Semantic Web: XML-Based Internet and Information Visualization. Springer, Berlin (2003)
7. West, T.G., Uhl, J.J., Martinez, J.: Seeing the unseen: visualization in education. In: ACM SIGGRAPH 2006 Educators Program. ACM, Boston (2006)
8. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
9. Aigner, W., Bertone, A., Miksch, S.: Tutorial: Introduction to Visual Analytics. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 453–456. Springer, Heidelberg (2007)
10. Hartmann, F., Bauer, E.K.: Bildersprache. Otto Neurath - Visualisierungen. Facultas Universitätsverlag (2002)
11. Fensel, D.: Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer, Berlin (2001)
12. Smith, B.: Ontology. In: Floridi, L. (ed.) The Blackwell Guide to the Philosophy of Computing and Information, pp. 155–166. Blackwell, Malden (2003)
13. Swartout, B., Patil, R., Knight, K., Russ, T.: Toward Distributed Use of Large-Scale Ontologies. In: Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW 10), Banff, Alberta, Canada (1996)
14. Kriglstein, S., Motschnig-Pitrik, R.: Knoocks: New Visualization Approach for Ontologies. In: Proceedings of the 12th International Conference on Information Visualization London (2008)
15. Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology visualization methods—a survey. ACM Computing Surveys 39(10), 10 (2007)
16. Sintek, M.: Ontoviz Tab (last access: 2008-06-04), `http://protege.cim3.net/cgi-bin/wiki.pl?OntoViz`
17. Shneiderman, B.: Tree visualization with tree-maps: 2-d space-filling approach. ACM Transactions on Graphics 11(1), 92–99 (1992)
18. Storey, M.A.D., Wong, K., Fracchia, F.D., Mueller, H.A.: On Integrating Visualization Techniques for Effective Software Exploration. In: Proceedings of the IEEE Symposium on Information Visualization (InfoVis 1997). IEEE Computer Society Press, Los Alamitos (1997)
19. Plaisant, C.: The challenge of information visualization evaluation. In: Proceedings of the Working Conference on Advanced Visual interfaces (AVI 2004), Gallipoli, Italy (2004)
20. Bruls, M., Huizing, K., van Wijk, J.J.: Squarified Treemaps. In: Proceedings of the Joint Eurographics and IEEE TCVG Symp. on Visualization, TCVG 2000 (1999)
21. Shneiderman, B., Plaisant, C.: Designing the User Interface, 4th edn. Addison Wesley, Reading (2004)
22. Pohl, M., Herbst, I., Reichl, F., Wiltner, S.: Students' Attidudes Towards Novel Interfaces in E-Learning. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 738–747. Springer, Heidelberg (2007)

Agile User-Centered Design Applied to a Mobile Multimedia Streaming Application*

Zahid Hussain[1], Martin Lechner[1], Harald Milchrahm[1], Sara Shahzad[1], Wolfgang Slany[1], Martin Umgeher[1], and Peter Wolkerstorfer[2]

[1] Institute for Software Technology, Technical University Graz, Austria
m3@ist.tugraz.at
[2] CURE - Center for Usability Research & Engineering, Austria
wolkerstorfer@cure.at

Abstract. Mobile computing is leading a revolution. Multimedia consumption on mobile devices is increasing day by day. The most important factor for the success of such applications is user acceptance. Additionally, the success of a software development project is associated not only with tools and technologies but also depends on how much the development process is both user-centered and developer-oriented. We are working on a project to develop a multimedia streaming application for mobile phones. The paper describes our adopted development process: the integration of Extreme Programming (XP) – one of the popular agile methods – with User-Centered Design (UCD) and shows how the integrated process facilitates user-orientation and at the same time preserves the social values of the development team. The paper also presents a summary of a recently carried out usability study.

Keywords: Agile Methods, Extreme Programming, Usability, User-Centered Design, Mobile Application.

1 Introduction

The most important factor for the success of a software application is user acceptance. "An inherently usable and technically elegant application cannot be considered a success if it does not satisfy the end-users' needs. End-users are often left out of the development process" [1]. Agile development processes involve a customer as a business representative who is responsible to specify the business value of user requirements, but this customer needs not necessarily to be a real end-user.

Agile methods are becoming popular nowadays. Being a lightweight agile method, Extreme Programming (XP) has advantages of: on-time delivery,

* The research herein is partially conducted within the competence network Softnet Austria (www.soft-net.at) and funded by the Austrian Federal Ministry of Economics (bm:wa), the province of Styria, the Steirische Wirtschaftsförderungsgesellschaft mbH. (SFG), and the city of Vienna in terms of the center for innovation and technology (ZIT).

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 313–330, 2008.

co-located team, relying on the team members' knowledge rather than documentation, optimized resource investments, short release cycles, working high quality software, tight customer integration, incremental design, constant communication and coordination, rapid feedback, continuous refactoring, pair programming, and test driven development [2], [3], and [4]. "XP is a collection of well-known software engineering practices. XP aims at enabling successful software development despite vague or constantly changing software requirements. The novelty of XP is based on the way the individual practices are collected and lined up to function with each other" [4]. It is also a people-oriented process with many social activities.

"Usability measures the quality of a user's experience when interacting with a product or system; UCD is an approach for employing usability" [5]. UCD, also called human-centered design, is an approach to user interface design which is based on information about the people who will use the product. UCD process focuses on users through the planning, design, and development of a product [6]. Holzinger emphasizes that every software practitioner should be aware of different usability methods and apply them according to specific situation of a project [7].

There already exist approaches of integrating agile methodologies and Usability Engineering (UE) / UCD [8], [9], [10], [11], and [12]. Memmel et al. point out that "When UE becomes part of agile software engineering; this helps to reduce the risk of running into wrong design decisions by asking real end users about their needs and activities" [12].

Focus of both methodologies on users makes it possible to integrate them [13]. The integrated process allows to combine benefits of both methodologies and makes it possible to reduce the shortcomings of each because XP needs to know its true end-users and UCD benefits from a flexible and adaptive development methodology which runs throughout the project life-cycle [14]. We integrate XP and UCD in our project [1], where we are developing an application that enables a user to perform content-based search for Audio-Video (AV) contents and play the streamed content on a mobile phone [15]. The end-users are indirectly involved in the process by our use of different HCI instruments like user studies, personas, usability expert evaluations, usability tests, and automated usability evaluations [16]. Usability of a mobile application is an important ongoing research issue. Numerous studies address UE / UCD issues for mobile applications [17] [18] [19]. We conduct various usability studies and in this paper a summary of one of the studies is presented.

The next section describes the similarities between XP and UCD. Section 3 describes the project environment. Section 4 describes the adopted process. Section 5 provides the details of a usability study. Section 6 concludes the paper.

[1] A preliminary version of this paper describing the adopted process has been published in the proceedings of the PPIG 2008, The 20th Annual Psychology of Programming Interest Group Conference, Lancaster University, UK, September 2008.

2 Similarities between XP and UCD

The core values of XP and UCD [20] are applied to solve different issues: In XP, a simple implementation fulfilling the minimum requirements of the application is created and iteratively extended, while UCD tries to continuously improve the usability of the user interface. However, when comparing some of the core values it seems obvious that the two development processes can benefit from each other's practices.

2.1 End-User

One of the core practices of XP is to have a *Customer on Site* who is co-located with the programmers in order to answer domain-specific questions and give feedback on the system. This practice can be matched well with the testing of prototypes with actual users as proposed by UCD if the customer is also the real end-user or when developers are directly exposed to end-users.

2.2 Continuous Testing

Constant and extensive testing is at the heart of XP. It is mainly embodied by two practices: *Continuous integration* runs all existing automated tests whenever the code base is changed or extended in order to check if the changes caused any undesired side effects. Most of these tests emerge from *test-driven development*: in the first run, automated tests checking the desired behavior are created; then the actual behavior is implemented and can right away be evaluated with the tests. This usually is done only for pure behavioral code but can be extended to user interfaces: tests can check the expected behavior of an interface, and these tests can be run whenever the code is changed.

The end-user tests of UCD are a valuable source for test targets: an unexpected user action that caused a problem in the application can be replicated as an automated and continuously evaluated test to ensure that the problem, after solving it once, does not reappear.

2.3 Iterative Development

Both XP and UCD propagate an iterative procedure [20] [2] of design and development [6]. An XP project yields *small releases* (another core XP practice) on a regular and frequent base (usually a few months). Each release version is based on the previous one, incorporating new features and fixing bugs of the predecessor. Inside a release time frame, work is organized in "iterations" (usually taking one to four weeks). On an even smaller scope, many feedback-and-change cycles take place, especially in conjunction with test-first development and *refactoring* (the practice of changing source code in order to improve its quality without changing its functionality).

UCD also proposes a design–test–modify circle for developing user interfaces. The scope of iterative development in XP and UCD differs: releases and iterations in XP are mainly organizational units, while refactorings are just considered a development tool; on the other hand, UCD's iterative user interface refinement is a more explicit process as its involvement of external persons (the test users) makes it more complex. Nonetheless, iterative interface development of UCD fits well into the iteration principle of XP because both approaches are aware of the value (and necessity) of evolutionary development.

3 Project

We are working in a project where we develop an application that enables a user to perform content-based search for AV content and play it on a mobile phone. The project started in summer 2007 and will end in 2010. The application enables a user to search not only in the metadata but also in the spoken words of the AV clips. This content includes radio and TV archive material, like documentaries or other recordings of historical, political and cultural importance, discussion programs, movies, music videos, audio books, etc. The application is being designed keeping in mind the social interaction of users and provides some Web 2.0 features [21].

In addition to this, one goal of the project is the analysis of agile software development methods, particularly XP, and to devise a usability test procedure for mass applications on mobile devices with emphasis on UCD and iterative user-interface design.

The team consists of six full-time regular members having different social and cultural backgrounds, five developers and a product manager who plays the role of the "on-site customer" from XP.

The customer communicates with the project partners who come from various domains, including UI design, usability research, telecommunication, content providing, and software-hardware infrastructure. Developers also directly communicate with the engineers of a partner usability research center regarding usability issues. The usability engineers working for our project are also active in UCD research with the team.

3.1 Application Features

The user interface of the current prototype comprises the following main features. The application's main screen provides the features: "Search", "Top rated", and "Most recent" clips. It provides links to the "Channel" and "Media Feeds" pages as well as a link to the "Clip Detail" page when one clicks on the title of a clip. The application also implements a few other Web 2.0 features like "Recommended" and "Most viewed" clips.

Search: "Search" allows to search the whole AV content by entering keywords. It displays the search results ordered by broadcast date (if any). For each result item, the clip's title, link, description, duration, originating channel (if any), and a representative thumbnail image are shown. The user can play a clip by clicking on the respective link.

Channel: "Channel " allows to browse the schedule of TV and radio channels. A channel lists the original program schedule of the current day, but users can browse to the previous day's schedule or can search within the channels. Users can select the date and the time (either Morning, Afternoon, Prime time, or Night) where the system then displays the list of clips in the selected time period, ordered by broadcast time. The resulting items are shown in a similar format as in the "Simple Search" result list with rating stars and a channel icon for each clip in the schedule. The top of the page also provides a dropdown list for selecting channels.

Media Feeds: The "Media Feeds" feature is intended to provide the users with a facility to create and consume a constantly updated stream of clips based on the users' search criteria. This media feed can be sent to a friend by SMS or email.

Clip Detail: The system provides the "Clip Detail" page when users click on the title of the clip. Users can rate a clip and add a comment or view all comments. With the "Tell a friend" feature, users can send a clip to their friend by SMS or by email.

4 Process

The following sub-sections describe the process which is followed for the user-centered design of our application.

4.1 Approach to User-Centered Design

User-interface design plays an important role in the acceptance of a web based application by users. The overall process of our approach to UCD is based on evaluating the usability of the application in small iterative steps. This helps us to gain insights into the functional and cognitive requirements of real users. We design prototypes of the user-interface of the system and test them throughout the development process. As a result the fidelity of the prototypes increases and evolves.

The work flow presented in Figure 1 illustrates the iterative design approach to incorporate UCD into our XP process. From a broad perspective, the application development cycle starts from defining the user stories, then comes to mock-up designing and at the end to the actual implementation. The process is executed as follows:

- Different feature-related user stories of the application are created by the customer along with partners.
- Developers create different paper mock-ups for each of the required features to collect ideas and to present them to the customer.
- The customer decides which of the mock-ups best suits his needs or suggests modifications to the mock-ups.

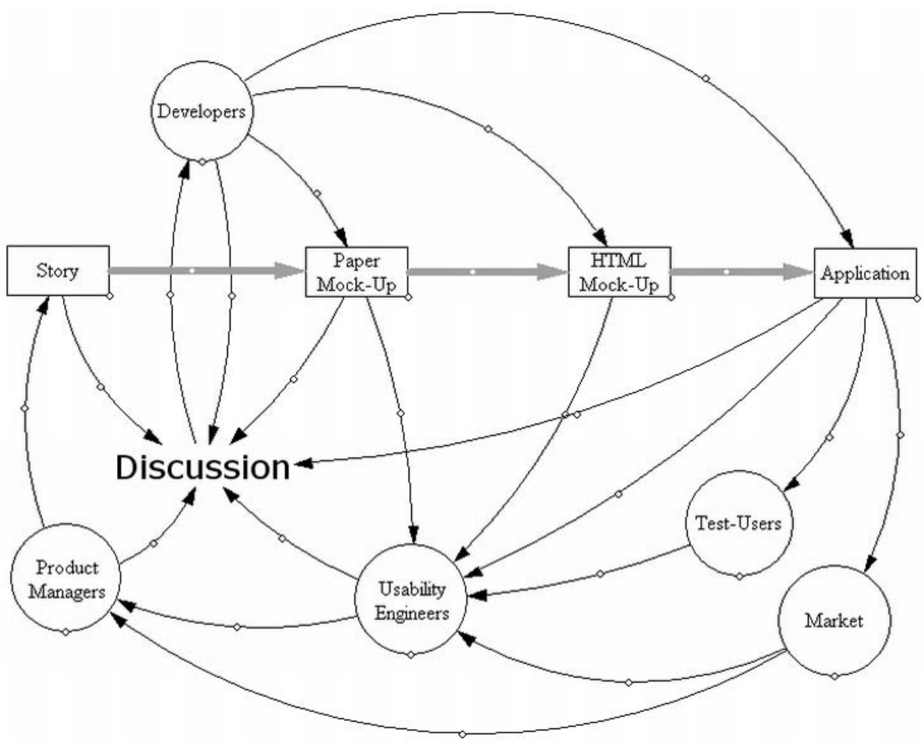


Fig. 1. Iterative UI design workflow [21]

- A final mock-up is derived according to customer's wishes which then serves as the basis for the actual implementation.
- Once an implementation mock-up of a feature or group of related features is finished, the usability engineers are asked to give feedback on it.
- After incorporating the feedback given by the usability engineers into the application, end-user tests are conducted on the application by the usability engineering team.
- The feedback on the application from the usability engineers as well as from the test-users is taken as an input for further refinements in the UI design of the application.
- The results are then incorporated into automated tests (test driven development) which are used as an executable specification for the actual implementation.

This feedback-and-change cycle provides insights into whether the user-interface design is meeting the usability criteria. As the application development is done in short iterations, the developers are able to refactor the system continuously according to the feedback derived from the parallel, as well as iterative, UI design process. Hence, the system evolves according to the needs of the end user and the specifications derived from actual usage.

4.2 Choosing the Type of Mock-Up

We make use of two different types of mock-ups; low fidelity paper mock-ups and high fidelity implementation mock-ups. The benefit of using paper mock-ups for the interaction design is that they can be designed and modified quickly. For simple interaction designs, a low fidelity paper mock-up suffices as a basis for further discussions and the implementation. An additional advantage is that it is easier to criticize simple and rough mock-ups compared to ones which look neat and perfect from the graphic design perspective [22]. But for some features a high fidelity mock-up is required to clearly visualize the interface. As we have the benefit of an on-site customer co-located with the development team all the time, for those tasks a quick implementation mock-up is designed and immediately shown to the customer. That implementation mock-up is then modified based on the immediate feedback of the customer. Would our customer be not co-located with us all the time, it would be difficult to take maximum advantage out of this quick feedback-and-change cycle.

4.3 Frequency of End-User Tests

The end-user tests are made on an on-demand basis, that is, when the customer says that now is the appropriate time from the business point of view to run a usability test with test-users. Also, when there is enough amount of new functionality added to the application it becomes effective to perform usability tests and then only to continue with further development. It would have been wonderful if user tests could be made on regular basis, like at the end of each release, but considering the expenses and resources required for it we have kept it only on an on-demand basis. Therefore, the expensive part of involving real users is done more effectively.

4.4 Integration of HCI Instruments

Figure 2 describes the way how usability engineers have integrated HCI instruments (User studies, personas, extended unit tests, usability tests, and usability expert evaluations) into the XP process. It shows the interplay of the HCI instruments with the XP process. When applied correctly in various phases of the project, the instruments are designed to reach the goal of improved software quality not only in terms of technical quality but also in terms of usability [16]. End-users are integrated in two different ways. On one side, user studies are taken into account to develop personas [23] which then specify direction of development by guiding customer to identify user stories, and also at the end of a development iteration the vision about the users is broadened which helps in extending the personas. This serves as an indirect end-user input for the development process. On the other side, feedback from usability tests performed by test-users as part of the usability evaluations serves as a direct input for further enhancement and development of the application [16].

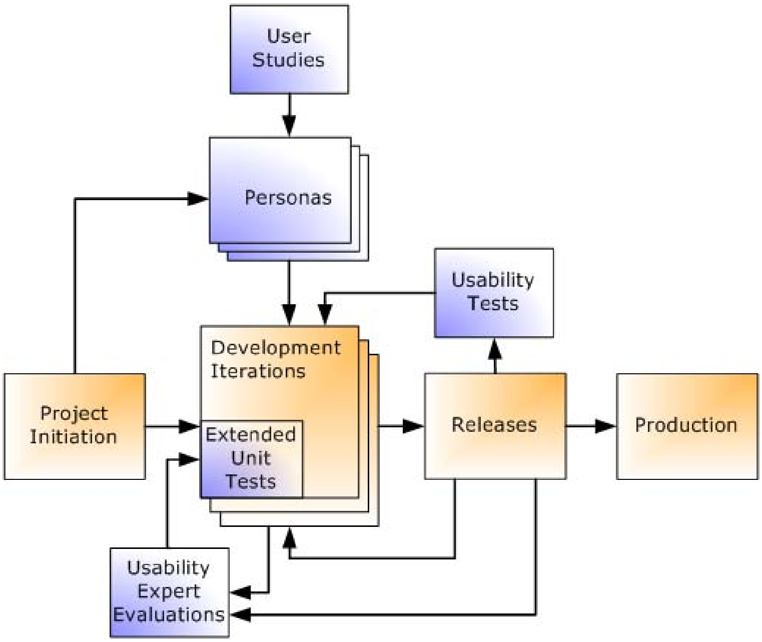


Fig. 2. The integration of HCI instruments into XP [16]

4.5 Testing Issues

A big issue in mobile user-interface design practice is that current approaches are not sufficient for mobile phones [24]. For designing any software, use of good UCD practices ensure that the product is accepted by the users [25] thus supporting the use of UCD approach for UI design. To enhance it further, we provide high fidelity implementation prototypes to our usability engineers for user testing. Paper prototypes are good and sufficient for verifying non mobile-based product requirements, but in case of applications for mobile phones they are not sufficient for finding out and solving usability issues related to detailed interaction on the small device with its limited user input capabilities [25]. Therefore, in our case the application is tested on mobile phones and not on any web based simulator for understanding the interaction issues concerning the use of mobile phone interface [25].

4.6 Communication and Collaboration

Communication between stakeholders is an important characteristic of software development. Communication and collaboration between customers, business partners, developers, and other stakeholders enhance overall team efficiency [1]. The value of communication is expressed by the XP practices of pair programming, metaphor, informative workspace, simple design, on-site customer, the planning game, and coding standards. [26]. Other factors in communication are the use of whiteboards, positioning and sharing of desk facilities to facilitate pair programming, stand–up meetings, developers buying-in to the concepts of the rules and practices of XP, and collective code ownership [27]. We sit side by side in a spacious room having enough space for private workplaces as well as for three separate pairing stations. This seating arrangement has promoted effective interaction in the team and has helped in resolving technical issues on the spot [28]. The teams' XP room is equipped with six whiteboards which are used to record the XP stories agreed at release and iteration planning meetings. Story cards are physically stuck to the whiteboards in prioritized order with adjacent notes written on the board. Various graphs showing architecture and velocity of the project are also drawn on the whiteboards. By looking at the whiteboards, anyone can see the current status of the project.

Email, phone calls, and video conferencing are the tools used in routine communication with the usability engineers and with other partners. Personal visits to and by project partners are also made by the product manager and by other team members whenever necessary.

4.7 The Planning Game

We hold two types of planning meetings: release based meetings and iteration based meetings. A release lasts for three months whereby within a release, an iteration lasts for two weeks. Project partners attend release meetings and through discussions identify and define user requirements called user stories in XP [29]. "The parallel with the UCD approach is obvious here: an understanding and appreciation of the users and their requirements" [1]. These stories are written down on story cards and are prioritized by the project partners. Developers then estimate the time required for implementing the stories.

At the beginning of each iteration, an iteration meeting is held and is attended only by the team members including the product manager. The product manager selects and prioritizes stories which fit in the current iteration depending upon the available velocity. Then, developers break down the stories into detailed tasks and estimate them. Finally, the product manager defines the acceptance criteria for each story and task.

Before and after implementing these stories, continuous feedback is obtained from the usability engineers. Then, these stories are modified according to the feedback of end-users and the usability engineers. Once "again this is a common step with UCD approaches; an understanding of the user goal and the tasks to achieve that goal. Addressing a requirement in terms of the user and their goals focuses development upon what is needed" [1].

4.8 Pair Programming

This practice has helped us in spreading and sharing the project-specific knowledge and improving the technical skills of the developers. We also applied the practice of working in pairs with the product manager [29]. The product manager pairs with a developer for writing customer-acceptance tests, thus exposing the customer to the process and the internal status of the application which helps in better understanding and implementing the end-users' requirements. This also has enhanced the enthusiasm of the team members to work in a collective and collaborative team environment.

4.9 On-site Customer

In UCD, "all activities are focused on providing business value through ensuring a useful, usable and engaging product. The customer is not only defined as the project stakeholder, but the end user as well" [1]. The Manifesto for Agile Software Development [30] does not clearly demand end-users as customers. In our process, the product manager plays the role of an "on-site customer" and communicates with the various stakeholders. The end-users are indirectly involved by the usability engineers who provide the HCI instruments.

5 Usability Study

Usability tests are carried out to evaluate the running prototype. One of the usability studies was executed in January 2008 with 10 respondents using a mobile phone. The classical task-based usability test method was used [31]. Each respondent was asked to execute 5 different tasks. Tasks were carried out on a Nokia N95 mobile phone. To gather general feedback and general opinions, two interviews were carried out: One before and one after the task session (pre- and post-interview). Each task was accompanied by task specific post-questionnaires. Interview sessions lasted about 1 hour. For the tests the device's standard browser as well as opera-mini 3.0 were used (the first is incorporating a web-like mouse pointer, the latter a link marker to navigate through the interface).

After the test, respondents had to judge three different visual design paper-prototypes. We used the AttrakDiff questionnaire [32] to capture the attitudes of the users towards the application in terms of graphical design, enjoyment, and aesthetics. After the tasks, the AttrakDiff questionnaire was filled-out.

The results of the tests suggest two main improvements:

Improvements of Layout and Design: Main improvements should be made concerning the visual appearance of the site:

a) The actual site, menu, and navigation layout is not optimal. Through the use of the color blue as text color and background color at the same time, equal text sizes throughout the interface and different alignments, the site's hierarchy is not visible for users.

b) The actual layout does not incorporate visually attractive design elements and is rated as pragmatic but monotone with a lack of stimulating elements (Attrakdiff questionnaire).

Figure 3 shows the prototype of the home page.

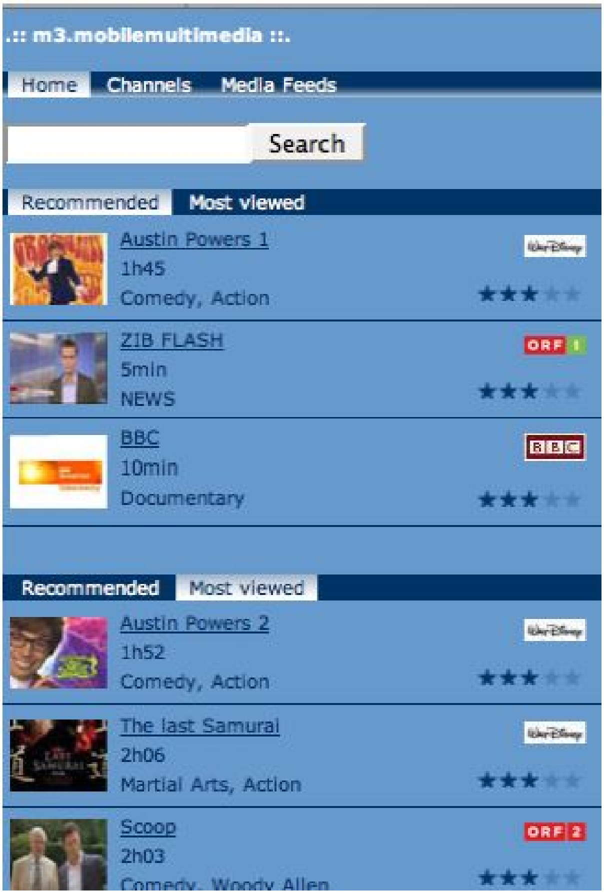


Fig. 3. The prototype of the home page

Improvements of the Prototype's Usability: On the "Channel" web page, a web-like calendar function to select dates should be integrated (the actual function will not be usable for greater amounts of data). All navigation menu elements should be separated from content menu elements ("Home" vs. "Watch"). Furthermore, interactive elements ("Rate", "Comment" etc.) should be placed on a separate page instead of on the bottom of a description page. Figure 4 shows the recommended prototype of the Channel page showing the calender. Figure 5 shows the Menu entries without any visual separation.

For the further development of the prototype the sub-site "Media Feeds" should be separated into two categories introducing the sites "create Media Feed" and "watch Media Feed". Special attention should be given to feedback mechanisms which at the moment do not support the user (feedback of search queries, display of media feed search results).

Fig. 4. The prototype of the Channel page showing the calender

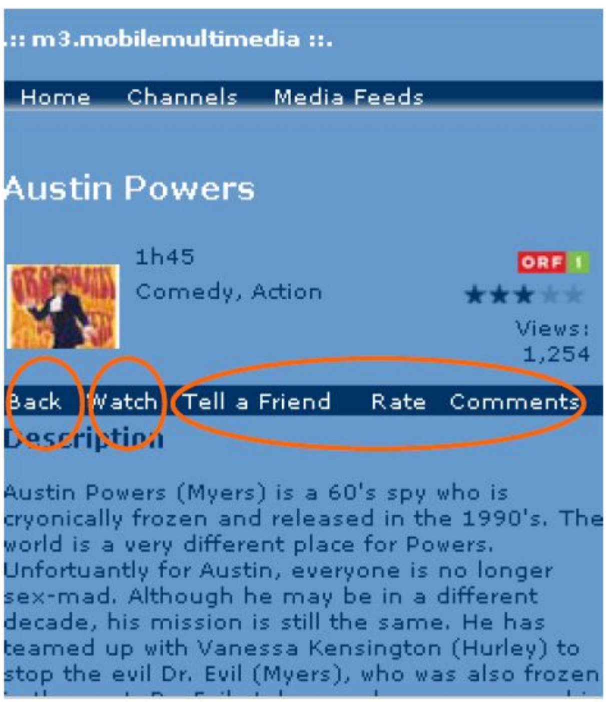


Fig. 5. The menu entries without any visual separation

From the mock-ups of three different designs, the AttrakDiff results suggest that a yellow design was most liked by the respondents. It was also suggested that the blue design may be used in the future but the following improvements should be made.

– Accentuate contrast on whole site.
– Avoid light blue text on darker blue backgrounds.
– Introduce visually attractive design elements that increase the attractiveness of the site.
– Eliminate monotony by introducing more colors.

Two of the developers also observed the usability study session which gave them a chance to realize the impressions of actual end-users and their feelings as well as actual usage. This helped in guiding the development according to the wishes of end-users.

5.1 A Task Example

Here a task example is presented. The task is: "Find the detailed description of a given movie, write a comment and rate it".

Facts on Task: The task was completed without any greater difficulties by all respondents. On the "Home" page and on the "Channel" page respondents used the heading to find the detailed description and the video's thumbnail to watch the video. Respondents did not encounter much problems on the "Clip Detail" page. The prominent position of the links "Comments", "Rate" and "Tell a friend" – Figure 5 – on top of the description page helped respondents to understand which possibilities are offered. On the "Clip Detail" page there are two interaction paradigms that were both understood: Clicking on the link "Tell a fried" opens a new page. This didn't cause any problems for users. The functions "Rate" and "Comment" are placed at the bottom of a "Clip Detail" page and users had to scroll down or use a link to jump down. In reference to both described paradigms, user comments indicate that the longer the list of comments is, the more uncomfortable the site is to browse. Further, the task uncovered that on the mobile interface respondents did not recognize that they were scrolling down the page when using the anchor-links "Comment" and "Rate". To get back to the top of the site they pushed the "back" button. This did confuse some of the respondents as they jumped back to "Home" although their intention was to get to the top of the "Clip Detail" page. Of course this depends on how the browser implements the "back" functionality.

A solution that incorporates interactive functions ("Rate", "Comments", "Tell a Friend") on a separate page is recommended. Furthermore:

– Back Button: A dedicated back button should be integrated on top of the page. This is where basic navigation elements are expected.

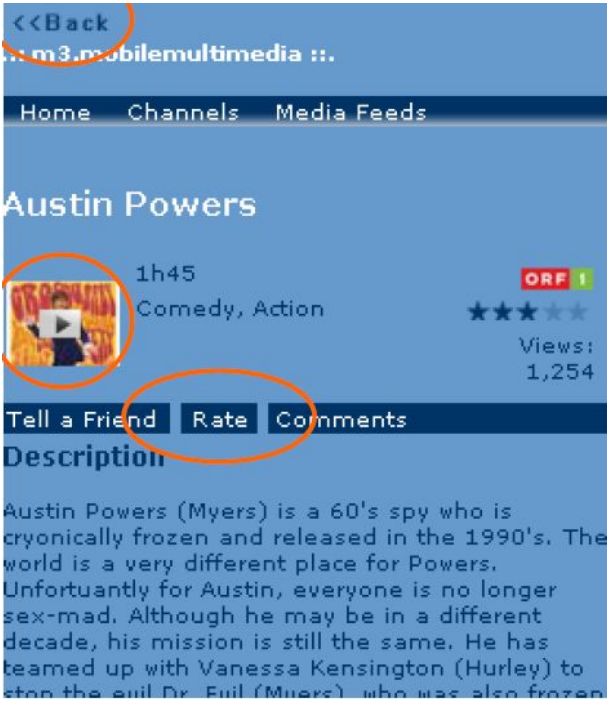


Fig. 6. Improvements of menu layout and arrangement

– Watch Button: A "watch" button should be designed and integrated consistently. An additional watch button – if necessary – should be placed on a particular spot on the site and not be integrated in the navigation menu. The "Watch" button should be visually highlighted.
– Tell a friend, Rate and Comments: These elements describe "interactive functions" on the site and therefore should be kept together and aligned to the left side of the page.

Figure 6 shows the recommended menu layout and arrangement.

Respondents' Feedback/Comments:

– All respondents indicated that in their opinion the "Clip Detail" page provides a good overview.
– The design of the "Comments" and "Rating" section is good and intuitive. Too many comments on one page should be avoided as the page would get too long (1 respondent).
– Comments should be ordered in chronological sequence, beginning with the most recent entry (1 respondent).
– The space on top of the "Clip Detail" page (heading) should be used in a better way. This would provide more space for description texts (1 respondent).

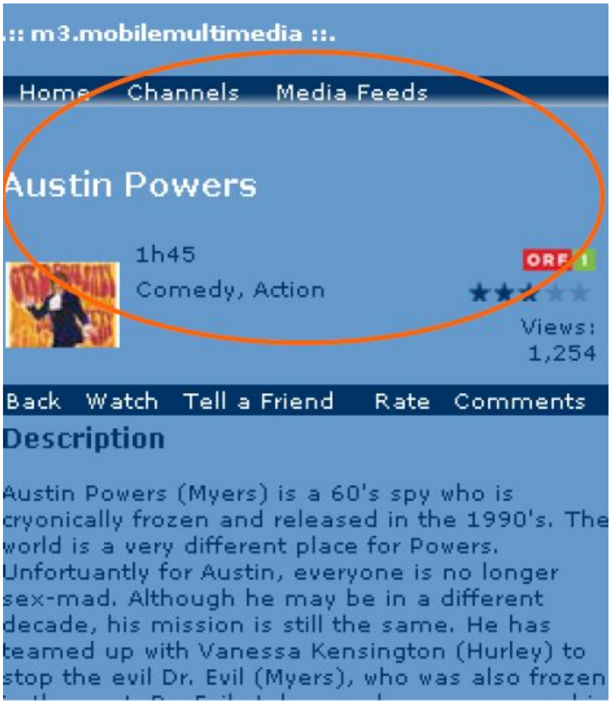


Fig. 7. Use the space on top of the "Clip Detail" more efficiently

– It should be possible to select which information is sent to another person via the "Tell a friend" function (the video's description, the video itself, etc.). Radio buttons should be used to specify one out of different possibilities (1 respondent).

Figure 7 shows the space on top of the "Clip Detail" page which should be used more efficiently.

6 Conclusion

Agility is an invitation to adapt, to mold, and to reshape the software development methodology according to the requirements of a project. Being a lightweight agile process, it is easy to extend the XP process with additional practices. Our XP process fits well into the UCD approach because of the many overlapping principles (iterative development, end-user incorporation, testing) of both methodologies. The usability engineers in the project are well integrated into the development team. It would have been even better if one of the usability engineers had been present physically with the team all the time, as face-to-face communication is more helpful in quickly resolving design issues. We could not conduct usability tests frequently with end-users due to time and budget constraints but mitigated it with usability expert evaluations. The whole development process is influenced when each feature of the application is assessed from users' perspective. This addresses the problems which arise when the system requirements are gathered only by discussions with stakeholders [33]. The involvement of all stakeholders, particularly end-users in the process, can increase the chance of success of a project.

We continuously try to optimize our process as long as the project lasts and will provide further insights whether the process has been able to enhance usability of the application. In October 2008, we will conduct a contextual mobile multimedia content usability study which will give insights into mobile HCI concerning the coherence of content types, consumption times, and consumption contexts. The benefits of the integration of end-users indirectly in the form of HCI instruments, co-location, communication, and planning meetings have contributed to improve not only our process but also led to increase the overall morale of the team.

References

1. McNeill, M.: User centred design in agile application development, http://www.thoughtworks.com/pdfs/agile_and_UCD_MM.pdf
2. Beck, K.: Extreme Programming Explained: Embrace Change, 1st edn. Addison-Wesley Professional, Reading (1999)
3. Beck, K., Andres, C.: Extreme Programming Explained: Embrace Change, 2nd edn. Addison-Wesley, Boston (2004)
4. Abrahamsson, P., Warsta, J., Siponen, M.T., Ronkainen, J.: New directions on agile methods: a comparative analysis. In: ICSE 2003: Proceedings of the 25th International Conference on Software Engineering, Washington, DC, USA, pp. 244–254. IEEE Computer Society, Los Alamitos (2003)
5. Usability.gov: Step-by-Step Usability Guide (Last visited: 18.08.2008), http://www.usability.gov/
6. W3C: Notes on user centred design process (UCD) (Last visited: 18.08.2008) (April 2004), http://www.w3.org/WAI/EO/2003/ucd
7. Holzinger, A.: Usability Engineering for Software Developers. Communications of the ACM 48(1), 71–74 (2005)
8. Holzinger, A., Errath, M., Searle, G., Thurnher, B., Slany, W.: From extreme programming and usability engineering to extreme usability in software engineering education (XP+UE→XU). In: COMPSAC 2005: Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC 2005), Washington, DC, USA, vol. 2, pp. 169–172. IEEE Computer Society, Los Alamitos (2005)
9. Ferreira, J., Noble, J., Biddle, R.: Agile development iterations and UI design. In: Agile 2007, pp. 50–58. IEEE Computer Society, Los Alamitos (2007)
10. McInerney, P., Maurer, F.: UCD in agile projects: dream team or odd couple? Interactions 12(6), 19–23 (2005)
11. Constantine, L.L., Lockwood, L.A.D.: Usage-centered software engineering: an agile approach to integrating users, user interfaces, and usability into software engineering practice. In: ICSE 2003: Proceedings of the 25th International Conference on Software Engineering, Washington, DC, USA, pp. 746–747. IEEE Computer Society, Los Alamitos (2003)

12. Memmel, T., Reiterer, H., Holzinger, A.: Agile methods and visual specification in software development: A chance to ensure universal access. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 453–462. Springer, Heidelberg (2007)
13. Göransson, B., Gulliksen, J., Boivie, I.: The usability design process - integrating user-centered systems design in the software development process. Software Process: Improvement and Practice 8(2), 111–131 (2003)
14. Toxboe, A.: Introducing user-centered design to extreme programming (May 2005), http://blog.anderstoxboe.com/uploads/16082005_XP_and_UCD.pdf
15. Hussain, Z., Lechner, M., Milchrahm, H., Shahzad, S., Slany, W., Umgeher, M., Wolkerstorfer, P.: Integrating Extreme Programming and User-Centered Design. In: PPIG 2008, The 20th Annual Psychology of Programming Interest Group Conference, 10th - 12th September 2008, Lancaster University, UK (2008)
16. Wolkerstorfer, P., Tscheligi, M., Sefelin, R., Milchrahm, H., Hussain, Z., Lechner, M., Shahzad, S.: Probing an agile usability process. In: CHI 2008: CHI 2008 extended abstracts on human factors in computing systems, pp. 2151–2158. ACM, New York (2008)
17. Benbunan-Fich, R., Benbunan, A.: Understanding user behavior with new mobile applications. Journal of Strategic Information Systtems 16(4), 393–412 (2007)
18. Holzinger, A., Searle, G., Nischelwitzer, A.K.: On some aspects of improving mobile applications for the elderly. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 923–932. Springer, Heidelberg (2007)
19. Kjeldskov, J., Stage, J.: New techniques for usability evaluation of mobile systems. International Journal of Human-Computer Studies 60(5-6), 599–620 (2004)
20. Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., Cajander, R.: Key principles for user-centred systems design. Behaviour & Information Technology, Special Section on Designing IT for Healthy Work 22(6), 397–409 (2003)
21. Hussain, Z., Lechner, M., Milchrahm, H., Shahzad, S., Slany, W., Umgeher, M., Vlk, T., Wolkerstorfer, P.: User interface design for a mobile multimedia application: An iterative approach. In: ACHI 2008, First International Conference on Advances in Computer-Human Interaction, Sainte Luce, Martinique, France, February 10-15, 2008, pp. 189–194. IEEE Computer Society, Los Alamitos (2008)
22. Sefelin, R., Tscheligi, M., Giller, V.: Paper prototyping - what is it good for?: a comparison of paper- and computer-based low-fidelity prototyping. In: CHI 2003: CHI 2003 extended abstracts on Human factors in computing systems, pp. 778–779. ACM Press, New York (2003)
23. Junior, P.T.A., Filgueiras, L.V.L.: User modeling with personas. In: CLIHC 2005: Proceedings of the 2005 Latin American conference on Human-computer interaction, pp. 277–282. ACM Press, New York (2005)
24. Subramanya, S., Yi, B.K.: User interfaces for mobile content. IEEE Computer 39(4), 85–87 (2006)
25. Kangas, E., Kinnunen, T.: Applying user-centered design to mobile application development. Commun. ACM 48(7), 55–59 (2005)
26. Hazzan, O., Tomayko, J.E.: Human aspects of software engineering: The case of extreme programming. In: XP, pp. 303–311 (2004)
27. Gittins, R., Hope, S.: A study of human solutions in extreme programming. In: Kadoda, G. (ed.) Proc. PPIG 13, pp. 41–51 (2001)
28. Law, A., Charron, R.: Effects of agile practices on social factors. In: HSSE 2005: Proceedings of the 2005 workshop on human and social factors of software engineering, vol. 30, pp. 1–5. ACM Press, New York (2005)

29. Hussain, Z., Lechner, M., Milchrahm, H., Shahzad, S., Slany, W., Umgeher, M., Vlk, T.: Optimizing Extreme Programming. In: ICCCE 2008: Proceedings of the International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, pp. 1052–1056. IEEE, Los Alamitos (2008)
30. Manifesto: Manifesto for agile software development (2001), http://agilemanifesto.org/
31. Rubin, J.: Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. John Wiley & Sons, Inc., New York (1994)
32. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus, G., Ziegler, J. (eds.) Mensch and Computer 2003: Interaktion in Bewegung, Stuttgart, B. G. Teubner, pp.187–196 (2003)
33. Jokela, T., Abrahamsson, P.: Usability assessment of an extreme programming project: Close co-operation with the customer does not equal to good usability. In: Bomarius, F., Iida, H. (eds.) PROFES 2004. LNCS, vol. 3009, pp. 393–407. Springer, Heidelberg (2004)

Adaptivity and Personalization in Ubiquitous Learning Systems*

Sabine Graf[1] and Kinshuk[2]

[1] National Central University, Graduate Institute of Learning and Instruction,
No. 300, Jhongda Rd., Jhongli City, Taoyuan 32001, Taiwan
sabine.graf@ieee.org
[2] Athabasca University, School of Computing and Information Systems,
1 University Drive, Athabasca, AB T9S 3A3, Canada
kinshuk@ieee.org

Abstract. Ubiquitous learning is an emerging research field, allowing students to learn anywhere and anytime. Adaptivity and personalization issues play an important role in ubiquitous learning systems, enabling those systems not only to allow learners to learn whenever and wherever they want, but also in a personalized and adaptive way, considering the situation, characteristics, and needs of learners. In this paper, an overview of adaptivity and personalization issues in ubiquitous learning systems is provided and current and future research directions are discussed.

Keywords: Adaptivity, personalization, ubiquitous learning systems.

1 Introduction

Ubiquitous learning environments overcome the restrictions of classroom or workplace-restricted learning and extend e-learning by bringing the concepts of anytime and anywhere to reality, aiming at providing people with better educational experience in their daily living environments. Use of devices such as mobile phones and personal digital assistants (PDA) allows new opportunities for learners by being intensely connected. Therefore, educational content can be accessed and interaction can take place whenever learners need it, in different areas of life, regardless of space and time.

While "ubiquitous learning" is a growing research area, aspects of adaptivity and personalization become more and more important. Incorporating adaptivity and personalization issues in ubiquitous learning systems (ULS) allows these systems to provide learners with an environment that is not only accessible anytime and anywhere, but also accommodating to the individual preferences and needs of learners. Being aware of and considering the current context of the learners as well as, for example, their different prior knowledge, interests, learning styles, learning goals, and

* The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 96-2520-S-008 -007 -MY2 and NSC 097-2811-S-008-001-. The authors wish to acknowledge the support of iCORE, Xerox and the research related gift funding provided to the Learning Communities Project by Mr. Allan Markin.

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 331–338, 2008.

so on, leads to a more effective, convenient, and successful learning experience in the ubiquitous learning environments.

This paper is written in the context of the workshop on “Adaptivity and Personalization in Ubiquitous Learning Systems”, aiming at giving an overview in this area and highlighting recent research. Section 2 introduces ubiquitous learning and Section 3 discusses adaptivity and personalization aspects in the context of ULSs. In Section 4, we give a brief introduction of the workshop papers and Section 5 concludes the paper with a brief discussion on future outlook.

2 Ubiquitous Learning

In this section, we give an overview about ubiquitous learning, describing its origin and definition and subsequently discussing the characteristics and features of ULSs.

2.1 Origin and Definition of Ubiquitous Learning

Ubiquitous learning has its origins in ubiquitous computing. David Lay describes ubiquitous computing as “a vision of computing power ‘invisibly’ embedded in the world around us and accessed through intelligent interfaces” [1]. The term “ubiquitous computing” has been introduced by Mark Weiser [2], who argued that the most profound technologies are those that are invisible and used by people unconsciously to accomplish everyday tasks. In a ubiquitous computing environment, many small computers are embedded in daily life objects, enabling these objects to support and assist people in tasks about work, education, and daily life. Important features of ubiquitous computing environments are the wireless communication between objects as well as the sensors, which allow the objects to sense user information and environment information in the real world and provide users with personalized services [3].

A ubiquitous learning system (ULS) supports learners through embedded and invisible computers in everyday life [4]. Such environments allow students to learn at any time and any place, encouraging them to more experiential learning [1] such as learning by doing, interacting and sharing, and facilitate on-demand learning, hands-on or minds-on learning and authentic learning [5].

Two other terms, namely mobile learning and pervasive learning, are often used in similar context with ubiquitous learning. Ogata and Yano [4] describe the difference of these terms based on Lyytinen and Yoo’s work [6] about ubiquitous computing, considering the level of embeddedness and mobility. In mobile learning, learners carry their mobile devices with them, which act as their learning environment. Thus, mobile learning has a high degree of mobility but a low degree of embeddedness, since, apart from the web server, no other computers than the students’ own mobile devices need to be involved in the learning process. Pervasive learning, on the other hand, takes place in an environment where small computers are embedded and communicate mutually in order to provide context-based information. Therefore, pervasive learning has a high degree of embeddedness but a low degree of mobility since learning depends on the location where the computers are embedded. Ubiquitous learning combines mobile and pervasive learning and assumes that computers are embedded in everyday objects. In ubiquitous learning, students move with their mobile devices which support their learning process by communicating with the embedded objects and other devices in the surrounding environment.

There exist also broader definitions of the term "ubiquitous learning". While Hwang, Tsai, and Yang [7] agree on the abovementioned definition by Ogata and Yano, they describe that a broader definition of ubiquitous learning focuses on the aspect of learning at any time and any place. According to this broad definition, an environment which allows students to learn wherever they are and whenever they want, can be considered as a ubiquitous learning environment. Furthermore, Ogata and Yano point out that for a broad definition of ubiquitous learning, both mobile learning and pervasive learning, are in the category of ubiquitous learning [4].

However, in this paper, we refer to the narrow-sense definition of ubiquitous learning given by Ogato and Yano [4], combining mobile and pervasive learning.

2.2 Characteristics and Features of Ubiquitous Learning Systems

Chen et al. [8] described five characteristics of mobile learning environments, from which Ogata and Yano [5] derived the following characteristics for ubiquitous learning: *permanency* means that students' work and the learning processes are recorded; *accessibility* means that learners can request and access their documents and data anywhere; *immediacy* allows students getting any information immediately, independent of their location; *interactivity* allows learners to interact with experts, teachers, and peers; and *situating of instructional activities* means that learning can be embedded in daily life, in a natural and authentic form. Bomsdorf [9] proposed additionally *adaptability* as the sixth characteristic of ULSs, referring to the function that learners get the right information at the right place and in the right way.

Hwang, Tsai, and Yang [7] also discussed the characteristics of ULSs, focusing more on aspects of context-awareness and adaptation. According to them, a ULS is *context-aware*, meaning that it can sense the learners' situation and the situation of the learners' current real world environment. Furthermore, a ULS is able to offer *adaptive support* based on the students' learning behavior and contexts in the real and cyber world. Moreover, the system can actively provide *personalized support* in the right way, at the right place, and at the right time, based on the personal and environmental contexts in the real world, as well as the learner's profile and learning portfolio. Another characteristic is that ULSs enable *seamless learning*, allowing students to move around in the real world. Furthermore, a ULS is able to *adapt the learning material according to the functions of the mobile device* the student is using.

A ULS can support students in many ways, offering them an environment where they can learn through experience in the real world, supported and guided by the system, which is able to adapt and personalize its interaction and suggestions to the learner. ULSs can support individual and collaborative learning. For individual learning, the system is able to *interact with students*, providing them with an active and student-centered way of learning [7]. The system can provide *navigation support* and guide students, based on their current location, to places where they can learn through experience or conduct location-dependent learning activities. Furthermore, the system is able to *present learning material* and suggest learning activities to the learners in a context-aware way, showing adaptive content or activities at the right time and the right place in order to facilitate a more authentic learning experience. For collaborative learning, ULSs can *support learners in finding peers*, with whom they can meet virtually or in real world and build learning groups, *or experts* whom they can ask for help and advice. Furthermore, when emphasizing on collaborative learning, aspects of social networks and Web 2.0 technologies can play an important role to facilitate communication, interaction, and knowledge sharing among students.

As stated, for example, by Hwang, Tsai, and Yang [7], Bomsdorf [9], and Yang [10], adaptivity and personalization are important functions of ULSs. Yang [10], for example, stated that ubiquitous learning is characterized by providing intuitive ways for identifying right collaborators, right contents, and right services in the right place at the right time based on the learners' surrounding context. Personalization and adaptivity deal with what "right" means for each learner. Therefore, the functions of personalization and adaptivity interweave with the features of ULSs mentioned above, making the interaction between the system and students, navigation support, content presentation, and searching for peers or experts personalized and adaptive.

In the next section, we discuss general issues about adaptivity and personalization as well as show how adaptivity and personalization can be applied in ULSs.

3 Adaptivity and Personalization

Adaptivity and personalization in learning systems refers to the function enabling the system to fit the students' current situation, needs, and characteristics, taking into account, for example, their knowledge level, learning styles, cognitive abilities, current location, motivation, interests, preferred language, and so on. While adaptivity focuses on the aspect that learners' situation, needs, and characteristics are considered by the system automatically, personalization is a more general term and highlights the customization of the system, including also issues which can be adapted and specified by learners themselves such as the color of the interface, the preferred language, or other issues which make the environment more personal.

Different aspects have to be considered when aiming at providing students with adaptive and personalized courses, material, or activities. One dimension incorporates *what kind of information* about the learner is used for adaptation. Another dimension considers *what can be adapted* in the system based on the specific information. In the following sections, these two dimensions are discussed in more detail.

3.1 Which Information Is Used for Adaptivity and Personalization?

Many research works about adaptivity and personalization have been carried out in the context of hypermedia and web-based learning systems, focusing for example on the consideration of students' level of knowledge, learning styles, cognitive abilities, motivation, and so on. In ULSs, this research is very relevant with respect to content presentation. Furthermore, information about the students' and the environments' context is a crucial factor for providing adaptivity and personalization in ULSs.

The necessary information for providing adaptivity and personalization in ULSs can be retrieved from different sources. Hwang, Tsai, and Yang [7] pointed out 5 types of situation parameters, which represent different kinds of information. Through the sensors of the ULS, information about the students' context and the environmental context can be gathered. Information concerning the students' context includes, for example, their current location, the time of their arrival and issues such as heartbeat or blood pressure. The environmental context includes information about the environment of a sensor, such as location, temperature, and information about approaching objects/people. Furthermore, information can be gathered from the students' interaction with the system via the mobile device, including for example stored documents, given answers to questions, and certain settings the learner made in his/her user interface. Moreover, the system can access a database, where personal data of students and environmental data are stored. Personal data can include the student's learning styles, course schedule, prior knowledge, progress in the course and so on. Environmental data provide more detailed information about the environment, such as a schedule of arranged learning activities or notes for using the site.

3.2 What Can Be Adapted and Personalized?

Another dimension refers to *what* can be adapted and personalized in a system. As mentioned above, a ULS can support learners by interacting with them, guiding them to suitable places, presenting them with learning material and activities, and helping them to interact with peers and experts. In the following paragraphs, examples for adaptivity and/or personalization for each of these features are discussed.

With respect to interaction between the system and the learner, ULSs can be actively involved in the students' learning activities [7]. Through their context-awareness, ULSs can provide personalized hints at the right time or suggest suitable learning activities for a learner, depending on the current environment and the needs of the student. An example for adaptive and personalized interaction between the system and the learner is shown in the language learning system JAPELAS [11]. JAPELAS aims at teaching foreign students Japanese polite expressions. When a learner starts talking to another person, the system gives a suggestion about the level of polite expression based on hyponymy, social distance, and situation through receiving information from the other person's mobile device and from sensors of the own device. Another example is the language learning system TANGO [12], which detects objects around the learner, using RFID tags, and involves these objects in learning activities, for example, asking the student to close a window or move a can from one place to the other.

In order to help learners to navigate to locations where learning can take place, personalization and adaptivity aspects deal mostly with location-awareness and planning suitable learning activities. For example, a ULS can generate a navigation path according to students' prior knowledge in order to remedy students' lack of knowledge [13]. Hwang, Tsai, and Yang [7] describe a similar scenario where a ULS asks a student to go to a specific place to observe and identify a plant.

Regarding content presentation, different methods exist for providing students with adapted and personalized learning material. These methods determine how learning material is presented differently for learners with different characteristics and needs. Brusilovsky [14] classified these methods regarding to their aim into two groups, namely for adaptive presentation and adaptive navigation support, whereby navigation in this context means how students navigate through the learning material. Many research works have been conducted, especially in the domain of adaptive educational hypermedia and web-based systems, focusing on generating courses and material that fit the students' profile and portfolio. Furthermore, for ULSs, adaptation of content presentation for a particular mobile device is another important issue [7]. Examples for applying adaptivity and personalization aspects for content presentation in ULSs are proposed by Bomsdorf [9] and Graf et al. [15]

Another feature of ULSs includes interaction among learners as well as between learners and teachers/experts. Adaptivity and personalization can be used for improving asynchronous and synchronous communication in a ULS. Regarding asynchronous communication, components such as discussion forums, question and answer services, and knowledge sharing services can be made more personal and adaptive to the learners' situation, characteristics, and needs. For example, considering where a student asked a specific question or generated a document can provide additional data, which can be useful when peers search for information. Furthermore, ULSs can help students to communicate synchronously by assisting them in forming learning groups or showing them who might be able to answer their questions. For example, Martin et al. [16] presented a location-based application which gives information about people who are close to the learner. Furthermore, a ULS can provide suggestions for building learning groups based on the students' location as well as other students' characteristics, as proposed, for example, in [15].

4 Topics of the Workshop

In this section, we outline the papers presented at the International Workshop on Adaptivity and Personalization in Ubiquitous Learning Systems, aiming at introducing current research in the respective area. Each of these papers contributes towards the development and the effectiveness of adaptive and personalized ubiquitous learning environments, focusing on different aspects.

The paper by Thalmann is a review paper on the adaptation criteria used in adaptive hypermedia systems, showing what kind of information is applied for providing adaptivity and personalization. Furthermore, suggestions are provided for preparing learning material with respect to the identified adaptation criteria.

The paper by Glavinić and Granić deals with adaptivity and adaptability issues for user interfaces of learning systems, concerning adaptation of knowledge presentation and interaction style as well as focusing on adaptation to mobile devices.

The next two papers discuss development issues. The paper by Hussain, Lechner, Milchrahm, Shahzad, Slany, Umgeher, and Wolkerstorfer deals with the development process of applications for mobile devices. In this paper, the integration of extreme programming and user centered design in the development process is proposed and its benefits are demonstrated. Mödritscher and Wild proposed an approach for personalized e-learning based on web application mashups, focusing on how the learning process can be adapted by allowing learners integrating various web-based tools into one aggregated view and supporting them in building and sustaining learning communities.

The paper by Zakrzewska deals with improving the formation process of learning groups. Zakrzewska presented an approach for clustering students according to their learning styles and usability preferences. This technique can contribute to the formation of learning groups in ULSs by suggesting learners other peers who are in the same cluster and therefore can learn together from the same material.

The next two papers deal with facilitating and supporting learners with little experience with technology. Ziefle investigated which instruction format is beneficial for navigation performance in mobile phones as well as whether one instructional format can be used in general or the instruction format should be tailored to specific information needs and user abilities. Morandell, Hochgatterer, Fagel, and Wassertheurer presented a study about the usage of personalized photo-realistic avatars for elderly people and people with mild cognitive impairment, aiming at improving the acceptance and user-friendliness of ambient/ubiquitous assisted living environments.

5 Conclusions

This paper aimed at providing an overview about adaptivity and personalization issues in ULSs, by presenting an introduction about ubiquitous learning and discussing the possibilities of adaptivity and personalization in ULSs. Subsequently, we introduced current research in this area, presented at the International Workshop on Adaptivity and Personalization in Ubiquitous Learning Systems.

Ubiquitous learning is an emerging and promising research field, which can benefit strongly by considering personalization and adaptivity aspects, making the support provided to the learners more convenient and effective. As can be seen from the proposed research works, many areas such as mobile learning, ambient assisted living, human-computer interaction, and adaptive hypermedia need to contribute in the development and effective usage of adaptive and personalized ULSs.

As a conclusion from the past and current research directions, one issue seems to be of special importance for future research. ULSs can detect and store a huge amount of information about learners and their environments. However, currently only few kinds of information are considered in a single ULS when providing adaptive and/or personalized support. A great potential lies in considering and combining the huge amount of information, which can be detected by ULSs, and using it for providing richer and more effective adaptivity and personalization for learners. This includes especially the combination of context information from ULSs, such as location or data about the environment, with information in the learners' profile, such as his/her interests, motivation, cognitive abilities, and learning styles.

References

1. Lay, D.: Ubiquitous computing. In: Emerging technologies for learning, vol. 2, pp. 64–79. British Educational Communications and Technology Agency (2007)
2. Weiser, M.: The Computer for the 21st Century, September 2001, pp. 94–104. Scientific American (2001)
3. Hwang, G.-J.: Criteria and Strategies of Ubiquitous Learning. In: Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC 2006), pp. 72–77. IEEE Computer Society, Los Alamitos (2006)
4. Ogata, H., Yano, Y.: Context-Aware Support for Computer-Supported Ubiquitous Learning. In: Proceedings of the International Workshop on Wireless and Mobile Technologies in Education, pp. 27–34 (2004)

5. Ogata, H., Yano, Y.: How Ubiquitous Computing can Support Language Learning. In: Proceedings of the International Conference on Knowledge Economy and Development of Science and Technology, pp. 1–6 (2003)
6. Lyytinen, K., Yoo, Y.: Issues and challenges in ubiquitous computing. Communications of ACM 45(12), 63–65 (2002)
7. Hwang, G.-J., Tsai, C.-C., Yang, S.J.H.: Criteria, Strategies and Research Issues of Context-Aware Ubiquitous Learning. Educational Technology & Society 11(2), 81–91 (2008)
8. Chen, Y.-S., Kao, T.-C., Sheu, J.-P., Chiang, C.-Y.: A Mobile Scaffolding-Aid-Based Bird-Watching Learning System. In: Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE 2002), pp. 15–22. IEEE Computer Society Press, Los Alamitos (2002)
9. Bomsdorf, B.: Adaptation of Learning Spaces: Supporting Ubiquitous Learning in Higher Distance Education. In: Mobile Computing and Ambient Intelligence: The Challenge of Multimedia, Dagstuhl Seminar Proceedings 05181 (2005)
10. Yang, S.J.H.: Context Aware Ubiquitous Learning Environments for Peer-to-Peer Collaborative Learning. Educational Technology & Society 9(1), 188–201 (2006)
11. Yin, C., Ogata, H., Yano, Y.: JAPELAS: Supporting Japanese Polite Expressions Learning Using PDA(s) Towards Ubiquitous Learning. The Journal of Information and Systems in Education 3(1), 33–39 (2004)
12. Ogata, H., Akamatsu, R., Mitsuhara, H., Yano, Y., Matsuura, K., Kanenishi, K., Miyoshi, Y., Morikawa, T.: TANGO: Supporting Vocabulary Learning with RFID tags. In: Electronic proceedings of International Workshop Series on RFID, Tokyo (2004)
13. Chang, A., Chang, M.: Creating an Adaptive Mobile Navigation Learning Path for Elementary School Students Remedy Education. In: Proceedings of the International Conference on Interactive Computer Aided Learning, Villach, Austria (2006)
14. Brusilovsky, P.: Adaptive hypermedia. User Modeling and User-Adapted Interaction 11, 87–110 (2001)
15. Graf, S., MacCallum, K., Liu, T.-C., Chang, M., Wen, D., Tan, Q., Dron, J., Lin, F., McGreal, R.: Kinshuk: An Infrastructure for Developing Pervasive Learning Environments. In: Proceedings of the IEEE International Workshop on Pervasive Learning (PerEL 2008), pp. 389–394. IEEE Press, Hong Kong (2008)
16. Martín, S., Sancristobal, E., Gil, R., Castro, M., Peire, J.: Mobility through Location-based Services at University. International Journal of Interactive Mobile Technologies (iJIM) 2(3), 34–40 (2008)

Instruction Formats and Navigation Aids in Mobile Devices

Martina Ziefle

HumTec, RWTH Aachen University, Theaterplatz 14
D-52062 Aachen, Germany
Ziefle@humtec.rwth-aachen.de

Abstract. Three different instruction formats were examined respecting their usefulness for the navigation through hierarchical menus in mobile phones. 56 middle-aged adults had to solve four mobile phone tasks twice consecutively. Before completing the first trial, they were supported by instruction formats which contained different spatial knowledge types [28]. The first form was a procedural step-by-step instruction delivering landmark and route knowledge. The second instruction consisted of a visualized menu tree, in which the menu path to be taken was marked. By this, mainly survey knowledge but also route information was provided. The third instruction format also used the visualized menu tree. However, in addition to the menu path which had to be taken, also functions' labels were given, conveying landmark, route and survey knowledge. Further, a condition was examined in which no help was given. Dependent variables were navigation effectiveness and efficiency. Overall, the step-by-step condition, the instruction type most often used in technical leaflets, had the smallest effect, especially for users with low spatial visualization abilities. With instruction forms which use diagrammatic visualizations and provide survey knowledge, navigation performance considerably improved.

Keywords: Instruction format, navigation aid, navigation effectiveness, efficiency, landmark knowledge, route knowledge, survey knowledge.

1 Introduction

Information and communication technologies have interpenetrated all professional and private fields in the last decades. A prominent role in this context play mobile small screen technologies, e.g. mobile and smart phones, communicators, navigation devices and electronic organizers, which show continuously increasing rates of growth each year [26]. Beyond their ubiquity, these technologies have fundamentally changed the nature of social, economic and communicative pathways in modern societies. Communication and information are present everywhere and at any time and seem to overcome physical as well as mental borders. Though, the effective and successful integration of mobile technologies and broad acceptance of these technologies in all fields of daily life impose considerable challenges on modern societies.

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 339–358, 2008.

1.1 The Problem

If critically looking at current developing trends, a number of crucial factors come into fore that may severely impede the barrier free integration of future communication technologies into private and professional societal structures.

(1) *User diversity*: Contrary to former times, when the utilization of information and communication technologies (ICT) was restricted to technology-experienced expert groups, nowadays broader user groups have access to ICT. Thus, the organization of professional and private activities, events and transactions heavily depends on the ease of using these devices by a diverse user population. However, recent studies showed that the usage of small mobile technologies imposes considerable difficulties and that these difficulties are not equally strong for all users. Menu navigation performance was found to be strongly influenced by a number of individual factors, which differentially benefit or hamper navigation [e.g. 21, 12]. Among the individual factors, age, gender, spatial orientation and computer-experience were identified to play a prominent role. Users with high spatial visualization abilities were distinctly advantaged compared to users with low spatial abilities in completing computer tasks [e.g.31], using hypertext [e.g.19] or mobile devices [e.g.2,4,5,8,37,38]. Also, prior computer experience is a crucial factor for navigation efficiency [e.g.1,12,32]. Moreover, gender is increasingly discussed to play an important role in the explanation of technical performance. Women usually report lower levels of computer-related self-efficacy [e.g.4] and a higher computer anxiety [e.g.10], which, in turn, may reduce the probability of active computer interaction and lead to a generally lower computer-expertise level. Likewise, noticeable differences in computer experience between males and females were reported [e.g.3,23], which then mediate gender differences in technical performance. As the usage of (mobile) technologies is increasingly less voluntary, more and more technically inexperienced and/or aged users are urged to use these devices. Usability issues of interface and communication designs are thus of prominent interest.

(2) *Function complexity and device miniaturization*: Mobile devices are typically small-sized with a small communication display but a huge functionality, providing on-the-go lookup and entry of information, quick communication and instant messaging anytime and everywhere. The mobile character of these devices in combination with the small communication window represents a still higher usability demand compared to large display technologies. The limited screen space is extremely problematic for providing optimized information access and the question of how to “best” present the information on the small display. Only few items can be seen at a time and users navigate through a menu whose complexity, extension and spatial structure is not transparent to them as it is most of the time hidden from sight. As a consequence, users are urged to memorize the functions’ names, their relative location within the menu and keep up orientation. Disorientation in handheld devices’ menus is a frequent problem [e.g. 21,30,31,35,36,37,39], especially for aged users [e.g. 2,21,22,35,36,37] or those with little computer-related knowledge and experience [e.g. 3,4,5,29].

(3) *New application fields*: Aggravating the situation, the application fields and the contexts, in which mobile technologies are used, changed considerably over the last

decade, and, still, they will tremendously change. The traditional functionality of cell phones, effectuating simple calls, plays merely an inferior role. Rather, the devices are used as controllers for technical processes at home (e.g., programming TV), as mobile alternatives to traditional computers, as basic communication devices (e.g., Internet access), and as intelligent robots that administrate personal concerns (e.g., managing accounts). Mobile technologies are assumed to especially support seniors in their daily lives and keep up independent living. As such, the devices are used as e.g. for medical monitoring, as navigation and memory aids. But mobile technologies are also increasingly used in smart homes and ambient intelligent environments, in which devices are communicating with remote computers, sometimes integrated in clothing [e.g.18], furniture or walls [e.g.23]. Taken together, given these developments, usability demands are more important than they ever were. As long as information designs of technical devices are not easy to use and learn technical innovations can not have sustained success.

1.2 The Duties for a Human-Centered Design

Design approaches must therefore take the user-perspective seriously. This claim includes that also instruction leaflets and technical manuals are ergonomically designed and can be widely understood by users. This is the focus of the present experimental study.

Though, the creation of usable and understandable manuals is a challenging issue. On the one hand, manuals must meet demands of user diversity, thus even users with a different upbringing and education level as well as users with a limited domain-knowledge should be supported adequately. Moreover, the manuals should give users a proper understanding of the device structure and to prevent them from getting lost in the menu, which is especially important for users with low spatial visualization abilities. On the other hand, adults do not want to spend much time reading manuals [e.g.11], presumably because the study of text heavy instructions may be conceived as time-consuming and bothersome. Furthermore, many users report that instructions and manuals usually do not tell them what they need to know, but provide them with a huge amount of "unnecessary" and "useless" information. Accordingly, research confirmed the ineffectiveness of detailed textual instructions, and proposed diagrammatic instructions instead [e.g. 25]. However, diagrammatic instructions also have fallacies: the most severe failure is the lack of unambiguously picturing the initial step of a subtask including the location and object of a required action (e.g.11,14,15,33]. Transferring this concept to the cognitive ergonomic design of instructions for menu-driven small screen devices, as the mobile phone, it is decisive to tell the user where within the menu a certain task has to be accomplished [e.g.6,8,37], thus providing spatial clues. Thus, it may be assumed that processes of spatial orientation take place when users interact with hypertext and navigate through the different nodes and links [e.g.12,17,20] as well as with hierarchical menu structures [e.g.7,9,30].

According to spatial visualization theories [28], three forms of spatial knowledge are proposed to be crucial for proper spatial orientation in natural environments: Landmarks (outstanding points in the environment), routes (the ways connecting different landmarks) and survey knowledge (a graphical and structural outline of a specific terrain). If when navigating through the hierarchical menu structure of a mobile phone comparable processes to the ones involved in spatial orientation in the natural environment take place, then instructions should support users with the different kinds of spatial knowledge. To build up survey knowledge a map of the menu structure could be provided. For large displays the efficiency of such a map has already been demonstrated [e.g.9] and also for small screen devices such as the mobile phone [6,8]. A very recent study [8] examined the usefulness of instruction types for the navigation efficiency in mobile phones, using children as participants. Children are generally assumed to be especially technology prone and experienced, and to easily master the usage of technological devices. Though, in contrast, it was found that children are very sensitive to the cognitive demands imposed by the devices and showed considerable performance losses in suboptimal interface designs [e.g.8,32,34,35,36,37,38]. In order to learn if children benefit from spatial maps, an instruction type delivering survey knowledge was contrasted to a conventional step-by-step instruction, containing route knowledge. Both instructions were compared to a free exploration of the menu which should provide all kinds of spatial information. The findings revealed that the benefit of instruction formats depended on the children's age and the ability to process spatial information (which is carried by age). The younger the children (8 - 9 years), the lower were the benefit by spatial maps *and* by the free exploration. With increasing age, children were able to fully exploit the spatial clues of the map and also to integrate the spatial knowledge they gained from exploring the menu. For the younger children, a sequential aid in form of a step-by-step instruction was more helpful for them.

1.3 Questions Addressed

Summing up, it is of considerable impact that instruction formats and manuals of technical devices are well designed and that they compensate the specific knowledge gaps users experience when navigating through menus of small screen devices. As disorientation is a rather frequent phenomenon in small screen devices, resulting in a lot of detouring within the menu, especially spatial orientation clues should be provided preventing users from loosing their bearings within the menu.

So far, comparably little is known regarding the question how appropriate diagrammatic instructions have to be designed for the "average user", i.e. middle-aged adults, which cannot rely on a mature technical understanding, and which do not have much experience with the handling of small screen devices. As exactly this user group will be the typical end users of modern mobile technologies in novel application contexts, it was of interest to find out if this user group is able to benefit of spatial information about the menu structure delivered by diagrammatic instructions. On the one hand, the survey knowledge should help them to built up a mental representation of the system's structure [e.g.17,27,28,34,37], which in turn enables users to increase navigation efficiency. On the other hand, diagrammatic instructions could also impose considerable cognitive load, as participants have to understand abstract information and to transfer the spatial clues into a proper mental representation of the menu. The present study therefore examines the usefulness of different instruction formats on navigation efficiency. Another point at issue was the sustainability of the different instructions, assessing learnability effects. Three different forms of instructions with varying spatial knowledge clues were examined and compared to each other. In order to learn if any instruction type is superior for navigation efficiency than no help, a control condition was also examined, in which no navigation aid was given. Understanding in how far individual user characteristic may interact with the benefit by different instruction formats, spatial ability was psychometrically assessed and related to performance outcomes.

2 Method

2.1 Experimental Variables

As *independent variable* the type of instruction format was examined.

1. The first type of instruction was a step-by-step instruction containing the different menu functions that had to be selected one after another in order to solve a task. The knowledge provided by this instruction is landmark knowledge as well as route knowledge (names of the functions to be visited one after another), without revealing the menu structure. Fig. 1 shows one example of such an instruction. Participants were instructed that the functions shown in the step-by-step diagram are the key terms they had to memorize, as they are the cardinal points in the menu they should follow. This instruction was presented for a period of 20 s prior to each task in the first trial.

Fig. 1. Step-by-step instruction type

2. The second form of instruction was a menu structure visualized in a graphical tree where the name and location of the functions that have to be selected to solve a task are given. This instruction type provides primarily survey knowledge, but contains also information on landmarks (the functions' names) and route knowledge (interconnections between functions). Figure 2 (left side) shows an extract of a menu tree containing the path and the menu functions. In the experiment, this information was plotted on a sheet of paper, separately for each task, and pinned on the wall. Participants were instructed that this structural map informs them about all the different paths and branches of the phones' menu as well as their interconnections. They were also told that the functions shown in the map were the key terms they had to memorize, as they are the cardinal points in the menu they should follow.
3. The third instruction type was again the graphical tree map (as in the second instruction type), however without any function labels. Thus, an "empty tree" was presented. By this, primarily structural information (survey knowledge and route knowledge) were delivered. Participants were instructed that this structural map informs them about all the different paths and branches of the phones' menu as well as their interconnections (Figure 2 right side).

As dependent variables four different performance measures were analyzed. As effectiveness measure, (1) the number of tasks successfully solved was determined. Furthermore, measuring efficiency of menu navigation, (2) the time needed to process the tasks, (3) detour steps (number of keystrokes carried out that were not necessary

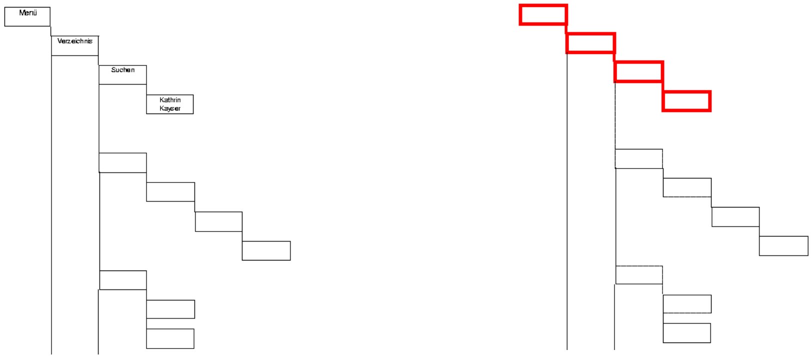


Fig. 2. Instruction types: Tree with functions (left side) and empty tree (right side)

when solving the task the shortest way possible) and finally, (4) hierarchical returns to higher levels within the menu indicating that participants felt to have taken the wrong path in the menu hierarchy and needed to go back to a known position.

2.2 Apparatus and Procedure

A mobile phone corresponding to the Nokia 3210 model regarding menu and keys was simulated as software solution, run on a PC and displayed on a touch screen (TFT-LCD Iiyama TXA3841, TN, 15' with touch logic by ELO RS232C). It was a general question which phone model was appropriate for experimental purposes. The software prototype mirrored exactly the menu structure and the navigation keys given in the real phone. However, the keys and the display of the mobile phone were enlarged compared to the original model in order to exclude motor and visual difficulties. Avoiding biases and concealing the real brand the simulated phone had a neutral design. Three menu functions were presented on the screen at a time. In order to measure effectiveness and efficiency in detail, user actions were recorded online on the keystroke level. Thus, frequency and type of keys used, time spent on tasks and the navigational route could be reconstructed in detail. In order to assure a comfortable and relaxed body posture while completing the tasks, participants sat on a table and worked on the touch screen, which was fixed to the edge of the table.

Each participant was allocated to one of the four experimental conditions. In the three instruction conditions, the specific type of instruction participants would thereafter use was presented and explained referring to a sample instruction. It was carefully ensured that all participants understood exactly what they had to do in each task. Instructions were shown for a period of 20 seconds after being explained what they had to do in the respective tasks and before starting to process each single task.

2.3 Participants

56 participants from a wide professional range volunteered in the experiment, in an age range between 32 and 65 years of age (M=49.6 years, SD=8.7). Half of them (N = 28) were females (M=46.8 years; SD=9.4), the other half (N = 28) were males

(M=47.0 years; SD=8.3). As we aimed at a "normal" group, both academic and non-academic education levels were present. All of them were quite frequent phone and computer users, however, it was taking care of that participants did not reveal to be technical experts or professionals for small-screen-devices. It was a specific interest in this study if the "average" user would benefit from diagrammatic instructions (compared to no instructional aid) and if specific spatial knowledge demands can be met by the instructions. In all four conditions, gender and age was balanced. In a pre-experimental questionnaire, the previous experience with technology (mobile phone, PC, DVD) was assessed. Statistical testing revealed no differences within technical experience across groups. Also the four groups had a comparable spatial ability.

2.4 Assessing Spatial Visualization Ability

After participants had completed the phone tasks twice consecutively, they were requested to complete a spatial test of mental rotation [3DW,16]. The test contained seven mental rotation tasks and was presented In Figure 3, a test item is shown. Participants had to decide which of the cubes (A-F) has the same sides and the same spatial orientation than the target cube (x). To do so, they had to mentally rotate the different alternatives (A-F) and to compare it with the target cube. A maximum of seven points could be reached.

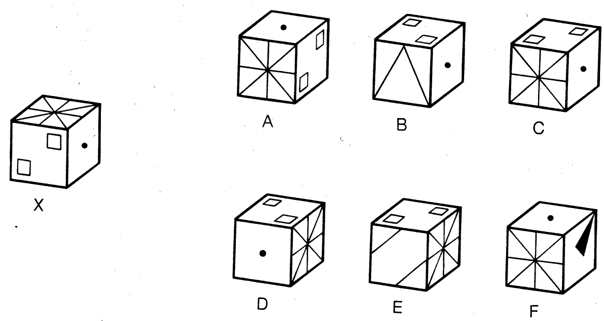


Fig. 3. Exemplary test item from the mental rotation ability test [16]

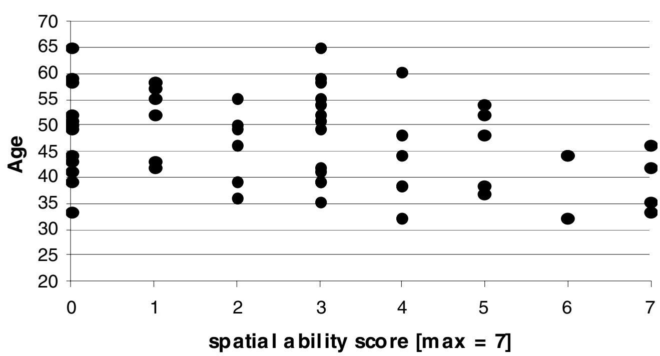


Fig. 4. Spatial ability scores as a function of age of participants

Participants did not have a time limit to complete this test (no speed test). Though, the perceived ease of processing these items considerably varied among participants as did the time they needed to complete the task.

Spatial ability scores are given in Figure 4. As can be seen there, only few participants succeeded to solve all tasks correctly. Conversely, there were a considerable number of participants who did not even solve one of the tasks correctly. Also, there was a significant correlation between spatial ability scores and the age of participants ($r = 0.456$; $p<0.01$) showing that with increasing age spatial ability significantly decreases. Though male participants had, on average a higher spatial score (M= 2.9 out of seven; SD=2) than female participants (M=2.5; SD=2.3), differences did not reach statistical significance.

2.5 Experimental Tasks

Four basic and very common mobile phone tasks (calling a number stored in the phone book, sending a text message, hiding one's own number in the display, making a call divert) had to be solved by the participants. In total, 36 steps were necessary to solve the four tasks on the shortest way possible. In order to measure learnability, all four tasks were presented twice. In the second run, participants were not given any instructional help. This was of specific significance as it was a central interest to analyze the sustainability of the respective instruction type. After a short break, participants completed the tasks a second time, in the same order than in the first trial. For each task, a time limit of 5 minutes per task was set.

3 Results

Data were analyzed by MANOVA procedures and by ANOVA analyses with repeated measurements, assessing learnability effects. The significance level was set at 5%. Spatial ability and gender were treated as between-subject variables in order to learn if the usefulness of instruction formats interacts with these individual factors. Regarding spatial ability, by median split two groups across the participants groups were formed. By this, participants with "high" spatial ability score were contrasted to participants with "low" ability score.

First, the main effect of instruction format are reported, followed by effects of spatial abilities and gender. After that, learnability effects on navigation efficiency are addressed as well as interacting effects between the extent of learnability and the type of instruction format. This will answer the question if there are namable differences between the different instruction formats and, still more important, about the sustainability of the instruction formats over time.

3.1 Effects of Instruction Formats on Navigation Performance

First, effects if the different instruction formats are reported. Descriptive outcomes can be found in Figure 5. There, the effectiveness (upper left diagram) the mean effectiveness, the time participants needed to process the task (upper right diagram), the number of detour steps (lower left diagram) and the number of returns to higher levels in the menu (lower right diagram) is pictured.

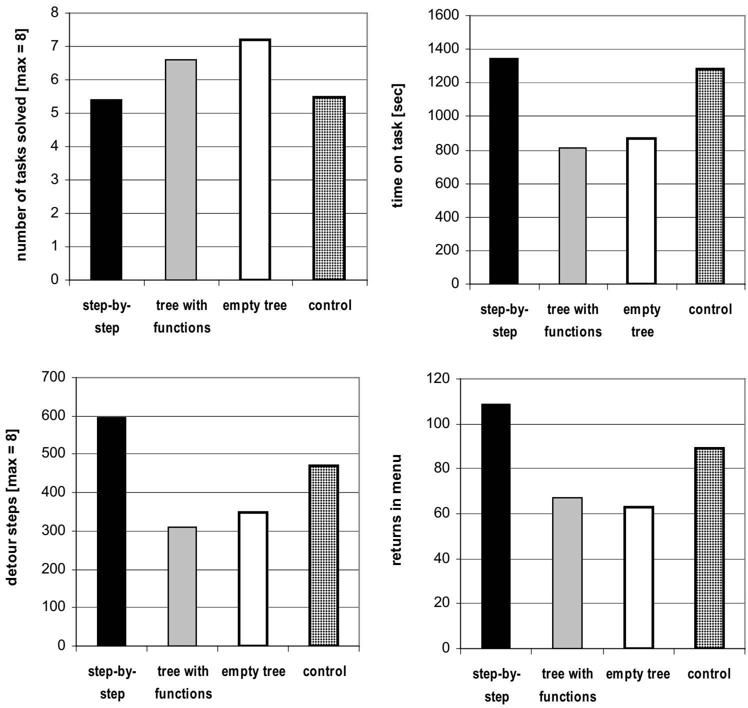


Fig. 5. Navigation effectiveness and efficiency in the four experimental conditions

Statistical testing revealed a significant main effect of instruction format on all dependent measures: task effectiveness ($F\,(3,41)=3.2$; $p<0.05$), time on task ($F\,(3,41)=2.7$; $p=0.05$), number of detour steps ($F\,(3,41)=5.8$; $p<0.05$) as well as on the number of returns to higher levels in the menu ($F\,(3,41)=11.2$; $p<0.05$). Among the instruction formats, the step-by-step condition was the instruction format which performed worst. Taking both trials together, participants solved, on average, about 6 tasks out of eight successfully and needed about 19 minutes to process the tasks. In addition, 477 detour steps and 90 returns to higher levels in menu hierarchy were carried out, hinting at a considerable difficulty to find the targeted function. Both map conditions performed significantly better. Supported by the empty tree, the most abstract instruction format, participants solved on average, 7.2 tasks (out of eight), 14 minutes processing times, only 320 detour steps and executed 69 returns in the menu. The tree which was providing landmark knowledge in addition (function names), also yielded a benefiting effect: It took, on average, 14.9 minutes to solve 6.5 of the eight tasks successfully. Participants executed 371 detour steps and about 67 hierarchical returns to find the targeted function. The question if any instruction format is better than having no help has to be answered in the negative. There was no significant performance difference between the control condition and the step-by-step condition.

Two things should be noted in this context: One is that the step-by-step-condition is the most frequently used instruction format in conventional leaflets of technical devices, thus corroborating the information designs of menu-driven small screen devices to be suboptimally designed. The second point refers to the overall navigation performance. It should be kept in mind that the phone used here was a mass model with a medium complexity and that overall not more than 72 steps had to be carried out in order to solve the four tasks twice on the most direct path. The fact that average users carried out between 320 and 480 detour steps and, for reorientation, went back about 70 times to higher levels in menu hierarchy, shows the enormous cognitive difficulty of users to orientate within the menu of small screen devices.

3.2 Effects of Spatial Ability on Navigation Performance

The spatial ability revealed to be a prominent factor for navigation performance, confirming outcomes of previous studies. Participants with high spatial abilities (scores above the median) showed a significant higher navigation performance than participants with a lower spatial ability score.. The main effect of spatial ability was found across all dependent variables (task effectiveness: $F(3,41)=24.7$; $p<0.05$; time on task: $F(3,41)=26.2$; $p<0.05$; number of detour steps: $F(3,41)=5.8$; $p<0.05$ and number of returns in menu hierarchy: $F(3,41)=11.2$; $p<0.05$). Descriptive outcomes for all dependent variables are visualized in Figure 6.

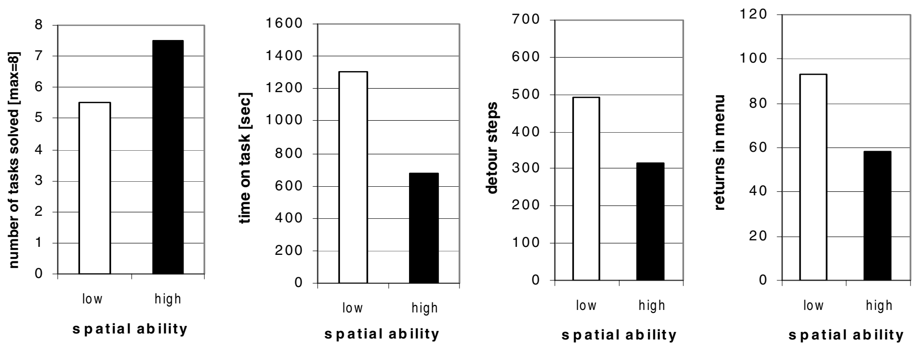


Fig. 6. Effects of spatial ability on navigation performance. White bars indicate performance of users with spatial ability scores below, black bars above the median.

3.3 Effects of Gender on Navigation Performance

Another analysis was run to learn if gender differences can be identified regarding navigation performance. In the literature, there are a number of studies according to which female users show a lower technical performance compared to men, but there are also studies, which could not identify performance differences. Performance differences between males and females are assumed to be mediated by the lower technical experience of female users, a lower technical self-competency and lower spatial abilities. In our study, female users did not differ within the extent of technical experiences and spatial abilities. Therefore, one could expect that gender differences in navigation performance would not be there.

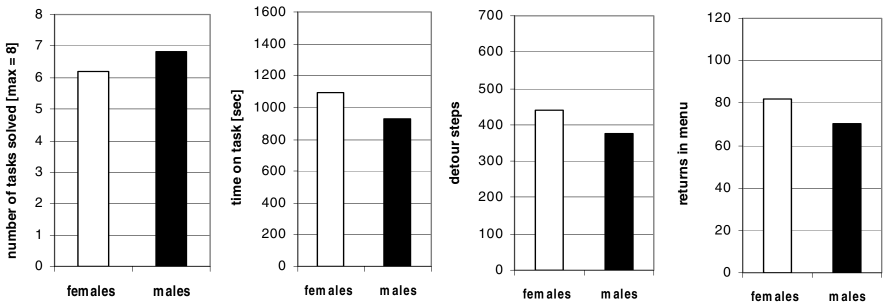


Fig. 7. Gender effects on navigation performance

Accordingly, statistical testing did not reveal a significant omnibus effect, but a marginally significant effects were found for task effectiveness F (3,41)=3.9; p<0.1) and time on task F (3,41)=3.1; p<0.1). In Figure 7, descriptive outcomes for all dependent measures are visualized.

Taken together, the gender effect was not clear (at least compared to the clear effects of instruction format and spatial ability), but though cannot be fully ruled out.

3.4 Learnability Effects on Navigation Efficiency

A special focus of the present study was laid on learnability effects, the question if the navigation performance improves from the first to the second trial (Figure 8). In all dependent variables a profound and significant learning effect was revealed (Task effectiveness: F (1,41)=3.3; p<0.1; time on task: F (1,41)=55,9; p<0.05; detour steps: F (1,41)=24.8; p<0.05; returns in menu hierarchy: F (1,41)=9.5; p<0.05).

Interestingly, the nature of learnability effects seemed to mainly concentrate on an increase within the efficiency with which participants navigated through the menu. They navigated considerably faster in the second run and also, with lesser detouring within the menu. In contrast, participants could not increase their task effectiveness in the same scale.

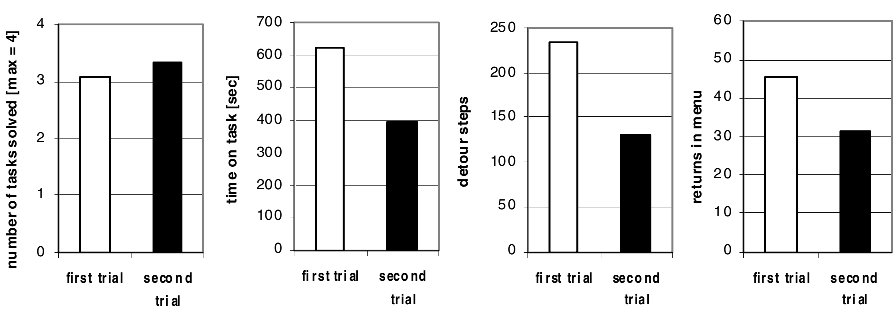


Fig. 8. Learnability effects on navigation effectiveness and efficiency from the first (white bars) to the second run (black bars)

3.5 Interaction of Learnability and Instruction Formats on Navigation Efficiency

Do instruction formats differ within learnability effects? To answer this question navigation performance in the first trial compared to the second trial was analyzed for each of the instruction formats, separately. Figure 9 illustrates the outcomes.

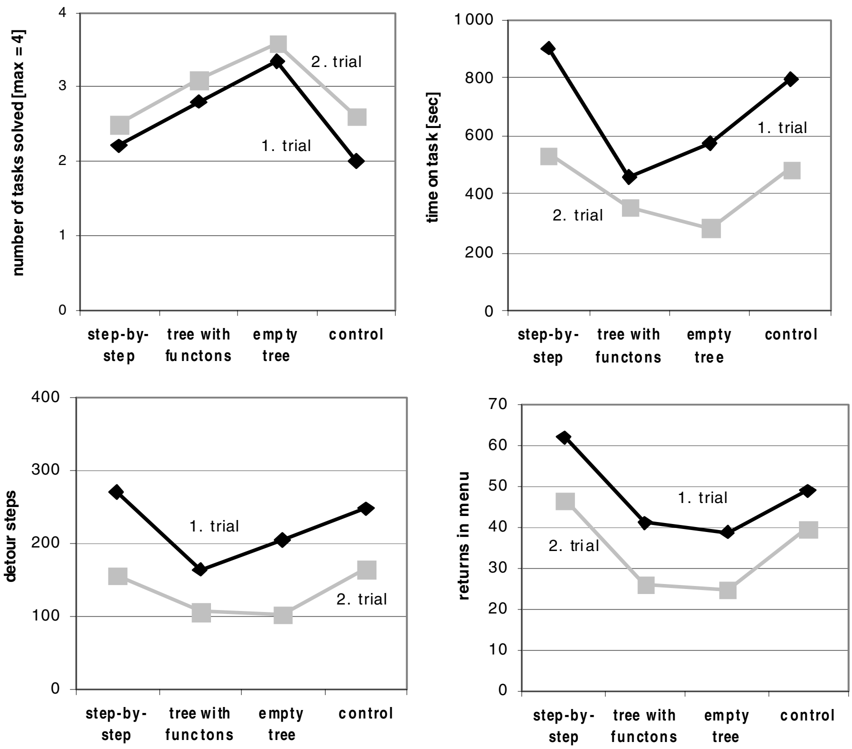


Fig. 9. Learnability effects on navigation effectiveness and efficiency in the first compared to the second trial depending of different instruction formats

Beyond the distinct main effects for instruction format (tree conditions are superior) and learnability (performance increases from the first to the second run), generally, no significant interacting effect between instruction formats and learnability were identified. Though, for the time on task ($F(1,41)=3.8$; $p<0.05$) and the number of detour steps ($F(1,41)=2.8$; $p=0.05$) interacting effects of learnability and instruction format reached statistical significance. A closer insight revealed that the learnability effect from the first to the second run was stronger in the step-by-step condition compared to other instruction formats. However, it should be noted that performance in the step-by step condition still was considerably low (significant main effect) compared to the performance in the tree conditions.

3.6 Interacting Effects between Users' Spatial Ability, Instruction Formats and Learnability Effects

It is a crucial question if we can assume that individual factors, as e.g. users' spatial abilities mediate effects of different instruction formats. If so, then we should carefully select specific formats depending on individual user profiles. As spatial abilities have revealed to be a prominent mediator of performance in menu navigation of small screen devices, it was a crucial question if there are interacting effects between users' ability to process spatial information, the type of instruction format and learnability effects. ANOVA analyses with repeated measurements showed marginally significant effects for this three-way interaction ($p<.1$). In order to understanding the nature of the interaction, in the following the outcomes are visualized in detail for each of the dependent measures, directly contrasting participants with high and low spatial abilities. First, task effectiveness is looked at (Figure 10).

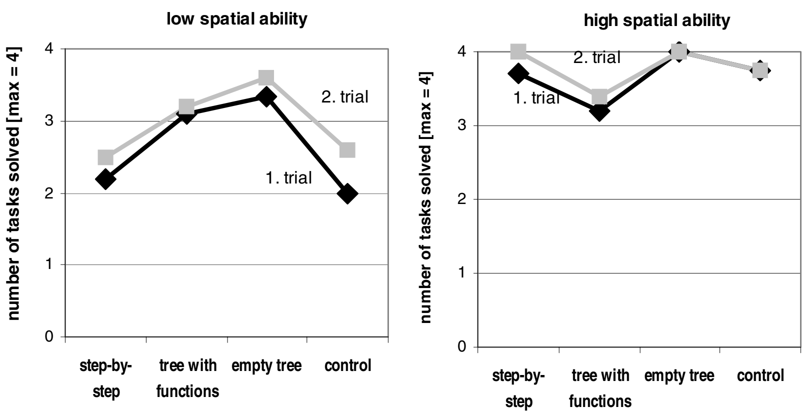


Fig. 10. Task effectiveness in the first and the second run, depending on the type of instruction format and spatial abilities. Left: Low spatial ability group, right side: high spatial ability group.

As can be seen from figure 10 (right side), persons with high spatial abilities show high success rates independently from the type of instruction. Also, learnability effects are quite small, as the performance was already nearly perfect in the first run. The picture changes when looking on the group with lower spatial abilities (left side). Their task effectiveness strongly depends on the type of instruction format. Participants which were supported by diagrammatic structural information containing survey knowledge sow a higher task success compared to the step-by-step-instruction and the control group.

Regarding the time needed to complete the task (Figure 11), also a marginally significant three-way interaction was found (F (1,41)=2.4; p<0.1).

Beginning again with the high-spatial-ability group (Figure 11, right side), we do not see any interacting effect between learnability and instruction format. Independently of the type of instruction, the time needed to process the tasks was higher in the first compared to the second run, showing a solid learnability effect.

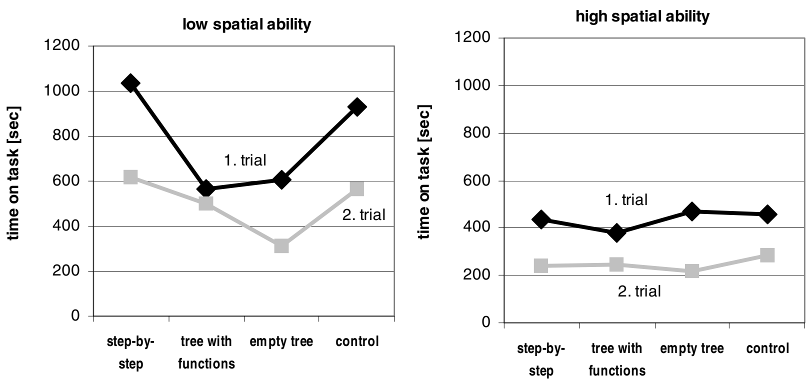


Fig. 11. Time on task in the first and the second run, depending on the type of instruction format and spatial abilities. Left: Low spatial ability group, right: High spatial ability group.

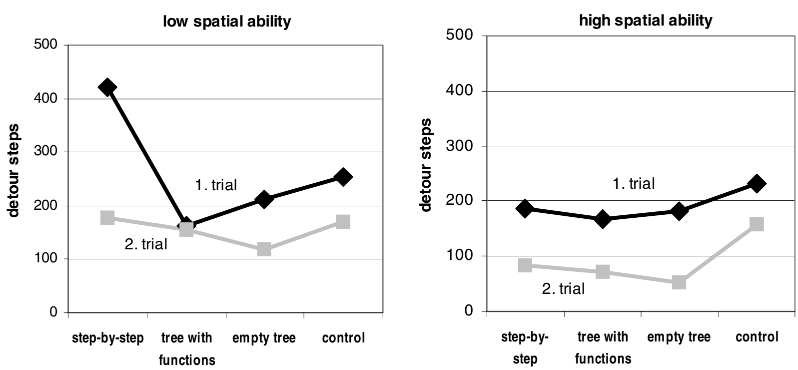


Fig. 12. Number of detour steps carried out in the first and the second run, as a function of the type of instruction format and spatial abilities. Left: Low spatial ability group, right side: High spatial ability group.

However, participants with lower abilities reacted quite differentially depending on the type of instruction (Figure 11, left side). The step-by step-condition and the control condition yielded the lowest performance. However, whenever participants with low spatial abilities are supported by an instruction format containing survey knowledge, navigation performance increases. This shows that the map conditions do compensate the specific knowledge gaps of participants having only low spatial abilities.

For the number of detour steps (Figure 12), a similar pattern was found revealing a marginally significant three-way-interacting effect (F (1,41)=2.4; p<0.1).

Participants with high spatial abilities (Figure 12, right side) showed a similar detouring behavior, more or less independently from the type of instruction format, but carried out significantly fewer detour steps compared to the first run. Again, it is the low spatial group (Figure 12, left side), which showed differential effects of instruction formats. Performance is low for this group, when they received the step-by-step format or do not get any help. But performance is considerably advantaged whenever instruction formats deliver survey knowledge.

Finally, the number of returns in menu hierarchy is looked at (Figure 13). For this measure, no three-way interacting effect was found. However, for the sake of completeness, the detouring behavior in both groups and experimental conditions are illustrated.

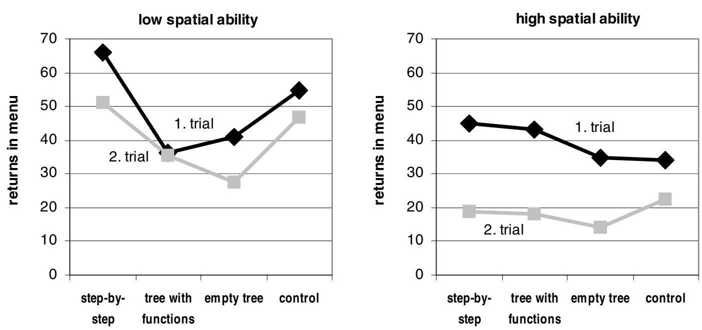


Fig. 13. Number of hierarchical returns in the first and the second run, depending on the type of instruction format and spatial abilities. Left: Low spatial ability group, right side: High spatial ability group.

4 Discussion

The present research was guided by two major goals. One goal was concerned with the question which instruction format would be beneficial for navigation performance in a rather typical user group, which does not possess a sophisticated technical understanding and experience, but rather represents the typical “average” user. The other goal was to find out if we can rely on a “design-for all”- approach, or if we need to tailor instruction formats according to specific information needs and user abilities. As disorientation is a rather frequent problem when using mobile small-screen devices with a hierarchical menu [2,4,5,8,19,36,37], the usefulness of instruction formats were under study, which contained different types of spatial knowledge. Based on the theoretical framework of spatial orientation [28], we varied three types of spatial knowledge within the instruction formats. One type represented an sequential aid (step-by-step-instruction), which tells users which step they have to do one after another. This type provides mainly landmark knowledge and resembles the conventional navigation aid in leaflets of technical devices. The second type contained more spatial information. A diagrammatic map of the menu structure was presented to participants, with the function names as additional clues. This type delivered survey knowledge (knowledge about the structure of the menu), landmark knowledge (function names) as well as route knowledge (interconnections within the menu). A third instruction type again used a visualization of the menu tree, however, in this type the function names were not given, but only the path, which had to be taken by participants, was marked. Accordingly, this type contained survey and route knowledge.

The results showed that a procedural step-by-step-instruction had by far the smallest effect. Participants in this condition showed the lowest task effectiveness and needed more time to complete the tasks. Also, the detouring behavior within the menu was much larger than in those conditions, in which the map of the menu was given in a diagrammatic form. Still more critical, the step-by-step condition showed even larger impedimental effects than in the control condition in which users did not receive any instructional aid. In contrast, instruction formats containing structural knowledge, as it was present in both map conditions, enabled the participants to come along much better with the tasks. Task success was much higher and the detouring in the menu was distinctly reduced, taken from the smaller number of detour steps and returns to higher levels in menu hierarchy in these conditions. It is an interesting finding that the map with functions, which contained the richest spatial information (survey, route and landmark knowledge), did not yield a better performance compared to the map, in which only the path to be taken was marked (delivering survey and route knowledge). This shows on the one hand that the function names are less decisive for a proper menu orientation if users receive information about the menu structure. On the other hand it may be concluded that survey knowledge in combination with route knowledge is essential for users in order to properly orientate within the menu.

Based on the strong evidence for the considerable impact of users' characteristics on navigation performance in mobile devices [e.g.13,17,20,21,31,36,37], the impact of users' abilities was related to performance measures. It was of high ergonomic interest whether an interface can be created that enables weaker users to use mobile phones competently, and, further, to reduce disorientation.

Especially, spatial abilities had been identified to play a crucial role for navigation performance [e.g.12,13,17,20]. But what makes spatial abilities so important for menu navigation? It is assumed that spatial abilities specifically advantage navigation performance by supporting users in constructing a proper mental representation of the systems' structure [e.g. 27]. The mental representation on its part helps users to harmonize their relative position in the menu and, at the same time, to keep the menu structure in mind while navigating through the system. Thus, by having an appropriate model and a structural concept of the mental 'room' that has to be navigated through, performance is increased. In the mobile phone, where the overall structure of the menu is not transparent, and the screen size is very limited, spatial abilities may be even more crucial, because users have to develop a mental representation of the structure when navigating through the functions.

It is an important finding of the present study that the specific benefit of survey knowledge was distinctly larger in persons which have lower spatial abilities. Supported by the map, task success was higher, the time on task was reduced and considerably fewer detour steps and returns to higher levels in menu hierarchy were carried out. Thus, we can assume that the map compensates their specific knowledge gap and helps users with low spatial abilities to develop a proper mental representation of the menu structure. By this disorientation in the menu is considerably reduced.

Regarding the question which of the instruction formats shows the largest learnability effect, it was found that participants in all conditions did improve navigation performance from the first to the second trial, showing a solid learnability effect. Even if this finding is promising, it should be kept in mind that, still, the step-by-step condition, the instruction type frequently used in conventional leaflets of technical devices, led to the smallest performance, especially in those users which need to be especially supported.

A final note is concerned with the overall performance of the participants observed here. One might have expected that the mobile phone tasks would have been processed without too much friction losses. Even if this might be true for the task effectiveness in the participant group with high spatial abilities (as most of the participants solved slightly more than seven out of eight tasks), it should be taken into account that in our study participants were given 5 min for each task, plenty of time compared to the time periods users take in real life. Looking at efficiency measures however, it was found that typical end users of mobile small screen devices carried out a lot of detouring routes instead of solving the tasks the shortest way possible. Given the reference of 36 steps that were necessary to solve the four tasks on the most direct path, this seems to be quite substantial. Conceivably, this shows that mobile phones - at least in the way being currently designed – impose a high cognitive load on users.

From the data of this study the usage of diagrammatic instruction formats can be insistently recommended for a wide range of users. It is to be noted that the information about the menu structure and the knowledge about the hierarchical nature and as well as information about interconnections between menu categories are the most important features that should be transported to users.

However, the findings presented here are limited to hierarchically structured menu types of mobile phones. Therefore, future studies should continue in this line of research and examine appropriate instruction formats also for non-hierarchical small-screen device menus. In contrast to the exclusively hierarchically structured data in cell phones, PDAs and smart phones provide both, network structures as well as hierarchical menu parts and it can be assumed that in hypertext structures, which do have a much higher function complexity in addition, disorientation is even more likely to occur.

References

1. Argarwal, R., Prasad, J.: Are differences Germane to the acceptance of new Information technologies? Decision Sciences 30, 361–391 (1999)
2. Arning, K., Ziefle, M.: Cognitive and personal predictors for PDA navigation performance. Behaviour and Information Technology (in press)
3. Arning, K., Ziefle, M.: Assessing computer experience in older adults: Development and validation of a computer expertise questionnaire for older adults. Behaviour and Information Technology 27(1), 89–93 (2008)
4. Arning, K., Ziefle, M.: Understanding differences in PDA acceptance and performance. Computers in Human Behaviour 23(6), 2904–2927 (2007)
5. Arning, K., Ziefle, M.: Barriers of information access in small screen device applications: The relevance of user characteristics for a transgenerational design. In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 117–136. Springer, Heidelberg (2007)
6. Bay, S.: Cellular phone manuals: Users' benefit from spatial maps. In: Cockton, G., Korhonen, P. (eds.) Proc. SIGCHI 2003, pp. 662–663. ACM Press, New York (2003)
7. Bay, S., Ziefle, M.: Design for all: User characteristics to be considered for the design of phones with hierarchical menu structures. In: Luczak, H., Zink, K.J. (eds.) Human Factors in Organizational Design and Management – VII, pp. 503–508. IEA Press, Santa Monica (2003)
8. Bay, S., Ziefle, M.: Landmarks or surveys? The impact of different instructions on children's performance in hierarchical menu structures. Computers in Human Behaviour 24(3), 1246–1274 (2008)
9. Billingsley, P.A.: Navigation through hierarchical menu structures: Does it help to have a map? In: Proceedings of the Human Factors and Ergonomics Society 26th Annual Meeting, 1982, pp. 103–107 (1982)
10. Busch, T.: Gender differences in self-efficacy and attitudes toward computers. Journal of Educational Computing Research 12, 147–158 (1995)
11. Caroll, J.M., Mack, R.L.: Exploring a Word Processor. Human Computer Interaction 1, 283–307 (1985)
12. Downing, R.E., Moore, J.L., Brown, S.W.: The effects and interaction of spatial visualization and domain expertise on information seeking. Computers in Human Behaviour 21, 195–209 (2005)
13. Egan, D.: Individual Differences in Human-Computer-Interaction. In: Helander, M. (ed.) Handbook of Human-Computer-Interaction, pp. 543–568. Elsevier, Amsterdam (1988)
14. Frese, M., Alrecht, K., Altmann, A., Lang, J., Van Papstein, P., Peyerl, R., Pruemper, J., Schulte-Goecking, H., Wankmueller, I., Wendel: The effects of an active development of the mental model in the training process: experimental results in a word processing system. Behaviour and Information Technology 7(3), 295–304 (1988)
15. Frese, M., Brodbeck, E., Heinbokel, T., Mooser, C.: Errors in training computer skills: On the positive function of errors. Human Computer Interaction 6(1), 77–93 (1991)
16. Gittler, G.: Three-dimensional Cube Test, Götingen, Germany, Beltz (1999)
17. Kim, H., Hirtle, S.C.: Spatial metaphors and disorientation in hypertext browsing. Behaviour and Information Technology 14(4), 239–250 (1995)
18. Leonhardt, S.: Personal health care devices. In: Mukherjee, S., Aarts, E., Roovers, R., Widdershoven, F. (eds.) AmIware: Hardware Technology Drivers of Ambient Intelligence, pp. 349–370. Springer, Dordrecht (2006)

19. Lin, D.-Y.: Age differences in the performance of hypertext perusal. In: Proceedings of the Human Factors and Ergonomic Society 45th Annual Meeting, pp. 211–215. Human Factors and Ergonomics Society, Santa Monica (2001)
20. Norman, K.L.: Spatial visualization - a gateway to computer-based technology. Journal of Special Education Technology 12(3), 195–206 (1994)
21. Pak, R.: A further examination of the influence of spatial abilities on computer task performance in younger and older adults. In: Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, pp. 1551–1555. Human Factors Society, Santa Monica (2001)
22. Park, D.C., Schwarz, N.: Cognitive Aging. Buchanan, Philadelphia (1999)
23. Prante, T., Streitz, N., Tandler, P.: Roomware: Computewrs disappear and Intreraction evolves. Computer 37(12), 47–54 (2004)
24. Rodger, J.A., Pendakhar, P.C.: A field study of the impact of gender and user's technical experience on the performance of voice activated medical tracking application. International Journal of Human-Computer Studies 60, 529–544 (2004)
25. Rodriguez, M.A., Polson, P.G.: Creating usable wordless instructions for performing complex one-time tasks: Effects of violating the rules. In: Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, New Orleans, USA, pp. 941–945 (2004)
26. Shiffler, G., Smulders, C., Correira, J.M., Hale, K., Hahn, W.M.: Gartner dataquest market databook (2005) (accessed September 2006), http://www.gartner.com
27. Sein, M.K., Olfman, L., Bostrom, R.P., Davies, S.A.: Visualisation ability as a prediction of user learning success. International Journal of Man-Machine-Studies 39, 599–620 (1993)
28. Thorndyke, P.W., Goldin, S.: Spatial learning and reasoning skill In Spatial orientation. In: Pick, H.L., Acredolo, L.P. (eds.) Theory, research, and application, pp. 195–217. Plenum, New York (1983)
29. Tuomainen, K., Haapanen, S.: Needs of the active elderly for mobile phones. In: Stephanidis, C. (ed.) Universal access in HCI: Inclusive design in the information society, pp. 494–498. Mahwah, NJ (2003)
30. Vicente, K.J., Hayes, B.C., Williges, R.C.: Assaying and isolating individual differences in searching a hierarchical files system. Human Factors 29(3), 349–359 (1987)
31. Westerman, S.J.: Individual differences in the use of command line and menu computer interfaces. International Journal of Human Computer Interaction 9(2), 183–198 (1997)
32. Ziefle, M.: The influence of user expertise and phone complexity on performance, ease of use and learnability of different mobile phones. Behaviour and Information Technology 21(5), 303–311 (2002)
33. Ziefle, M., Bodendieck, A., Kuenzer, A.: The impact of user characteristics on the utility of adaptive help systems. In: Khalid, H.M., Helander, M.G., Yeo, A.W. (eds.) Work with Computing Systems, pp. 71–76. Damai Sciences, Kuala Lumpur (2004)
34. Ziefle, M., Bay, S.: Mental models of Cellular Phones Menu. Comparing older and younger users. In: Brewster, S., Dunlop, M. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 25–37. Springer, Heidelberg (2004)
35. Ziefle, M., Bay, S.: How older adults meet cognitive complexity: Aging effects on the usability of different cellular phones. Behaviour & Information Technology 24(5), 375–389 (2005)
36. Ziefle, M., Schroeder, U., Strenk, J., Michel, T.: How young and older users master the use of hyperlinks in small screen devices. In: Proc. SIGCHI 2007, pp. 307–316. ACM Press, New York (2007)

37. Ziefle, M., Bay, S.: How to overcome disorientation in mobile phone menus: A comparison of two different types of navigation aids. Human Computer Interaction 21(4), 393–432 (2006)
38. Ziefle, M., Bay, S., Schwade, A.: On keys' meanings and modes: The impact of navigation key solutions on children's efficiency using a mobile phone. Behaviour and Information Technology 25(5), 413–431 (2006)
39. Ziefle, M., Arning, K., Bay, S.: Cross-platform consistency and cognitive compatibility. In: Richter, K., Nichols, J., Gajos, K., Seffah, A. (eds.) The Many Faces of Consistency in Cross-Platform Design. CEUR-WS.org, pp. 75–81 (2006), http://CEUR-WS.org/Vol-198/
40. Ziefle, M., Bay, S.: Transgenerational Designs in Mobile Technology. In: Lumsden, J. (ed.) Handbook of Research on User Interface Design and Evaluation for Mobile Technology, pp. 122–140. IGI Global (2008)

HCI Research for E-Learning: Adaptability and Adaptivity to Support Better User Interaction

Vlado Glavinić[1] and Andrina Granić[2]

[1] University of Zagreb, HR-10000 Zagreb, Croatia
Faculty of Electrical Engineering and Computing (FER)
Department of Electronics, Microelectronics, Computer and Intelligent Systems (ZEMRIS)
vlado.glavinic@fer.hr
[2] University of Split, HR-21000 Split, Croatia
Faculty of Science
Department of Computer Science
andrina.granic@pmfst.hr

Abstract. The paper provides an overview of our research being performed in the area of interfaces for e-learning, with the emphasis on application of intelligent methods. We especially address the efforts being done in providing a suitable interaction for a class of intelligent e-learning systems – Intelligent Tutoring Systems (ITSs), within the research framework of our home institutions. The specific issues being thus studied include adaptivity topics related to knowledge representation, interaction style as well as mobility support, along with related usability assessment methods. The dependence on intelligent handling of the former stems from the expected more efficient and effective solution of problems related to the diversity of users of Information and Communication Technology (ICT) supported education. In this respect we also shortly tackle and comment on a number of adaptable and adaptive interfaces for e-learning, which are in this context being conceptualized, implemented, and assessed both *in vitro* and *in vivo*.

Keywords: Adaptivity, e-Learning Systems, Intelligent Tutoring, m-Learning Systems, Usability Evaluation, User Interfaces, User Sensitive Design.

1 Introduction

E-learning as a concept, and especially the associated e-learning systems, has recently taken over a number of important tasks in order to help cope with a new situation in education. Specifically, democratization of knowledge acquisition is now targeting an increasingly diversified student population, here including enrolled, out-of-course, life-long, and students with special needs. The number of students/learners is constantly increasing, but this growth is being supported by a quasi-constant and anyway very constrained number of teachers/instructors. Within such a framework, along with providing students with a way to acquire knowledge, the need to objectivize and speed-up their evaluation is but a natural follow-up. The above trend of having to deal with an ever growing and dissimilar learners' collection is being closely followed by the separation, both spatial and temporal, of learners and teachers. All of this creates a

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 359–376, 2008.

situation requiring a massive application of e-learning systems that would exhibit (functional) autonomy and independence, along with a better "friendliness" towards their users, adapting to their educational needs, hence altogether showing some kind of intelligence.

The landscape as described in the above paragraph provides a natural motivation for dealing with user interface adaptation in e-learning systems, therefore complementing already present intelligent tools and procedures of the application area. Consequently, this paper presents our efforts in the area of adaptive and adaptable interfaces for e-learning systems, and especially for a class of intelligent e-learning systems, which is exemplified by intelligent tutoring systems and authoring shells. In this area our work has been focused on specific topics of knowledge presentation, interaction style, and mobility. Within this framework we have studied some of the interaction issues in the respective user interfaces of a number of intelligent e-learning systems, which have been operational in the educational process at two of the Croatian universities for some time already, hence providing us with objectivity and practical insight.

Specifically, in order to study the possibilities of adaptation of knowledge presentation we developed an adaptable interface featuring multiple views (*MUVIES*), which made it particularly appropriate for use as an educational system for novice users. With the *AKBB* prototype implementation, offering different interface styles in a self-adaptable way, we investigated style swapping basing on parameters pertaining to user characteristics, and which were inferred and quantified during the interaction. Regarding mobility, we are focusing on providing an intelligent mobile interface, which is planned to complement a prototype mobile intelligent tutoring system and, in lines with the main concepts of user sensitive design, relies on the use of personal agents and related user and device descriptions based on profiles.

We have also considered user sensitive research methods, by performing experimental studies, which aim to study in more detail the existence and level of interaction among users' individual differences and learning outcomes accomplished while using e-learning applications not restricted to be intelligent ones only.

The paper is organized as follows. In Section 2 we provide a short outline of e-learning systems, emphasizing intelligent ones and particularly the ITS class, as our work has been in part concentrated on the locally developed TEx-Sys family of ITSs [54]. In Section 3 we delve on the possibilities of adaptation of interfaces for e-learning systems, concentrating on the topics of knowledge and interaction style. Adaptation issues related to mobility are tackled in Section 4. In Section 5 we discuss the methods for usability assessment used and (possibly) refined in the previously mentioned research. Finally, in Section 6 we provide discussion on our results along with some afterthoughts and (the usual) plans for further work. Section 7 concludes the paper.

2 E-Learning Systems

E-learning is a general term used to denote computer-enhanced learning, as an instructional content or learning experience delivered or enabled by electronic technology [46]. E-learning systems can usually be subdivided into two groups. The first one comprises those systems that do not imply the use of methods and techniques pertaining to artificial intelligence, providing the students a "linear" information flow according to a predefined educational program.

The other group is formed by *intelligent* e-learning systems, as exemplified by *intelligent tutoring systems* (ITSs), which adapt to the needs and requirements of diverse learners according to their capabilities, previous knowledge and (partial) evaluation results [8] (it should be noted that "automatic" generation of ITSs is possible by using e-learning systems known as *authoring shells*, ASs [44]). Aside the functional adaptation, these highly interactive systems should accordingly be equipped with equally functional adaptive interfaces, in order to enhance the overall performance of the teaching/learning process. This is particularly important in the present times, as the aligning of the complete European higher education system to the postulates of the Bologna declaration makes it a must to support an increasingly autonomous and independent learning style, with students being assigned more and more personalized work to be done at home or anywhere else but in an "ex-cathedra" class. Adjustments to be enforced in user interfaces of (intelligent) e-learning systems should inherently include manipulations turning up from the implementation of universal access principles, as the distant user (learner/student) would necessarily use a variety of access devices of very differing functionality and processing power, as well as a range of underlying communication channels of different capacity and quality.

3 Adaptation in E-Learning System Interfaces

HCI research acknowledges that understanding users' needs is at the core of successful designs for information society technology (IST) products and services. In the emerging knowledge society for all, those system user interfaces are unquestionably important, which are more closely tailored to the way people naturally work, live and acquire knowledge. The role of an intuitive interface and a flexible interaction suited to different needs, preferences and interests becomes even more important for the users' success, as users with a wide variety of background, skills, interests, expertise, goals and learning styles are using computers for quite diverse purposes [3]. This leads to *user-centered design* [45] and *user sensitive design* [25], as the natural and most appropriate methodology developed out of it. The central concept of user sensitive design is an equal focus on user requirements and the diversity of such requirements in the population of intended users. Moreover, the aim is to be able to focus on extra-ordinary users as well as typical users [48].

Intelligent user interfaces (IUIs) emerge as a means for making systems individualized or personalized, thus enhancing their flexibility and attractiveness. They are intended to facilitate a more natural interaction between users and computers, by not attempting to imitate human-human communication, but augmenting the human-computer interaction process instead [4]. The intelligence in the interface can e.g. make the system adapt to the needs of different users, take initiative and make them suggestions, learn new concepts and techniques or provide explanation of its actions, cf. [33; 41]. Measuring the degree of adaptation to differing user requirements and needs is obviously emphasized in such a context. Consequently, the frequently cited indication of intelligence is the ability to adapt [43], which implies the ability to adapt the interaction outcome to the level of understanding and the interests of individual users. A sound framework for dealing with users' heterogeneity was already provided by [52], which envisaged:

- *adaptable* systems, which allow the user to control interface customization, and
- *adaptive* systems, which adapt the interface appearance and behavior to users' individual characteristics.

Adaptive interfaces generally rely upon the use of *user models* (UMs), collections of information and assumptions about single users which are subsequently used when adapting the system to an individual [38; 43; 7]. While some of the information in the user model may be relatively static and long-term, other information may be updated dynamically as the user interacts with the system. This information is used in various ways to provide adaptivity, i.e. to enable the system to adjust its functionality and/or the communication according to the needs of individual users, needs that may also change over time [9].

System intelligent/adaptive behavior strongly relies upon *user individual differences* [10; 11; 37; 7], and influences the degree of success or failure experienced by users as well. When considering adaptation of systems to individual use, user personality and cognitive factors have specifically to be taken into account because of their higher resistance to change. Additional "stable" knowledge about the user is conveyed through her/his long-term characteristics, including information on the level of expertise with computers in general, with the system in particular, as well as familiarity with the system's underlying task domain. Certain information related to users' preferences or current goals conveyed through short-term user characteristics should also be considered.

3.1 Adaptation of Knowledge Presentation

Adaptation of knowledge representation makes the first of the elements of system adjustment to expected user requirements, hence providing her/him the possibility of selecting the suitable perspective on the object of manipulation. This can eminently be granted by applying *multiple views* [49], a technique for concurrently presenting the user with different and semantically meaningful perspectives on a common internal representation of the same data, the individual views simultaneously reflecting possible changes performed upon it. In the context of intelligent tutoring systems multiple views show the following desirable features [16]:

- increased users' productivity, because of both easier processing/manipulating the domain knowledge base and easier accessing the domain knowledge, as well as being tutored by it, respectively;
- multiple concurrent views, resulting in a more complete insight into the domain knowledge base structure and an easier and faster executing of the job required;
- distinct views on the domain knowledge items, their relationships (i.e. semantics) and their creation/modification;
- structured templates for domain knowledge bases, since syntax-directed editing can be applied, resulting in an easier and faster creation of the domain knowledge base additionally experiencing in a smaller error rate;
- more intelligible and clear display of changes performed, as simultaneous updates of knowledge representation is performed through all the views.

In order to study the possibilities of adaptation of knowledge representation we therefore developed *MUVIES*, a *Mu*ltiple *Vie*ws user *customizable* (i.e. *adaptable*)

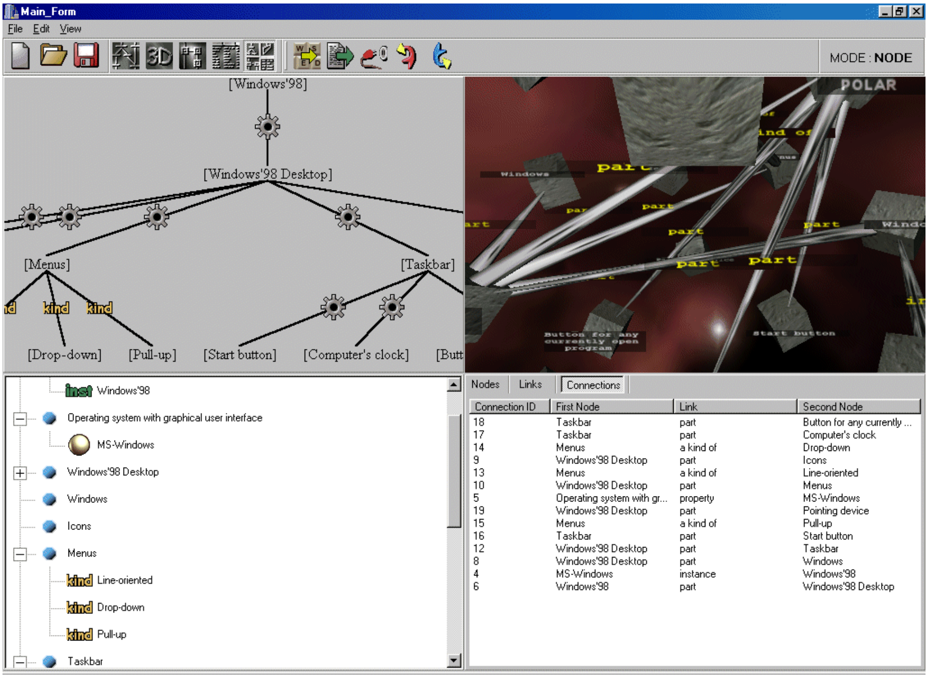


Fig. 1. Combined view in *MUVIES* (clockwise, starting from top left: graphical view of the semantic network structure, full three-dimensional view, list view and tree view)

interface for *TEx-Sys*, an intelligent hypermedia authoring shell [54] that used semantic networks as the means to represent knowledge. *MUVIES* displayed (i) a fully graphical view of the semantic networks' structure, (ii) a fully three-dimensional view of the domain knowledge, (iii) a Windows Explorer-like view, (iv) a list view as well as (v) a combined view. Fig. 1 shows the combined view, which incorporates all the other single views [16].

Other good features of *MUVIES* included (i) hiding the internal structure of the domain knowledge base, (ii) faster creation of the knowledge base supported by better interaction, and (iii) a richer and semantically more meaningful set of widgets provided by this new interface. On the whole, because of the above capabilities *MUVIES* was undoubtedly more acceptable to novice users, making it particularly appropriate for use as an educational system. In this context the game-like interaction offered by the full three-dimensional view certainly represents an early instance of application of *dynamic media* in e-learning [29].

3.2 Adaptation of Interaction Style

Aiming to further our investigation of adaptation effects, as a first step towards full interface adaptation, we introduced adaptivity in a characteristic part of ITS (more exactly, of AS) operation, that is *domain knowledge base generation*. The prototype implementation was named *Adaptive Knowledge Base Builder*, *AKBB*, and provided

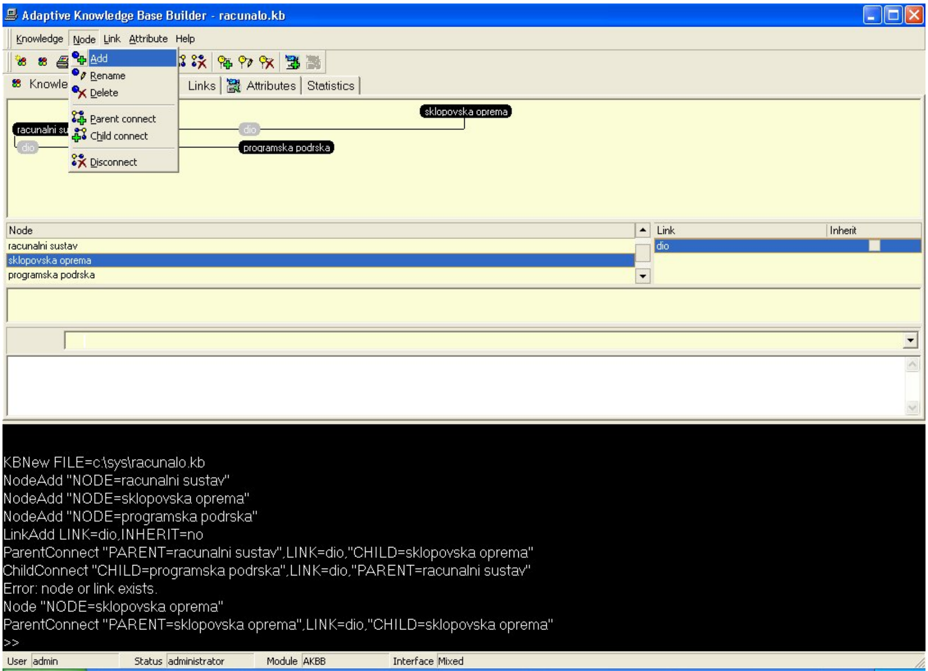


Fig. 2. Mixed mode interface style in AKBB

means for the development of specialized ITSs for particular domains of education [18]. AKBB offered three different interface styles, namely (i) a command, (ii) a graphical, and (iii) a mixed one, basing on knowledge about the individual user and the interaction session, which was dynamically acquired in run-time. Fig. 2 illustrates AKBB's mixed mode interface style.

AKBB is an example of a *self-adaptation* [9], where the system itself observes the communication, decides whether to adapt or not and generates and executes the adaptation as well. Specific values of the parameters that control style swapping rely on user characteristics. Those which are inferred and quantified during the interaction include the following ones:

- user level of experience in computer usage in general and in usage of AKBB itself; these characteristics are taken into account because of their influence on successful task accomplishment, what is based on general results of user analysis;
- cognitive and individual characteristic of the user, i.e. spatial ability, which has relevance to the usage of AKBB different interface styles.

As asserted by the "architecture" (or reference model) for adaptive user interfaces [5], AKBB used three models for its operation [24]:

- a user model, based on monitoring the user in run-time,
- a system model, storing system characteristics that are adaptive,
- an interaction model, defining the actual interface adaptation through parameter values obtained from the interaction, along with all the relevant inferences and

adaptations; five inferring as well as twelve adaptivity rules were included in the prototype.

Results of AKBB interface evaluation provided directions for its redesign, which included the following issues:

- the presentation of domain knowledge failed to convey in a transparent way the semantics of linked domain knowledge objects; in order to hide as much as possible the internal structure of the domain knowledge base, knowledge presentation should be redesigned, a possible solution being outlined in [17];
- the user model developed to support AKBB interface adaptation sadly relies on only a few user individual differences; in order to consider research and innovations in user sensitive research, the employed user model should be enhanced with personal characteristics that affect both learning and achieved learning outcomes;
- the interface should be refined in order to provide the users more control both by disabling automatic adaptation and by incorporating manual selection for swapping the operation mode;
- further research is needed to determine whether the AKBB adaptive interface is measurably better than a non-adaptive one and under what circumstances the benefit is more valuable than the apparent loss of control due to unexpected adaptations of the interface.

Following the rising interest in Web-based access to e-learning systems, and in the context of universal accessibility, we applied the preceding ideas (and experience) upon this topic, which resulted in a *distributed intelligent tutoring system with adaptive interface* named *AW-BITS* (Adaptive Web-Based ITS) [23]. AW-BITS is meant to provide an interface enabling system adaptation to user personal characteristics by both adaptive *presentation* (adapting the content of a page to a user according to the user model) and adaptive *navigation support* (in orientation and navigation by changing the appearance of visible links towards domain knowledge nodes).

3.3 User Sensitive Research Methods in E-Learning

As it is already noted before, the employment of user sensitive research methods is required in order to provide stronger foundations for designing effective e-learning systems. Experimental studies are thus performed, which aim to study in more detail the existence and level of interaction among users' individual differences and learning outcomes accomplished while using an e-learning application, e.g. [22; 20]. Identified personal user features, which are assumed to affect the learning process, are classified as follows:

- *personal* user characteristics, quite stable over time and independent from the system, where general personal characteristics (including characteristics that reflect internal psychological state) and previously acquired knowledge along with user abilities can be differentiated, and
- *system-dependent* user characteristics, which are the most changeable category of characteristics as related to particular system.

The experiment in [22], which is carried out through a five-step procedure, reveals some interesting results. It appears that there are no associations between intelligence and personality factors with learning outcomes; the only motivation to learn in addition to expectations of learning has statistically significant impact on knowledge acquired through interaction with the system. On the other hand, an expected highly significant correlation is found between students' background knowledge and their prior experience in using both computers and the Internet.

We consider this study to have provided us important directions for establishing an enhanced methodology which relies on research and innovations in user sensitive design. Further to the necessity to enlarge number and diversity of participants, we have found certain procedural issues which need to be refined as well. A set of guidelines to perform sound and instructive experimental studies to verify our hypothesis is presented in [19]. Due to still little evidence and missing empirical support, we are not further addressing the above in this paper.

4 Adaptation to Mobile Requirements

Mobile learning (*m-learning*), a subclass of e-learning providing knowledge ubiquitously and movably, effectively implements the concepts of *universal access* by supporting learning anywhere and anytime [55]. While e-learning, through ICT support for knowledge generation, storage, retrieval, delivery and visualization, strives to leverage the general knowledge level of (specific segments of) a population, m-learning provides mobile access for user knowledge retrieval, delivery and assessment. It should be noted that the mere usage of mobile devices generates the potentiality for including the widest possible parts of the population in the educational process, both in various, previously unforeseen situations, and in the most affordable way, hence reducing if also not avoiding the negative effects of the "digital divide". In this context the application of m-learning systems is quite natural, and foresees e-learning systems with support for client terminal mobility. These mobile devices (i.e. client terminals) through which the user accesses the system should in turn provide means to implement adaptability to her/his personality, both in the sense of interaction and of application, the latter referring to teaching methodology issues.

On the interaction side, m-learning heavily leans on portable devices of different capabilities that, in order of decreasing power include notebooks, tablet PCs – either "regular" or small form factor/UMPC (Ultra Mobile PC) ones, PDAs, smartphones and cellular phones. In this respect the less powerful mobile devices (e.g. "ordinary" cellular phones) mostly show restrictive factors such as small displays of usually low resolution, possibly only of a character-rendering type, as well as small phone-like numerical keyboards, what definitely hinders interaction.

As noted elsewhere for a period of time already [27], because of their very often prohibitive price tag high-end mobile devices are not that much popular amongst the common learner population, in spite of evident relative advantages regarding the user interface which in turn stems from better technical resources being offered by them. Therefore, a reasonable approach to m-learning introduction and operationalization would embrace cellular phones as the favorite mobile devices class to support this service [28]. This makes even more sense as all groups being addressed by m-learning, which cover a range between pupils/students in the regular education process and up to either employees seeking Life Long Learning (LLL) and/or jobless individuals requiring retraining, definitely possess and use cellular phones [31]. The elderly, as another possible target group for mobile applications which would include m-learning, should also be mentioned [30].

According to the recent annual report of the Croatian Agency for Telecommunications, Croatia in this respect shows a very good position, as the market penetration of mobile telephony amounts to 113.4% (12/2007 data), with an increase of 15% in the last year only. This fact places Croatia among EU countries (although not yet an EU member) at the quite high 12th place (even better than e.g. Austria!). On the other hand, the preponderant share (some 80%) of the mobile market is made by prepaid service users, which reasonably implies less prosperous people possibly possessing cheaper (and less capable) cellular phones. This in turn could be taken to imply a communication mostly based on messaging (SMS and MMS, with the former presently prevailing). However, high bandwidth mobile Internet access cover more than 49% of the national territory (data from 03/2008 referring to UMTS), and HSDPA has being offered since 06/2005, together providing a solid opportunity for more sophisticated m-learning services. To complete the picture, it should be noted that the number of phone connections being able to access the Internet using mobile networks amounts to 1.225 Mio (12/2007 data), some 48,000 of which use broadband access [50].

In accordance with the above our research within the context of m-learning interfaces has concentrated on mobile devices of the cellular phone class, advancing along two directions:

- insuring “technological” support to mobile (i.e. universal) access by specifically addressing limitations represented by lower-level and/or entry level mobile devices (specifically cellular phones), and
- making such mobile devices as usable as possible.

Regarding the former issue we are presently studying the ways to improve cellular phones' possible limitations concerning both limited (communication) channel capacity (e.g. supporting "pure" GSM devices)), by preprocessing knowledge contents to be transferred to client terminals (i.e. cellular phones) in order to overcome tedious and long-lasting user sessions with the remote m-learning server [39]. While this at first doesn't seem as a HCI-related issue, it certainly helps ensure a more pleasant working environment. The latter one focuses on improving the otherwise minimalistic cellular phones' interfaces by both trying to "augment" the display features of their screens and to "enrich" the ways of interaction [42].

The technology used herewith is strongly based on agents. As a matter of fact, the underlying m-system model and respective prototypes are being implemented as a multi-agent architecture [26], with the single agents relying on user and mobile terminal profiles databases, which provide for interface adaptivity to both mobile devices' capabilities and to user requirements. We plan to eventually develop a full-fledged intelligent interface that would blend with intelligent manipulation in the application (i.e. learning and teaching process) domain.

In the following some more detail is given to our research-in-progress regarding mobile interfaces.

4.1 Visibility Customization in Mobile Devices Adaptation

When addressing the problem of available display area, an acceptable balance between useful information and Mobile Device Application (MDA) controls is certainly

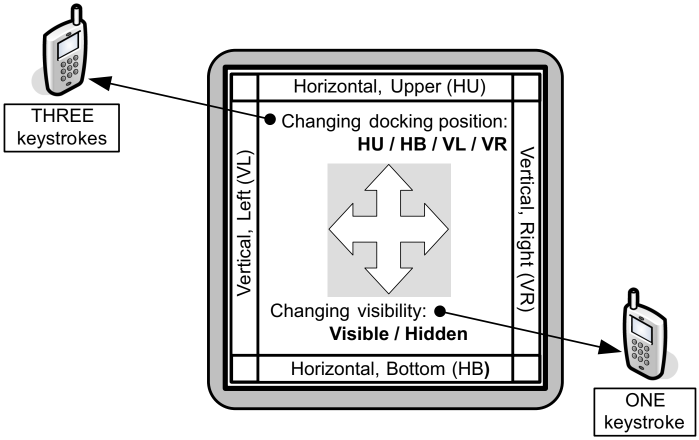


Fig. 3. Adjusting menu visibility and docking in an MDA

an important goal to achieve [13]. Usually controls in mobile devices are implemented as (hierarchical) menu systems, both character-oriented and graphical (iconic) ones, and can occupy quite a portion of the otherwise very limited resource; this implies the need for menu visibility control. Furthermore, the placement of controls, i.e. "menu docking", can also be a significant issue. While menus are customarily docked top-screen (horizontal upper position), a movable menu component offering adjustment to the other three positions (e.g. horizontal bottom, vertical left, and vertical right) means a level of flexibility that could show a leading edge.

Fig. 3 illustrates the above concepts in a simple mobile device environment, where menu visibility can be toggled by means of a keystroke, while menu docking can be achieved by three keystrokes [*ibid.*]. The above is basically a case of adaptable control (customization), and is tackled here in the context of functions to be inevitably included in a truly adaptive m-learning interface.

4.2 Mobile Interface Migration

Additional issues related to m-learning interfaces encompass lack of standardization and compatibility of mobile devices, along with limitations in accessing the Internet and the general problems with the communication bandwidth (price, availability and their relationship) [12]. Regarding the former topic, it is definitely important to expose m-learning users to a familiar look-and-feel, which would enforce an unobtrusive interaction, as the complexity resides in the MDA itself (i.e. the m-learning process). Such a seamless "device migration" in m-learning usage matches the criteria of ubiquitous learning, and can be implemented by separating the MDA proper (as the "application logic") from the user interface, this latter being adapted to a specific mobile device basing on the interpretation of an appropriate device-independent (usually XML-based) description by a "middleware" interface renderer tier, as illustrated in Fig. 4 [14].

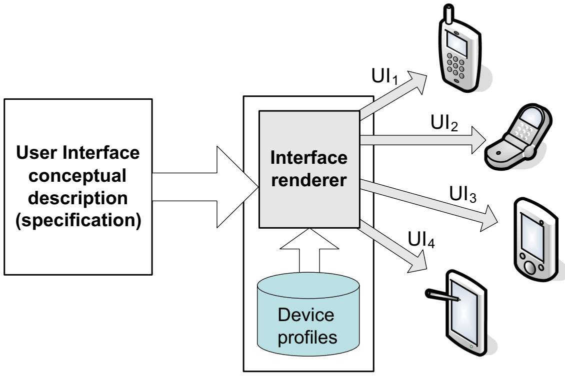


Fig. 4. Deployment of device-independent user interfaces

4.3 Towards Intelligent Mobile Interfaces: Personal Agents Mount the Charge

The ultimate objective to be addressed in this framework is clearly the automatic adaptation of the interface in accordance with both user preferences and level of knowledge, what leads to the use of agents and makes the m-learning system a *distributed agent-based system* that will perform adaptability of data services [15]. Within the framework of ITS-based education this can be achieved by assigning the tasks of classical ITSs to *personal agents*, which take over the eminently pedagogical chores (role of the human teacher, support to various education types, and search of the information space) but also those of adjusting the system to the device used for accessing it.

Adaptation to particular users is done by monitoring her/his interaction and running suitable adaptation algorithms, like (i) recency-based ones, monitoring the most-frequently used commands during the entire interaction, or (ii) frequency-based ones, calculating the frequency of usage of certain commands in shorter interactive cycles. By so doing, personal agents model a particular *user behavior profile* that could also anticipate future user objectives in specific tasks of the work process. As pointed out in Section 2, such intelligent interface can also include a suitable *user knowledge profile* [51].

Furthermore, it seems appropriate to confront the particular issue of m-learning interaction within a unified framework, which we denote as a *holistic approach to diversity in m-learning systems* that encompasses the three identified types of diversities, namely mobile devices, user domain knowledge and user interaction skills diversity [14]. This is in line with the main concepts of user sensitive design.

5 Usability Assessment for E-Learning System Interfaces

Usability evaluation plays a fundamental role in a human-centered design process, because it enables and facilitates design according to usability engineering principles. Basically, it is related to ease of use and ease of learning. There is no simple definition or meaningful single measure of usability. The majority of the employed assessment methods are usually grouped into (i) *usability test* methods, which involve end users, and (ii) *usability inspection* methods, which rely upon usability experts. Recent research has had a tendency to bring together these two main approaches, cf. [34]. Specifically, the approaches to *e-learning usability* range from those adapted to e-learning, e.g. [53], and up to those applying heuristics without special adjustment to the educational context e.g. [47]. As both an established set of heuristics and a joint evaluation methodology for e-learning systems do not exist yet, there is obviously a need for further research and empirical evaluation, e.g. [2].

Adaptive systems carry their particular set of issues regarding usability assessment. One of the key problems in their development is the inadequacy of available evaluation methods and techniques, and only a limited amount of empirical evaluations has been performed thus far. Moreover, there is still a lack of evaluation studies capable of distinguishing the adaptive features of the system from general usability [56]. Another crucial issue concerns what to quantify when evaluating adaptivity of adaptive systems, as parameters to be measured must be carefully considered in order to capture an exact indication of the interface's usability [32]. These, and possibly other issues, sometimes out-weight the benefits of adaptation [36].

Experimental results support the claims for adjustment to e-learning, e.g. [17]. Namely, aiming to evaluate how easily used and efficient the TEx-Sys family of systems is [54], we conducted an experiment employing a range of assessment methods, both empirical and analytic. The major strength in such an approach is the opportunity to supplement results from both the guideline evaluation and the empirical end user-based one, which is enhanced by users' feedback on their comfort while working with the system. The results go in hand with the assertion that we should not rely on isolated evaluations. Instead, usability assessment methods should be combined, obtaining different kinds of usability improvement suggestions.

The employed usability evaluation methodologies raised on the other hand a series of questions requiring further comprehensive research, the more so if employment of universal design within e-learning context is considered. [21]. Consequently, when designing an accessible and easy to use e-learning system, it is important to consider such key issues that include items encompassing learner-centered design paradigm, context of use approach, individualized approach, pedagogical framework as well as guideline framework. If so, then such an approach will be in accordance with the claim that in e-learning we do not need user interfaces supporting "doing tasks", but instead those supporting "learning while doing tasks", cf. [35].

Our recent research additionally explores the extent to which current and future *ubiquitous knowledge environments* [40] can be made sufficiently usable, accessible and smart, in order to support an inclusive information society, and the aspiration of universal access as well [1]. Using a range of converging methods to evaluate a random sample of such websites, we conclude that, while they act as substantial and functional repositories for knowledge, there still exist a potential for improvements, particularly in accessibility and smartness.

6 Discussion and Further Work

In the following we comment on the prototypes developed and tested, and provide some insights into our plans for further work in this area.

Especially in knowledge representation *MUVIES* provided us a relatively simple testbed for studying the appropriateness of multiple views as a technique to augment interaction of e-learning systems. With its concurrent presentation of different and semantically meaningful perspectives on a common internal representation of the same data the diversity of users could be addressed. The individual views simultaneously reflecting possible changes performed upon it proved valuable for a range of users covering elementary school pupils and up to university level students. The problems with its operation that were observed were mostly implementation rather than concept related, since the full three-dimensional view proved quite a "killer application" for most of client desktops in the common student PC laboratory.

As for adjustments of interaction style, *AKBB* worked quite well, adapting the different interface styles basing on knowledge about the individual user and the interaction session. The interface was run in a university level only environment, working properly in its somewhat crude adapting mode. As it could be remarked, the "swapping space", with its three different interface styles ask for a refinement, what would lead to a kind of flexible hybrid interface. In order to better conform to users' needs, we felt that both the set of parameters that control style swapping could be broadened (hence helping to better describe the user in the respective user profile) and the inference rules could be refined (so as to enable the system to make more graceful adaptations, i.e. in a less abrupt way).

Commenting on the user sensitive methods used, we note that the employed user model(s) should definitely be enhanced with personal characteristics that affect both learning and achieved learning outcomes, their identification being one of our next steps. Further research is also needed to determine whether an adaptive interface is measurably better than a non-adaptive one, and under what circumstances the benefit is more valuable than the apparent loss of control due to unexpected adaptations of the interface.

In the area of mobile interfaces, we continue our research targeting the implementation of an operable agent technology based m-learning system, to be tested in an in-class environment. As the first *in vivo* experiment, we plan to apply this system in "class participation" testing, sending the students (through their cellular phones) short test quizzes in order to check their involvement in the lectures. Other applications would include preparation for laboratory exercises, by running an elementary digital design virtual laboratory. For the moment being, we feel that a *MUVIES* style mobile interface is, although provably effective and efficient especially for running simulations (cf. [29]), out of consideration because of limitations inflicted by the use of "ordinary" cellular phones.

7 Conclusion

A number of important and mostly practical reasons have prompted the development and operationalization of e-learning systems since the introduction of the first computers until the present days, resulting in modern implementations that adjust their operation to users' needs, hence additionally making the educational process more efficient and effective. Along with these primarily functional achievements, user friendliness is definitely the issue computer scientists and psychologists/sociologists should strive to accomplish.

In this paper we outline the research directions in the field of interfaces for e-learning systems, which are being studied at our home institutions and witness our personal involvement and contributions, along with the results thus far achieved. In addressing the HCI research for e-learning systems we described some of the adaptivity issues of interest which have been of our primary concern, together with some related usability evaluation results. Regarding the former ones, we tackled adaptation of knowledge representation and of interaction style, as well as adaptation to mobile requirements, m-learning showing off as the next cornerstone of educational technology. Since conceptualization and development means nothing without a proper evaluation of the ideas conceived and artifacts realized, we also provided an overview of the applied usability assessment methods and techniques.

The main hypothesis of our work is that intelligent interaction as adaptation to user individual characteristics and needs, as experienced through her/his behavior during interaction, is the device for accomplishing a more efficient, effective *and* humane ICT-supported education, while this latter is in turn the principal means in achieving knowledge society. This is especially true from the standpoint of a developing transition country, as e.g. Croatia is [6].

On the other hand, e-learning and the respective e-learning systems are definitely "the one" instruments to overcome the present and foreseen problems in reforming its (higher) education system, particularly since only a comprehensive application of Information and Communication Technology (ICT) artifacts can insure the addressing of the widest possible learner community, along with bringing to the individual learners/students the necessary educational contents in a way that is tailored to their capabilities. Eventually this would lead to the creation of a body of knowledgeable and capable experts/specialists, who would be fully included in Life Long Learning, thus effectively and efficiently help realize the transition to the knowledge society.

Acknowledgments. This work has been carried out within projects 036-0361994-1995 *Universal Middleware Platform for Intelligent e-Learning Systems* and 177-0361994-1998 *Usability and Adaptivity of Interfaces for Intelligent Authoring Shells*, both within the program 036-1994 *Intelligent Support to Omnipresence of e-Learning Systems*, and funded by the Ministry of Science and Technology of the Republic of Croatia.

References

1. Adams, R., Granić, A.: Creating Smart and Accessible Ubiquitous Knowledge Environments. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 3–12. Springer, Heidelberg (2007)
2. Ardito, C., Costabile, M.F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T., Rossano, V.: An approach to usability evaluation of e-learning applications. Universal Access in Information Society 4(3), 270–283 (2006)
3. Benyon, D., Crerar, A., Wilkinson, S.: Individual Differences and Inclusive Design. In: Stephanidis, C. (ed.) User Interfaces for All – Concepts, Methods, and Tools, pp. 21–46. Lawrence Erlbaum Associates, Inc., Mahwah (2001)
4. Benyon, D., Murray, D.: Interacting with Computers. Special issue on intelligent interface technology 12 (2000)

5. Benyon, D., Murray, D.: Editorial. Special issue on intelligent interface technology: editor's introduction. Interacting with Computers 12, 315–322 (2000a)
6. Budin, L., Bajica, M., Carić, A., Čerić, V., Glavinić, V., Lovrek, I., Manger, R., Ursić, S.: Croatia in 21st Century. Information and Communication Technology, Office for the Strategy of Development of the Republic of Croatia, Zagreb, Croatia (in Croatian) (2001)
7. Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): Adaptive Web 2007. LNCS, vol. 4321. Springer, Heidelberg (2007)
8. Burns, H., Capps, C.: Foundations of intelligent tutoring systems: an introduction. In: Polson, M., Richardson, J. (eds.) Foundations of Intelligent Tutoring Systems, pp. 1–18. Lawrence Erlbaum Associates Publishers, Mahwah (1988)
9. Dieterich, H., Malinowski, U., Kühme, T., Schneider-Hufschmidt, M.: State of the Art in Adaptive User Interfaces. In: Schneider-Hufschmidt, M., Kühme, T., Malinowski, U. (eds.) Adaptive User Interfaces: Principles and Practice, pp. 11–48. Elsevier Science B.V. Publishers, North-Holland (1993)
10. Egan, D.: Individual Differences in Human-Computer Interaction. In: Helander, M. (ed.) Handbook of Human-Computer Interaction, pp. 543–568. Elsevier Science B.V. Publishers, North-Holland (1988)
11. Ford, N., Chen, S.Y.: Individual Differences, Hypermedia Navigation and Learning: An Empirical Study. Journal of Educational Multimedia and Hypermedia 9(4), 281–311 (2000)
12. Glavinić, V., Kukec, M., Ljubić, S.: Mobile Virtual Laboratory: Learning Digital Design. In: Luzar-Stiffler, V., Hljuz Dobric, V. (eds.) Proceedings of the 29th International Conference Information Technology Interfaces (ITI 2007), Cavtat –Dubrovnik, Croatia, pp. 325–332 (2007)
13. Glavinić, V., Ljubić, S., Kukec, M.: Transformable Menu Component for Mobile Device Applications: Working with both Adaptive and Adaptable User Interfaces. International Journal of Interactive Mobile Technologies (iJIM) 2(3), 22–27 (2008a)
14. Glavinić, V., Ljubić, S., Kukec, M.: A Holistic Approach to Enhance Universal Usability in m-Learning. In: 2nd International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2008), Valencia, Spain (2008b) (accepted for presentation)
15. Glavinić, V., Rosić, M., Zelić, M.: Agents in m-Learning Systems Based on Intelligent Tutoring. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 627–636. Springer, Heidelberg (2007)
16. Glavinić, V., Granić, A.: Interacting with Educational Systems Using Multiple Views. In: Kalpić, D., Hljuz Dobrić, V. (eds.) Proceedings of the 23rd International Conference 'Information Technology Interfaces (ITI 2001), Cavtat –Dubrovnik, Croatia, pp. 125–131 (2001)
17. Granić, A.: Experience with Usability Evaluation of e-Learning Systems. Universal Access in the Information Society (2008) DOI 10.1007/s10209-008-0118-z
18. Granić, A.: Foundation of Adaptive Interfaces for Computerized Educational Systems. Ph.D. Diss. University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia (in Croatian) (2002)
19. Granić, A., Adams, R.: User Sensitive Research in e-Learning: Exploring the Role of User Individual Characteristics. Special Edition on Universal Access in the Information Society on Innovations in User Sensitive Design, Research and Development (accepted for publication) (2008)

20. Granić, A., Stankov, S., Nakić, J.: Designing Intelligent Tutors to Adapt Individual Interaction. In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 137–153. Springer, Heidelberg (2007)
21. Granić, A., Ćukušić, M.: Universal Design within the Context of e-Learning. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 617–626. Springer, Heidelberg (2007)
22. Granić, A., Nakić, J.: Designing intelligent interfaces for e-learning systems: the role of user individual characteristics. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 627–636. Springer, Heidelberg (2007)
23. Granić, A., Glavinić, V.: Adaptive Interface of a Human Teacher Emulator. In: Salvendy, G. (ed.) Proceedings of the 11th International Conference on Human-Computer Interaction (HCII 2005)/3rd International Conference on Universal Access in Human-Computer Interaction: Exploring New Dimensions (UAHCI 2005), Mahwah, USA, pp. 1–10 (2005) (CD-ROM)
24. Granić, A., Glavinić, V.: Automatic Adaptation of User Interfaces for Computerized Educational Systems. In: Zabalawi, I. (ed.) Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2003), Sharjah, Dubai, pp. 1232–1235 (2003)
25. Gregor, P., Newell, A.F., Zajicek, M.: Designing for dynamic diversity – interfaces for older people. In: Jacko, J.A. (ed.) ASSETS 2002. 5th International ACM Conference on Assistive Technologies, Edinburgh, Scotland, pp. 151–156 (2002)
26. Hodžić, J.: Intelligent Content Adaptation in e-Learning Systems, Diploma Thesis (non-Bologna program), University of Zagreb, Faculty of Electrical Engineering and Computing (in Croatian) (2008)
27. Holzinger, A., Nischelwitzer, A., Meisenberger, M.: Lifelong-learning support by m-learning: example scenarios. eLearn Magazine 2005(12), 2 (2005)
28. Holzinger, A., Nischelwitzer, A., Meisenberger, M.: Mobile Phones as a Challenge for m-Learning: Examples for Mobile Interactive Learning Objects (MILOs). In: Tavangarian, D. (ed.) 3rd Annual IEEE Int'l Conf. on Pervasive Computing and Communications (PerCom 2005 Workshops), pp. 307–311 (2005)
29. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)
30. Holzinger, A., Searle, G., Nischelwitzer, A.: On some Aspects of Improving Mobile Applications for the Elderly. In: Stephanidis, C. (ed.) Coping with Diversity in Universal Access, Research and Development Methods in Universal Access. LNCS, vol. 4554, pp. 923–932. Springer, Heidelberg (2007)
31. Holzinger, A., Nischelwitzer, A.K., Kickmeier-Rust, M.D.: Pervasive E-Education supports Life Long Learning: Some Examples of X-Media Learning Objects. In: 10th IECEE World Conf. on Continuing Engineering Education, WCEE (2006) (last accessed 18.9.2008), https://online.tu-graz.ac.at/tug_online/voe_main2.getVollText?pDocumentNr=72401&pCurrPk=21958
32. Höök, K.: Evaluating the Utility and Usability of an Adaptive Hypermedia System. Knowledge-Based Systems 10(5), 311–319 (1998)
33. Höök, K.: Steps to Take Before Intelligent User Interfaces Become Real. Interacting with Computers 12, 409–426 (2000)
34. Hornbæk, K.: Current practice in measuring usability: Challenges to usability studies and research. International Journal of Man-Machine Studies 64(2) (2006)

35. Hsi, S., Soloway, E.: Learner-centred design: Addressing, finally, the unique needs of learners. In: Computer Human Interaction 1998, CHI 1998, Los Angeles, USA, pp. 211–212 (1998)
36. Jameson, A.: User Modeling Meets Usability Goals. In: Ardissono, L., Mitrović, A. (eds.) User Modeling: Proceedings of the Tenth International Conference. Springer, Berlin (2005)
37. Jennings, F., Benyon, D., Murray, D.: Adapting systems to differences between individuals. Acta Psychologica 78, 243–256 (1991)
38. Kobsa, A.: Supporting User Interfaces for All through User Modeling. In: 6th International Conference on Human-Computer Interaction HCI International, Yokohama, Japan, pp. 155–157 (1995), http://www.ics.uci.edu/~kobsa/papers/1995-HCI95-kobsa.pdf
39. Kukec, M.: Model of Semantic Based Preprocessing of Domain Knowledge to Support a Minimalistic e-Learning System Interface, MSc Thesis (non-Bologna program), University of Zagreb, Faculty of Electrical Engineering and Computing (research in progress) (2008)
40. Larsen, R.: Knowledge Lost in Information: Report of the NSF Workshop on Research Directions for Digital Libraries. NSF Post Digital Library Futures Workshop Chatham, Mass (2004) (June 15-17, 2003), http://www.sis.pitt.edu/~dlwkshop/report.pdf
41. Lieberman, H.: Introduction to intelligent interfaces (1997), http://web.media.mit.edu/~lieber/Teaching/Int-Int/Int-Int-Intro.html
42. Ljubić, S.: Model of an Adaptive Minimalistic Interface Supporting Universal Usability of e-Learning Systems, MSc Thesis (non-Bologna program), University of Zagreb, Faculty of Electrical Engineering and Computing (research in progress) (2008)
43. McTear, M.F.: Intelligent interface technology: from theory to reality? Interacting with Computers 12, 323–336 (2000)
44. Murray, T.: Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art. International Journal of Artificial Intelligence in Education 10(1), 98–129 (1999)
45. Norman, D., Draper, S.W. (eds.): User Centered System Design. Earlbaum, Hillsdale (1986)
46. Pantazis, C.: Executive summary: A vision of E-learning for America's workforce. Report of the Commission on Technology and Adult Learning, ASTD (2001) (last accessed: 2007-08-17), http://www.learningcircuits.org/2001/aug2001/pantazis.html
47. Parlangeli, O., Marchigiani, E., Bagnara, S.: Multimedia systems in distance education: effects of usability on learning. Interacting with Computers 12, 37–49 (1999)
48. Pullin, G., Newell, A.: Focussing on Extra-Ordinary Users. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4554, pp. 253–262. Springer, Heidelberg (2007)
49. Reiss, S.: PECAN: Program Development Systems that Support Multiple Views. IEEE Transaction on Software Engineering SE-11(3), 276–285 (1985)
50. Republic of Croatia, Croatian Agency for Telecommunications: Annual Report on the Operation of the Croatian Agency for Telecommunications and on the State and Implementation of Concepts and Goals of Regulation of the Telecommunications Market in the Republic of Croatia in Year 2007 (March 2007) (last accessed 3.8.2008), http://www.telekom.hr/lgs.axd?t=16&id=909
51. Rosić, M., Glavinić, V., Stankov, S.: Intelligent Tutoring Systems for the New Learning Infrastructure. In: Lytras, M.D., Naeve, A. (eds.) Intelligent Learning Infrastructures for Knowledge Intensive Organizations: A Semantic Web Perspective, ch. VIII, pp. 225–250. Idea Group Inc./InfoSci – Information Science Publishing, Hershey (2005)

52. Schneider-Hufschmidt, M., Kühme, T., Malinowski, U. (eds.): Adaptive User Interfaces: Principles and Practice. North-Holland/Elsevier Science Publishers B.V (1993)
53. Squires, D., Preece, J.: Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. Interacting with Computers 11, 467–483 (1999)
54. Stankov, S.: Isomorphic Model of the System as the Basis of Teaching Control Principles in the Intelligent Tutoring System, Ph.D. Diss., University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture (in Croatian) (1997)
55. Stephanidis, C.: Editorial. International Journal of Universal Access in the Information Society 1(1), 1–3 (2001)
56. Weibelzahl, S.: Problems and pitfalls in the evaluation of adaptive systems. In: Chen, S., Magoulas, G. (eds.) Adaptable and Adaptive Hypermedia Systems, pp. 285–299. IRM Press, Hershey (2005)

Personalized E-Learning through Environment Design and Collaborative Activities

Felix Mödritscher and Fridolin Wild

Institute for Information Systems and New Media,
Vienna University of Economics and Business Administration
Augasse 2-6, 1090 Vienna, Austria
{felix.moedritscher,fridolin.wild}@wu-wien.ac.at

Abstract. Over the last century, many theoretical frameworks and technological solutions for personalized e-learning have emerged. The underlying models, however, are often based on the practice that domain experts develop an adaptation strategy to personalize content or parts of a learning platform, which leads to different problematic aspects decreasing the feasibility or utility of such approaches. After giving a brief overview of the historical development and basic concepts of personalized e-learning, we outline the shortcomings of the traditional 'top-down, ex ante' models and present an alternative approach which deals with personal learning environments, web application mashups, learning activities and learner interactions, as well as pattern-based best practice sharing. Furthermore, a prototypic implementation for our 'learner-driven, bottom-up' approach to personalized e-learning, namely the 'Mash-UP Personal Learning Environment' (MUPPLE), is presented and discussed on the basis of a concrete scenario.

Keywords: Personal Learning Environments, Learning Environment Design, Learner Interaction Scripting, End-User Development.

1 Introduction

According to [1], personalized learning aims at "*tailoring the teaching to individual need, interest and aptitude*" to ensure the most effective knowledge transfer for each learner. In fact, the idea of adapting instructions to learner characteristics has been considered a success factor at least since the 4th century BC, and adaptive tutoring was a wide-spread method of education in the 18th century [2]. Moreover, a first prototypic implementation of an adaptive assessment tool was already reported in 1926, while in the last four decades many technology-based approaches and solutions to personalized learning have emerged [3].

Historically, personalized e-learning is founded on the aptitude-treatment interaction (ATI) research as well as macro and micro-adaptive instructional models. In practice, these streams lead to technologies like Computer-Managed Instruction (CMI), Intelligent Tutoring Systems (ITS), Adaptive Educational Hypermedia (AEH), but they also influenced Learning Management Systems (LMS) and e-learning standards. Referring to the relevant literature [4, 5, 6], personalized adaptive e-learning

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 377–390, 2008.

typically include four types of models: (1) the domain model to describe learning resources and knowledge domain, (2) the pedagogical (learner) model to characterize the learning context and learner states, (3) the didactical model to consider typical teaching aspects, like learning goals, course sequences, didactical requirements, etc., and (4) the adaptation model (rules) to specify the personalization strategies.

Commonly-known frameworks for personalized e-learning address different aspects, such as the conceptual idea (cf. the framework for adaptive e-learning by [4]), the architectural design of an AEH (cf. KnowledgeTree [7]) or a formalization of adaptive behavior (cf. the first-order logic by [8]). Younger trends also focus on standardizing personalization rules, e.g. by using XML-based specifications like IMS Learning Design [9]. However, all these frameworks and their underlying approaches prevent personalized adaptive e-learning from being utilized or utilizable in educational practice. In the next section we discuss the shortcomings of current approaches and examine alternatives based on new concepts and methods from the Internet. Section 3 summarizes our idea of a 'Mash-UP Personal Learning Environment' and a first prototypic implementation. Thereafter, section 4 describes personalization of learning with our prototype on the basis of an exemplary scenario.

2 Shortcomings of Personalized E-Learning and Novel Influences

In learning and research practice, traditional approaches to personalized e-learning still lack important issues as outlined in the following.

First of all, the current definition of personalized (adaptive) e-learning is often restricted to the context of one user interacting with instructions delivered by one system (e.g. the LMS) which, furthermore, contains a pre-defined, up-to-date learner model and automatically adapts to the learner (cf. [3, 4, 6, 10, 11]). In our opinion, this definition is not sufficient to meet the requirements of the real world, and the 'learning environment' includes all possible entities a learner interacts with and all influences on the learning process. More provokingly, we would even say that the learning environment comprises 'everything but the learner'. For the context of e-learning we restrict this definition to all tools on the computer utilized by a learner. Consequently, this point of view widens the scope of adaptivity and personalization. For instance, the knowledge about a learner, a so-called user model, might be distributed over several systems and known by a peer or the facilitator only. Additionally, adaptive behavior is not only observable in one specific system but has to be seen in connection with a learner utilizing various tools to connect to a learning network and collaborate with other actors on shared artifacts. As a result, adaptivity and personalization take place within the whole socio-technical system, the learning network, and not only in one educational system and on the basis of a specific learner model. The adaptation effects, however, are only visible at the frontend of this socio-technical system, precisely at the user interface displaying the learning tools.

Secondly, technology-driven personalized e-learning is based on ex ante, top-down modeling (e.g. modeling of learning styles and cognitive traits depicted in [12]), requiring technological and pedagogical experts to turn a valid adaptation strategy into functions and systemic behavior of an e-learning system, primarily following the paradigm of didactic-awareness [13] and considering aspects of adaptable courseware [14]. However, such strategies to automated adaptation of learning content and functions might lack of validity, e.g. if learners have to self-assess their learning style [15] or if wrong pedagogical assumptions have been made, as shown with a study on interactive media in [16]. Furthermore, new trends in the field of learning research (cf. [17]) try to address the perspective of learners and suggest a learner-driven approach to personalized e-learning, e.g. by analyzing learning behavior and recommending learning experiences of a learning community to learners.

Thirdly, researcher and developer in the field of personalized e-learning often build upon Learning Management Systems (LMSs) or Virtual Learning Environments (VLEs), aiming at adapting parts of such systems, like navigational elements [18] or systemic behavior by orchestrating services [5]. Some approaches even deal with opening up LMSs to externally provide adaptive behavior [19]. Concerning the technological realization, the implementation of personalization in monolithic learning environment requires high efforts in software development and evaluation and often ignores existing tools which are used by learners and might be more efficient in a certain learning context. Younger and more promising approaches focus on the perspectives of learners in terms of personal learning environments composing different learning services into a single user experience [20].

Finally, interoperability is of importance for both e-learning and personalized e-learning, particularly if learners actively contribute to the learning process. If not providing interoperability mechanisms [21] or considering standardized learning content [14, 22], personalisable courseware might get isolated or even be lost within learning platforms. On the other hand, standardizing adaptable courseware extremely increases the complexity and the effort of creating and planning a course, as indicated with a study in [23]. In addition to considering courseware design principles, a heterogeneous landscape of learning tools and services, collaborative learning activities, and learner interaction sequences, facilitators have to deal with learning design and, additionally, with selecting appropriate adaptation models and techniques [18] and extending their courseware, particularly on the basis of existing standards or beyond [14].

All in all, facilitating personalized e-learning experiences can be characterized with technological boundaries, restrictions of existing standards and specifications, as well as with more complexity and efforts for educators. New developments in the Internet, subsumed with the term Web 2.0 [24], aim at overcoming these problematic aspects of e-learning. The effects of the Web 2.0 on technology-enhanced learning have already been examined elsewhere, e.g. in [17, 25]. Following these principles, we propose that, in analogy to Web 2.0 principles, personalized e-learning should be based on '*the Web as a learning space*', allowing learners to use a variety of available tools and content. By providing '*rich learning experiences*' through more interactive user interfaces or community-enabling features, learners can collaborate with peers and actively participate in the learning process, e.g. using blogging or tagging functionality. On the basis of collaborative learning activities and of open, high-quality content (‘*the next Intel inside for learning*’), learners and facilitators can ‘*add value to their learning processes*’, for instance by commenting or tagging learning material or contributing content.

On the other hand, it is also possible to analyze learning behavior in order to ‘*harness the collective intelligence of a networked learning community*’ if personalized e-learning is not restricted to learner interactions with one specific system. Hereby, beneficial semantics for other peers can be provided either by the learners themselves, e.g. by sharing learning experiences with others, or through automated mining techniques trying to extract and exploit the ‘*network effects of a learning community*’, e.g. by recommending tools for learning activities. Finally, from a more technological point of view, these new influences from the Web 2.0 require new development methods for ‘*software above the level of a monolithic LMS*’, being based on ‘*lightweight programming models*’ like RESTful architectures [26], going beyond ‘*the software release cycles for LMSs*’, and considering complex socio-technical processes [27] in order to realize personal learning environments [20]. Thus, the ‘*long tail of software*’ [28] describes the Web 2.0 idea that learners design their own personal learning environments on the basis of available learning tools and services and according to related learning experiences of peers, if given.

Comparing traditional approaches to our idea of ‘**personalized e-learning 2.0**’, we identified significant advantages of this Web 2.0-driven development. Above and beyond, the personalization strategy is shifted from being implemented by a few domain experts and facilitators (ex ante, top-down modeling) to empowering learners to design their personal learning environments, to collaborate with peers, and to transfer learning experiences between facilitators and peers, overall leading to a learning network of actors, artifacts, and activities. Based on learner interactions, this networked community of online learners can be supported by more sophisticated strategies, e.g. for regulating collaborative learning or for reflecting the learning process (learner-driven, bottom-up, just-in-time modeling). Particularly, this would decrease the planning and implementation efforts of personalized e-learning, because modeling is less deterministic and the necessary models which are distributed and partially even outsourced can later be created, through involvement of the learners and on the basis of a valid learning activity design. Moreover, each model can be developed iteratively and, if stable enough, evaluated separately from other models.

While traditional personalized e-learning approaches seem to address a declarative design of procedural experiences, our bottom-up approach aims at designing procedures for creating declarative artifacts, evolving typical learn-by-heart or know-how-to-do goals (low-level learning objectives) to the learn-how-to-learn vision (higher-level learning objectives). Thereby, the idea of the 2.0-driven personalized e-learning also considers that learners can reuse learning experiences by adaptive sharing, cloning, or prototyping learning activities, instead of implementing and partially refining them. Last but not least, we foster the possibility that learners can bring in existing tools and content, instead of working with a given LMS and predefined resources to master teacher-given activities. Concluding this section, we believe the above-mentioned considerations to present a promising approach to personalized e-learning, particularly to enable lifelong learning beyond isolated learning contexts like higher education or workplace learning.

3 The Idea of a Mash-UP Personal Learning Environment

To show personalized e-learning 2.0 in practice, we present the basic concept as well as a first prototypical implementation of a 'Mash-UP Personal Learning Environment' (**MUPPLE**). Hereby, we start off with low-level aspects, such as the technological infrastructure, continue with learner interaction issues, and end up with high-level, learning-related issues behind our idea.

3.1 Technological Infrastructure

As a first step towards a technological infrastructure for a new generation of personalized e-learning solutions, we build upon the Web 2.0 idea of mashups. Hereby, we propose a so-called web application mashup [29] as one possible infrastructure for personal learning environments; learners may also use a portal-like platform with different widgets or different applications on their computer. Extending the idea of traditional mashups, a web application mashup allows displaying various web-based tools into one aggregated view within the browser. Such a solution approach needs to consider the following issues:

- Concluding from mashup visualization techniques [30], the display of different applications next to each other requires a certain (1) *cognitive support for users* (facilitators and peers!) in order to reduce their cognitive load on working with the system. In accordance with iGoogle, Netvibes or other providers of personalized websites, we realized a portal-like OpenACS component, namely the XoMashup application [29], which allows users to arrange tools along a grid layout.
- Addressing (2) *controllability* in the field of personalized e-learning [31], a web application mashup has to give the control over the arrangement of and interaction with the tools to a user. Therefore, our XoMashup component allows a user to rearrange, minimize, maximize, reload and close each window.
- Furthermore, it is possible to launch web applications and even add new ones to the mashup space. As usual, browser-based solutions do cause (3) *technical restrictions*. In our case, it is necessary to start full web applications with all its scripts and style-sheets as a part of the mashup page. Thus, we implemented our mashup solution on the basis of 'iframes'. This may be the only way to guarantee an own environment for each tool but may not be supported by all browsers. Further, the usage of iframes enforces the prevention of DOM operations which would reset the content of an iframe. Consequently, the grid-based windowing system of XoMashup is realized with absolute positioning and the manipulation of CSS directives.

All in all, the web application mashup solution allows learners to reuse existing (web-based) tools and services and can be considered as a technological infrastructure for our approach. Moreover, web application mashups are very flexible and, therefore, useful for many other application areas as well. Nevertheless, without some kind of underlying semantics like necessary models for personalizing the learning process, the XoMashup component would be nothing more than a personalizable (customizable) portal system, lacking pedagogical support for learners such as guidance or reflection.

3.2 Learner Interaction Model

Therefore, we built up a learner interaction model for describing how learners can design their personal learning environment and interact with it. The following aspects were taken into consideration for this model:

– In order to be (4) *independent of a subject domain*, we applied the Activity Theory model in a similar way as manifested for the INCENSE system [32]. Basically, we broke down the learning context into situations which describe the physical and social environment of learners. In such a situation, a learner is engaged in a so-called activity which consists of actions and objects (artifacts or other outcomes) and includes tools (or tool combinations) and other actors (facilitators or peers, even in multiple roles). Such a learning activity is meant to be our basic instructional entity where learners actively experience a domain and construct knowledge.
– This notion of a learning activity is (5) *simple and understandable for learners*, thus is considered to be of importance for a (6) *scrutable systemic behavior* [33] and a good basis for experiencing further personalization strategies.
– To enable (7) *reflective learning* [34](p.7), we decided to bind each action to one specific tool and one specific object in order to produce one outcome. Although different actors can work on the same action and even produce the same outcome, each learner only sees her own actions, and all started actions are visualized together with the corresponding object and tool.
– Additionally, these action-object-tool triples are recommended to peers on defining and starting new actions as a certain (8) *learner support*. However, learners are able to overwrite decisions and recommendations given by the system and may build up their personal learning environment by defining own actions and objects, bringing in own tools, and going through the actions in their own sequence to achieve the outcomes.
– Addressing (9) *learnability and efficiency*, our learner interaction model is implemented in the form of a domain-specific language called 'Learner Interaction Scripting Language' (LISL). Table 1 shows an example of a simple activity consisting of two actions. First, the learner is expected to record a short self-description with the tool VideoWiki (http://distance.ktu.lt/videowiki/), whereby the REST-based call for this action has to be specified and the URL for the object 'self-description' is determined by completing the action. Second, the learner should go through the self-descriptions of the peers by accessing a predefined URL, e.g. the collection containing all self-descriptions.

Table 1. Exemplary LISL code for the activity 'Getting to know each other' consisting of two action statements, each one bound to one object and one tool

- define action Compose with url http://[...]
- define action Browse
- define object 'self-description'
- define object 'descriptions of peers' with url […]
- define tool VideoWiki with url http://[...]
- Compose 'self-description' using VideoWiki
- Browse 'descriptions of peers' using VideoWiki

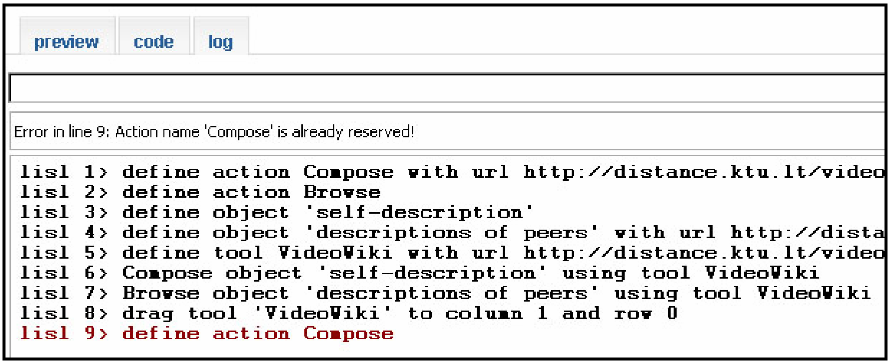


Fig. 1. Web-based LISL interpreter displaying the interpreted LISL script from Table 1 (red lines indicate errors, whereby the error message is shown below the command input field)

We build a web-based interpreter for LISL code, as can be seen in Fig. 1. This scripting approach allows experienced users to code their learning activities very efficiently, while novices can use web-based control widgets and dialogs which act as a wrapper for the LISL statements and are shown in the upcoming section.

3.3 Higher-Level Learning Paradigms

On top of the mashup infrastructure and the learner interaction model, higher-level learning paradigms address important issues for building and sustaining networked communities of learners:

– With regard to [35], the constructivistic-collaborative approach to adaptive e-learning deals with aspects of the (10) *active participation of learners*, motivational factors like self-esteem and (11) *collaborative activities*. Technologically, such considerations lead to a high degree of interactivity also found in games and simulations, to adaptation strategies for motivating learners, e.g. by pedagogical agents, or to adaptivity through collaboration, most prominently addressed by the field of Computer-Supported Collaborative Learning (CSCL). As already mentioned, the MUPPLE approach stresses a method to build and sustain a learning network of actors, artifacts, and activities, which increases the motivation to learn and aims at developing more complex competencies [36]. Similar to Web

2.0-driven platforms like Facebook, learners require facilities to get involved into collaborative activities and to regulate collaboration and social interactions with peers.
- Particularly for attracting new users, the success of MUPPLE highly depends on recognizable benefits for learners. Hereby, we foster the paradigm of (12) *best practice sharing* on the basis of activity patterns which can be provided by facilitators and learners to be shared within the community. Similarly to the idea of scripting collaborative activities [37], we apply the LISL scripting language to describe these activity patterns. As learning activities are encoded in form of LISL scripts, they can be exported into activity patterns and shared with other learners; vice versa, peers can use available best practices and create their own activities out of these patterns.
- Beside learner-driven best practice sharing, the bottom-up approach of MUPPLE supports the (13) *analysis of former learning scripts* in order to personalize different aspects of learning, e.g. by recommending action-object-tool triples to inexperienced learners or suggesting tool landscapes for certain activities. Such personalization strategies also address the learnability and efficiency of applying MUPPLE in practice.

4 Personalized Learning with MUPPLE

Based on our wider understanding of personalized e-learning, we build a first prototype considering the issues mentioned so far. Fig. 2 shows a screenshot of the learner's view on our exemplary activity 'Getting to know each other' which consists of the two actions 'Compose self-description' and 'Browse descriptions of peers' (cf. table 1). On the top, the header displays the activity currently opened. To the left-hand side, learners are supplied with an overview of their own activities and can navigate through them. By clicking on it, a learning activity is loaded and displayed in the content area; additionally, a branch with all action-object-tool triples included is opened simultaneously. The content area provides three different view modes of a MUPPLE page. By choosing one of the three tabs, a learner has a view on the web application mashup ('preview'), an editor for the LISL code of this page ('code') or the LISL interpreter ('log'). This structure of a MUPPLE page is related to principles of end-user development, which is closer examined in [38].

The LISL interpreter does not only show the interpreted code, but also highlights possible errors with detailed explanations and allows entering single lines of code (cf. Fig. 1). The preview mode, on the other hand, comprises an integrated view of all learning tools launched so far. Each tool is located within an own window (a so-called 'Mupplet') with the control elements mentioned in the last section ('reload original URL', 'minimize', 'maximize', and 'close') on the upper right side. Creation facilities on the left-hand side allow creating new activity pages from blank or from given patterns. Furthermore, it is possible to add action-object-tool triples to an opened MUPPLE page, whereas possible values are recommended on the basis of all other

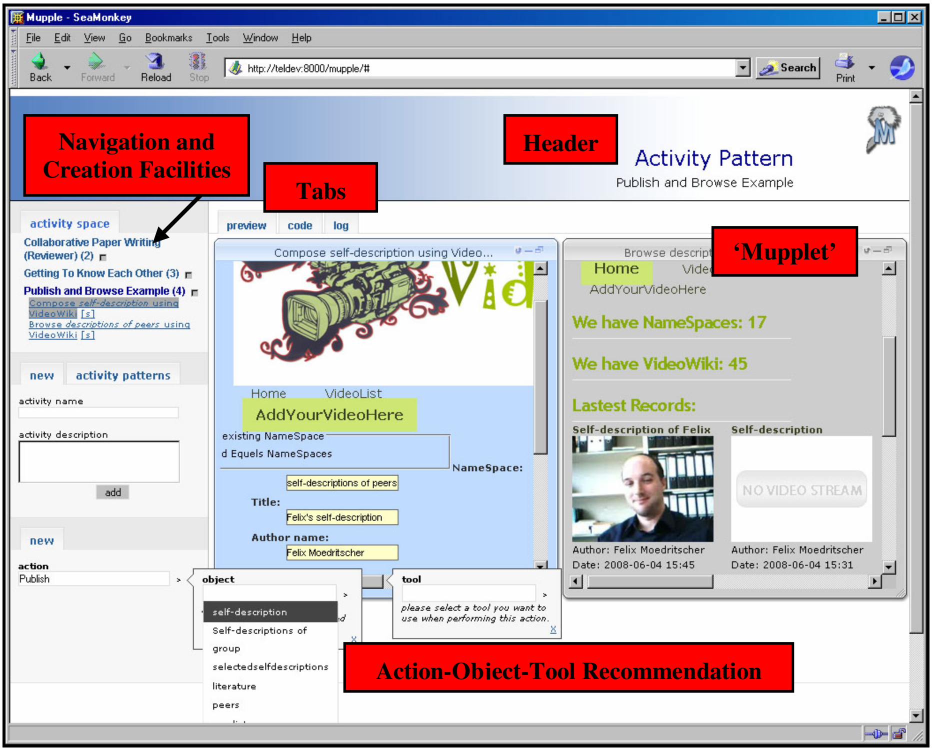


Fig. 2. User view on activity 'Getting to know each other', where the screen consists of a header (top), facilities for navigation and activity management (left), and the content area with its three tab-views (the mashup space, the LISL code editor, and the command-line interpreter).

activity pages. A ranking strategy has not been implemented for these recommendations so far. However, features like the closeness between activities (e.g. through derivation from the same pattern), the action-object-tool binding or social relations in the learner network may present useful factors to rank and filter values if there are too many of them.

In the following, a simple scenario is described to explain how personalization of the learning process takes place on working with the MUPPLE prototype: Consider a group of students distributed over various universities in different countries. To introduce the students to each other, the facilitator decides to use MUPPLE and predefines an activity pattern for this purpose. Based on former experiences with such a socialization exercise, this pattern includes the following action statements:

– Compose 'self-description' using VideoWiki
– Browse 'descriptions of peers' using VideoWiki
– Contact 'two peers' using WebMail
– Delete 'spam entries' using VideoWiki

After introducing MUPPLE to the students, the facilitator invites them to create a MUPPLE page from the given pattern. Now, each student instantiates a learning activity 'Getting to know each other' from the pattern and, furthermore, customizes it according to her own need. For instance, if students do not need the 'delete' action and might not even have appropriate permissions within the VideoWiki tool, they simply can remove this action from their pages. A possible result could be the MUPPLE page shown in Fig. 2. On the other hand, the facilitator who obviously has a different role in this collaborative activity uses the action 'delete' to clean up the collection containing the self-descriptions from spam regularly. With respect to controllability, a student might want to use a web-based chat tool to communicate with peers. So, she could bring in this tool and remove the original one, either by modifying the LISL code or using web-based control widgets. Finally, personalization takes place on creating new action statements if MUPPLE recommends action-object-tool triples from peers. Consequently, MUPPLE also might provide observation facilities to inform users about changes in the activities of peers or the facilitator, which are currently not realized.

Beside personalization within the learning activities, MUPPLE and its pattern-based best practice sharing method allow personalized learning in a broader sense. First, improved patterns can be derived from the learning activities experienced in practice, which, in our opinion, might increase the quality of the content and activities within MUPPLE. Second, the activity patterns are valuable for other learners and peer groups as well. For instance, other facilitators could reuse the learning experiences of the above-mentioned scenario if the actors share them in terms of an activity pattern. Hereby, the pattern-based approach might be useful to avoid the cold-start problem of learning platforms. Third, this best practice sharing method is also advantageous with respect to scrutability and privacy, two important issues for personalization. Activity patterns and their underlying semantic model are simple to understand, and exporting a LISL script easily allows filtering learner-specific lines of code. Here, the learners can regulate the amount of a (successful) activity they want to share. Above and beyond, personalization and (automated) adaptation effects take place within the learning environment (i.e. the learning network) while learners (peers and facilitators) use tools to collaborate on shared artifacts within their common activities.

Despite of the possibilities and strengths of our MUPPLE approach, a few disadvantages have to be outlined here. Primarily, these problems concern technological issues. First of all, it is necessary to have a high degree of interoperability between web applications, which now is not always the case. This specifically relates to single-sign-on procedures and communication channels to transfer both data and events from one application to the other. For example, the WebMail client requires authentication, so, currently, learners have to login separately in each application. Regarding communication channels, [21] propose a specification how to realize distributed feed networks with buffered-push capabilities. We intend to further investigate these means and will gain experiences on how they can be incorporated into LISL. It is planned to introduce additional 'connect' statements for combining tools with the abovementioned feed-based interoperability mechanism. We can think of other approaches, though, and we do not have a solution for the efficient communication of events. Secondly, the utilization of iframes causes problems in cross-domain scripting (cf. [39]). Finally, we are also aware that LISL and MUPPLE still lack important functions, especially in the area of regulating collaboration and privacy, and a comprehensive evaluation.

5 Conclusions and Future Work

In this paper, we stated that personalization and adaptivity is much more complex in real world learning situations, particularly if learners connect to a network of actors, artifacts, and activities. Therefore, we consider traditional approaches for personalized e-learning as not sufficiently to provide personalized learning experiences on the computer. As a consequence, we introduced our idea of 'personalized e-learning 2.0', taking into consideration that learners connect to a socio-technical network and collaborate on shared artifacts and outcomes. Although our prototypical implementation of a Mash-UP Personal Learning Environments (MUPPLE) is work in progress, we described how the learning process is adapted through environment design and collaborative activities in networked communities. As a conclusion, we underline the following three success factors being comprised in our model:

- First, we break up with traditional personalized e-learning models and consider the learning environment not to a pre-condition for, but an outcome of personalized e-learning. Therefore, we build upon learning environment design and our learner interaction scripting language to be able to describe, understand, and reproduce the outcome of learning.
- Second, the pedagogical model behind MUPPLE is very simple, so that learners can understand how personalization works. Additionally, this activity model is a solid basis for learning environment design and further personalization strategies, like the automated analysis of user behavior and network effects as well as the provision of recommendations or advanced regulation facilities.
- Third, the structure and domain-independency of these learning activities address higher-level learning objectives, independently of a subject domain, and enable best practice sharing as well as reflective learning, i.e. aiming at paradigms like learn-how-to-learn instead of learn-by-heart.

In total, we believe that personalized e-learning will proceed from an instructional design and top-down, ex ante modeling to a learner-driven, bottom-up, just-in-time adaptation of learning by considering the principles of Web 2.0. This is meant to be 'personalized e-learning 2.0'. Our future work will address interoperability issues of learning tools, regulation facilities for collaborative activities in learning networks as well as experiences in real-world learning settings, particularly to evaluate the utility of the MUPPLE approach for higher education and lifelong learning. Concerning personalized e-learning, we have to think of further strategies to recommend (advertise) pattern-based best practices to peers and to support learners in their collaborative activities within their learning network.

Acknowledgements

This work has been produced in the context of iCamp, a research and development project financially supported by the European Union under the ICT programme of the 6th Framework Programme (Contract number: 027163).

References

1. Heller, J., Steiner, C., Hockemeyer, C., Albert, D.: Competence-Based Knowledge Structures for Personalised Learning. International Journal on E-Learning 5(1), 75–88 (2006)
2. Sampson, D., Karagiannidis, C.: Kinshuk: Personalised Learning: Educational, Technological and Standardisation Perspective. Interactive Educational Multimedia 4, 24–39 (2002)
3. Park, O., Lee., J.: Adaptive Instructional Systems. In: Jonassen, D.H. (ed.) Handbook of Research on Educational Communications and Technology, pp. 651–684. Lawrence Erlbaum Associates, Mahwah (2004)
4. Shute, V., Towle, B.: Adaptive E-Learning. Educational Psychologist 38(2), 105–114 (2003)
5. Conlan, O.: The Multi-Model, Metadata Driven Approach to Personalised eLearning Services. University of Dublin, Dublin (2005)
6. Mödritscher, F.: Adaptive E-Learning Environments: Theory, Practice, and Experience. VDM, Saarbrücken (2008)
7. Brusilovsky, P.: KnowledgeTree: A Distributed Architecture for Adaptive E-Learning. In: Proceedings of the World Wide Web Conference (WWW), pp. 104–113 (2004)
8. Henze, N., Nejdl, W.: A Logical Characterization of Adaptive Educational Hypermedia. New Review of Hypermedia and Multimedia 10, 77–113 (2004)
9. Specht, M., Burgos, D.: Modeling Adaptive Educational Methods with IMS Learning Design. Journal of Interactive Media in Education (Adaptation and IMS Learning Design), 1–13 (2007)
10. Henze, N., Dolog, P., Nejdl, W.: Reasoning and Ontologies for Personalized E-Learning in the Semantic Web. Educational Technology & Society 7(4), 82–97 (2004)
11. Chen, C.-M., Lee, H.-M., Chen, Y.-H.: Personalized e-learning system using Item Response Theory. Computers & Education 44(3), 237–255 (2005)
12. Graf, S., Lin, T., Kinshuk, J.: The relationship between learning styles and cognitive traits - Getting additional information for improving student modelling. Computers in Human Behavior 24(2), 122–137 (2008)
13. Helic, D.: A Didactics-Aware Approach to Management of Learning Scenarios in ELearning Systems. Graz University of Technology, Graz (2006)
14. Mödritscher, F., García-Barrios, V.M., Gütl, C.: Enhancement of SCORM to support adaptive E-Learning within the Scope of the Research Project AdeLE. In: Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education (E-Learn), pp. 2499–2505 (2004)
15. Phillips, M.: Cognitive Style's influence on Media Preference: Does it matter or do they know? In: Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA), pp. 1023–1028 (2005)
16. Holzinger, A., Kickmeier-Rust, M., Albert, D.: Dynamic Media in Computer Science Education; Content Complexity and Learning Performance: Is Less More? Educational Technology & Society 11(1), 279–290 (2008)

17. Rollett, H., Lux, M., Strohmaier, M., Dösinger, G., Tochtermann, K.: The Web 2.0 way of learning with technologies. International Journal of Learning Technology 3(1), 87–107 (2007)
18. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction (UMUAI) 6(2-3), 87–129 (1996)
19. Gütl, C., Mödritscher, F.: Towards a Generic Adaptive System applicable for Web-based Learning Management Environments. In: Proceedings of the Annual Workshop of the SIG Adaptivity and User Modeling in Interactive Systems (ABIS), pp. 26–31 (2005)
20. Van Harmelen, M.: Personal Learning Environments. In: Proceedings of the International Conference on Advanced Learning Technologies (ICALT), pp. 815–816 (2006)
21. Wild, F., Sigurdarson, S.E.: Distributed Feed Networks for Learning. European Journal for the Informatics Professional (UPGRADE) 9(3) (2008)
22. Santos, O.C., Boticario, J.G.: Meaningful pedagogy via covering the entire life cycle of adaptive eLearning in terms of a pervasive use of educational standards: The aLFanet experience. In: Tomadaki, E., Scott, P. (eds.) EC-TEL 2006. LNCS, vol. 4227, pp. 691–696. Springer, Heidelberg (2006)
23. Mödritscher, F.: e-Learning Theories in Practice: A Comparison of three Methods. Journal of Universal Science and Technology of Learning (JUSTL) 0(0), 3–18 (2006)
24. O'Reilly, T.: What is Web 2.0. O'Reilly Media Inc. (2005) (2008-07-21), http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
25. Ebner, M., Holzinger, A., Maurer, H.: Web 2.0 Technology: Future Interfaces for Technology Enhanced Learning? In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 559–568. Springer, Heidelberg (2007)
26. Lathem, J., Gomadam, K., Sheth, A.: SA-REST and (S)mashups: Adding Semantics to RESTful Services. In: Proceedings of the International Conference on Semantic Computing (ICSC 2007), pp. 469–476 (2007)
27. Jones, S., Maiden, N.A.M.: RESCUE: An Integrated Method for Specifying Requirements for Complex Socio-Technical Systems. In: Mate, J.L., Silva, A. (eds.) Requirements Engineering for Socio-Technical Systems, pp. 245–265. Idea Publishing (2004)
28. Kraus, J.: The long tail of software: Millions of Markets of Dozens. Bnoopy: An entrepreneurship blog (2005) (2008-07-21), http://bnoopy.typepad.com/bnoopy/2005/03/the_long_tail_o.html
29. Mödritscher, F., Neumann, G., García-Barrios, V.M., Wild, F.: A Web Application Mashup Approach for eLearning. In: Proceedings of the OpenACS and.LRN Conference, pp. 105–110 (2008)
30. Spoerri, A.: Visual Mashup of Text and Media Search Results. In: Proceedings of the International Conference Information Visualization, pp. 216–222 (2007)
31. García-Barrios, V.M., Mödritscher, F., Gütl, C.: Personalisation versus Adaptation? A User-centred Model Approach and its Application. In: Proceedings of the International Conference on Knowledge Management (I-KNOW), pp. 120–127 (2005)
32. Akhras, F.N., Self, J.A.: System Intelligence in Constructivist Learning. International Journal of Artificial Intelligence in Education 11(4), 344–376 (2000)
33. Kay, J.: Stereotypes, Student Models and Scrutability. In: Proceedings of the International Conference on Intelligent Tutoring Systems (ITS), pp. 19–30 (2000)
34. Boud, D., Keogh, R., Walker, D. (eds.): Reflection: Turning Experience Into Learning. Routledge, Abingdon (1985)
35. Koper, R., Olivier, B.: Representing the Learning Design of Units of Learning. Educational Technology & Society 7(3), 97–111 (2004)

36. Mödritscher, F., García-Barrios, V.M., Gütl, C.: The Past, the Present and the Future of adaptive E-Learning: An Approach within the Scope of the Research Project AdeLE. In: Proceedings of the International Conference on Interactive Computer Aided Learning, ICL (2004)
37. Dillenbourg, P., Jermann, P.: Designing integrative scripts. In: Fischer, F., Mandl, H., Haake, J., Kollar, I. (eds.) Scripting computer-supported collaborative learning: Cognitive, computational and educational perspectives, pp. 277–302. Springer, New York (2007)
38. Wild, F., Mödritscher, F., Sigurdarson, S.E.: Designing for Change: Mash-Up Personal Learning Environments. eLearning Papers 9 (2008)
39. Jackson, C., Wang, H.J.: Subspace: Secure Cross-Domain Communication for Web Mashups. In: Proceedings of the International Conference on World Wide Web (WWW), pp. 611–620 (2007)

Avatars in Assistive Homes for the Elderly
A User-Friendly Way of Interaction?

Martin M. Morandell[1], Andreas Hochgatterer[1], Sascha Fagel[2], and Siegfried Wassertheurer[1]

[1] Austrian Research Centers GmbH - ARC
Biomedical Engineering / smart Biomedical systems
Viktor Kaplan-Strasse 2/1
A-2700 Wiener Neustadt, Austria
{martin.morandell,andreas.hochgatterer, siegfried.wassertheurer}@arcs.ac.at
[2] Berlin Institute of Technology
Strasse des 17. Juni 135
D-10623 Berlin, Germany
sascha.fagel@tu-berlin.de

Abstract. Designing user interfaces for elderly people has become a prominent research topic. In the context of Assistive Homes the development of user-friendly and alternative interfaces for the improvement of usability and acceptance is an important goal. Applying avatars shall provide notable benefits. This paper describes the implementation and evaluation of a "photo-realistic avatar for people with mild cognitive impairment and dementia". The findings of the user trials show positive effects such as raised user's attention, increased likeability but no effects on memory functions. Also other approaches from Germany and Spain with less anthropomorph-looking avatars are presented and their results discussed.

Keyword: Avatar, Usability, Dementia, Elderly, Assistance.

1 Background and Motivation

The successful application of a new technology depends on many factors; especially for the group of elderly end-users the interaction, its usability and even emotional feelings are very important criteria for it being accepted. These features become even more important as soon as technology starts playing a crucial role concerning safety and independence, e.g. in Ambient Assisted Living (AAL) environments where the acceptance of the applied technology is the crucial factor if the user can derive a benefit from the applyed technology or not. Beside the fact that the benefit of using new devices must be appreciable, in order to provide a motivation for its use [1], there can even be "a need for an emotional relationship" between the users and their Assistive Technology as described by Wu P. 2005 [2].

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 391–402, 2008.

There are many ways of possible interaction methods between the user and a Smart Home environment. Some applications use touch terminals or touch panels [3, 4] which seem to be easy to use especially for elderly users who have never used a computer before [5] and others put focus on the usage of TV sets with (user-friendly) remote controls since TV sets constitute a platform elderly are most familiar with and do use very frequently [6]. In many of these cases the output is symbol and/or text based [4] sometimes with additional recorded text messages [3]. Speech input is still rarely used and evaluated. This paper focuses on the communication towards the end-user (i.e. the output).

Even though feasibility studies on basic technical aspects of Smart Home environments are available from various sources (e.g.: the Gator Tech Smart House [7], the Independent Life Style Assistant [8]), the requirements for new strategies to make Information and Communication Technology (ICT) and in this case Smart Home technology likely to be used by the elderly population is underestimated and not observed sufficently. The key factor to success is the interface to the Smart Home's features and services. This was also stressed in the "Design Guidelines on Smart Homes" [9]:

> The user interface is the single component in such systems, upon which everything else will be judged. If the interface is confusing and badly designed, the system will be thought of in that way. Indeed, to make such systems appear simple is an extremely complex goal to achieve. It is, nonetheless, very important to do so.

Until today computers and their application programs have been operated by using predefined command sets. Although this is tried to be hidden more and more by applying different kinds of user interface technologies (Graphical User Interfaces, Tangible User Interfaces, Ubiquitous Computing) human-computer interaction is still limited since the user has to adapt to the computer and its commands. An inverse way would be to model the human character sets of communication by information technology [10].

In addition the paradigm shift from "the computer used as a tool" to "the computer used as an assistant", the vision of ICT for everyone without any barriers that might cause a "digital divide" and the request to make technology friendly, polite and fun to use [10] have caused a widespread use of avatars in many fields (e.g. internet pages, movies, computer games and kiosk situations).

An avatar can be seen as consisting of "body and mind" (compare to [10]):

- A software agent builds human models like emotional states, models of cognition and knowledge. The software agent can be seen as the mind of the avatar.
- These models are presented to the exterior using visually human models. This visualization can be seen as the body of the avatar.

In the field of Human-Computer Interaction for elderly end-users most user interfaces are still based on the WIMP model (window, icon, menu, pointing device), but in the recent past also embodied conversational agents [6, 11] have

(a) A realistic virtual character rendered on a common television set [6]

(b) A more human like avatar used within the ALADIN project [12]

(c) A cartoon like avatar used within the ALADIN project [12]

(d) A photorealistic talking head used in our approach

Fig. 1. Different Avatars

been applied as a user interface component as an attempt to make the interaction more natural.

In most projects these embodied conversational agents (=avatars) are not very photo-realistic (see fig: 1(a), 1(b)) [6, 11] and sometimes they are even presented in a cartoon-like way (see 1(c)) [12, 13].

1.1 The Avatar - A Multi Modal Output System

Speech and non-verbal gestures form an important part of human communication. Hence, the "digital divide" between a user and a system can be reduced by using an embodied conversational agent with appropriate verbal and non-verbal behavior. While speech input to a machine is often erroneous and limited due to insufficient quality of automatic speech recognizers, audio speech output by speech synthesisers has already achieved a reasonable quality with respect to intelligibility under normal conditions. However, the application of audio speech in systems for elderly end-users cannot be seen as a normal condition: The transmission of verbal information may be degraded by environmental noise, reduced capacity of hearing and reduced ability to concentrate. Additional visual speech and non-verbal output can enhance several aspects of quality of the interface:

1. As also known from natural speech [14] synthetic visual speech increases the intelligibility when added to audio speech [15].
2. Less effort is needed when listening to a talking head compared to pure audio speech [16].
3. The perception of prominence (e.g. to emphasise particular words in a sentence) can be clarified [17].
4. The identification of the intended expression to be transmitted can be enhanced [18].
5. An embodied conversational agent can improve user satisfaction and engagement and enhance the interaction with a system [19].

The face and the voice do not necessarily need to have the same degree of naturalness to achieve best results with respect to intelligibility; an avatar with artificial appearance combined with natural sounding speech can be highly intelligible [20]. However, the impact of the pairing of face and voice with respect to naturalness/artificialness is still in the scientific discussion: in [21] users prefer avatars with face and voice both being natural or both being synthetic whereas in [22] users prefer a more natural voice even with an obviously artificial face.

By all means the agent's behaviour must be coherent in the audio and in the video display to guarantee consistent information: In case of incoherent verbal [18] and non-verbal [17] gestures the information to be transmitted can be severely altered, i.e. the meaning that reaches the user is not the one intended by the system, whereas coherent multi-modal gestures lead to robust output.

1.2 Avatars for Assistive Homes - State of the Art

The idea of using avatars for the user interface of Assistive Homes for elderly (and especially for persons with Alzheimer's disease) was not often addressed by research [23, 24]. The application of avatars for subsets, e.g. for giving advice via the TV has been tested in more projects (e.g. by Krämer 2008 [11], Carrasco et al. 2007 [6] and Berry et al. 2005 [25]).

Two German Approaches. Krämer presents results of two studies where the effects of virtual assistants on elderly people have been investigated [11].

In the first study users (n=87, age 16-73) were asked to program a video recorder. One part of the group (n=71) was assisted by a virtual helper, the other part (n=16) had to use the handbook. Irrespective of age the users assisted by the avatar completed the task with less mistakes than those using the handbook.

In the second study presented in the same paper, a virtual assistant was evaluated that allowed naturally spoken commands to program a TV and VCR. The users (n=65) could select transmissions either by using natural voice or in the conventional way. The output varied: one group got text messages, another group got speech output, a third one got an anthropomorph agent and the fourth group could select one of these modalities. The result was that the acceptance and efficiency of the different styles did not vary much, but the test subjects using the avatar-like output were more likely to make naturally speaking input than the other groups.

Spanish Approach. In Spain avatar interfaces have been evaluated with the target group of mild cognitive impaired persons and persons with Alzheimer's disease [6, 26]. TV sets were used as front end in this approach and one of the implemented avatars could express some emotions. Besides lip movements also eye motion and blinks, head motion etc. were applied to support the life-like character of the avatar. Voice output was based on a text-to-speech synthesiser. As far as known these trials have only been carried out as laboratory trials. Assistive Homes with avatar interfaces implemented are still not in sight.

1.3 Expected Advantages Using Avatars in Ambient Assisted Living

There have been many studies related to possible advantages of the use of avatars instead of other (conventional) interface approaches. The expected benefits (some have been validated by the authors) are [27]:

- Social interaction: interactions between a human being and a machine are fundamentally social; interaction in a natural way (like with other human beings)
- User attention: animated characters are capable of attracting user's attention
- Naturalness: illusion of liveliness and interacting with a real person is generated
- More information in the transmitted message: facial expressions contain a lot of information
- Trustworthiness and believability: level of trustworthiness is increased by the personification of the agent; a realistic face is rated more intelligent, engaging and likeable

A further reason why the approach of using avatars is expected to be suitable especially for the elderly users and for laypersons is based on the fact that the user does not have to learn something new or adapt to new ways of communication [11].

2 Methods

In the following an approach and ongoing research performed by the Austrian Research Centers in cooperation with the Berlin Institute of Technology is described. Photo-realistic avatars (see Fig.2) are used as part of the Graphical User Interface for Assistive Homes for people with dementia.

Within a master's thesis [23] a prototype for a day structuring tool for people with dementia was developed using a photo-realistic avatar as output component. The idea behind using an avatar in the user interface for people with Alzheimer's disease was to bind the user emotionally to the system. This was tried to be stressed by making the avatar look like familiar and known informal or formal caregivers [24].

2.1 Technical Implementation

The avatars were created by the use of MASSY [28], a tool developed by Sascha Fagel at the Berlin Institute of Technology. MASSY generates videos of an articulating face from a photo of a real person. The synthetic articulation is realised by deforming a mesh based on MPEG FAPs defined on the face. A generic articulation model for German assures that the visible movements match the spoken phones. Avatars created with MASSY currently generate only verbal movements, i.e. the facial expression does not change and only the mouth region is animated. Animating a specific face to phonetically match a synthetic voice takes various advantages of audiovisual perception: while verbal information that is mainly transmitted by audio is supported by mouth movements, the face reflects many properties of the person's identity [29] such as expression [18] and age [30]. The technology to generate audiovisual speech from a photo makes it possible to easily and rapidly create avatars of specific persons which is necessary in a caregiver application for users with Alzheimer's disease.

These avatars were used as GUI component for the day structuring tool described in [31].

2.2 Trials Performed

The trials consisted of a pre-study and a usability study. Both were performed in a day care center for people with dementia as well as in a nursing home for elderly people, both located in Linz, Austria.

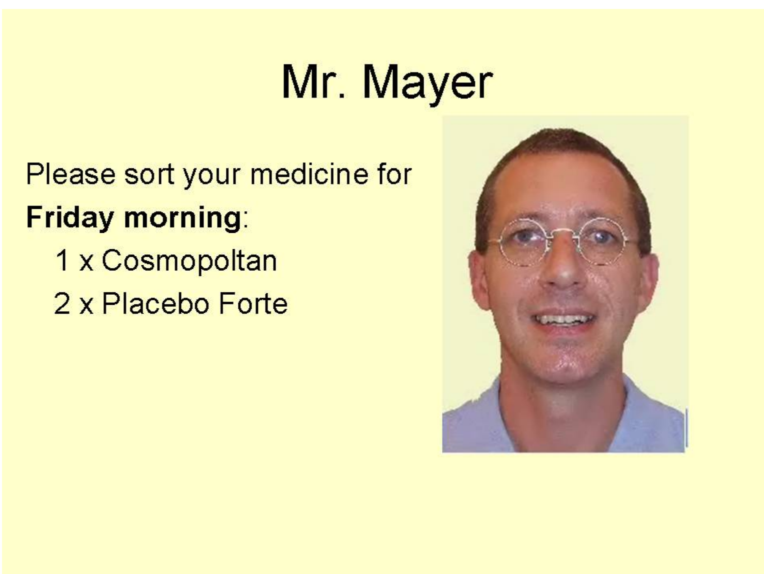


Fig. 2. Screenshot of the Avatar GUI showing a formal caregiver

In total 10 subjects accepted to participate in the project. Not all tests could be performed with all subjects due to illness, absence or giving up the participation in the project. The group (n=10) consisted of 6 female and 4 male subjects, 8 of them with a diagnosis of dementia (Mini Mental Rating from 15 to 23, average 20.75) and for two subjects dementia was not diagnosed but supposed. The average age when the tests were performed was 82. At the day care center tests and interviews were carried out in a separate room. In the nursing home, subjects were visited in their own living environment.

The pre-study consisted of three meetings with the following aims: Meeting 1: interview and memory game. Aim 1: To get to know the subjects and to get a feeling for their daily life situations. Meeting 2: Sorting Easter eggs into different nests following instructions given on a screen by text, graphics and human and synthetic speech. Aim 2: To test if information presented via a GUI with synthetic speech can be understood and processed by the users. Meeting 3: Selecting the preferred photo out of a set of familiar and unfamiliar persons. Aim 3: To identify the preferred head to create an avatar from.

Based on the outcomes of the pre-study, the following tests were performed using the talking heads created based on the photos selected in the third meeting.

Test 1: The aim of the first test was to find out, if talking heads were liked as advisors for certain tasks. The first task was to sort medicine. It was performed with 7 users. It was divided into 4 subtasks, two using text and an avatar and two subtasks based on text and speech only.

Test 2: The second test consisted of putting forms of different colors and shapes into the right order, following the instructions given by the GUI with or without the avatar. The aim was to find out if the success of task guidance differs when using avatars or not.

Test 3: Observations during the second test concerning how users reacted to new scenes built the basis for the third test. Subjects (n=5) were reading a text and were interrupted by the system, either just by voice and text or combined with a talking head.

The tests were performed using a computer screen mounted into a wooden frame with loudspeakers. The tests were done in a "Wizard of Oz" way.

3 Results

The pre-study showed that synthetic speech (in this case MBROLA [32]) can be applied for speech output and is understood by elderly users. Different styles of information presentation should be provided, combinable and adjustable to the preferences and needs of each single user. In the photo selection test of the pre-study (where the photos for the avatar were selected) it turned out, that only photos of familiar persons are preferred to be used as talking heads. Photos of relatives and known formal caregivers seem to be the preferred ones of the subjects.

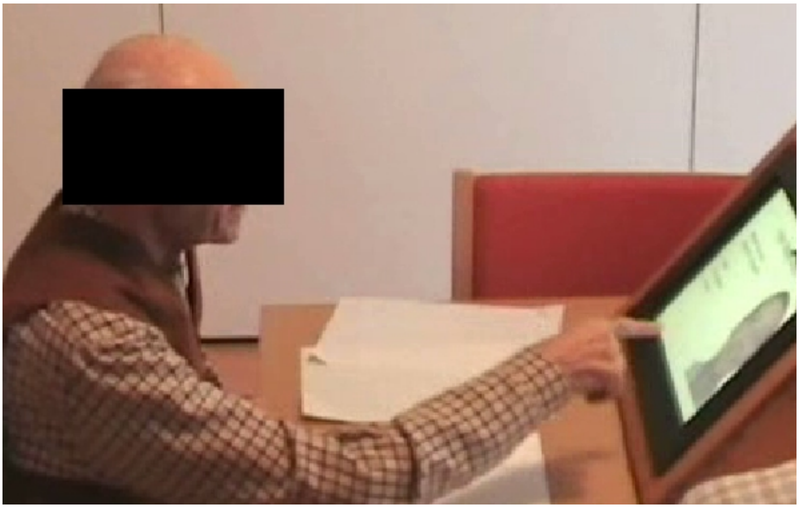


Fig. 3. The figure shows a test candidate talking to the avatar

The usability tests revealed the following results:

Test 1: In this test the avatar requested the subject (n=7) to sort medicine into different boxes. The test showed that for four users the visual impression of the talking head influenced the acoustic perception, leading to the impression that even though synthetic voice was used it was perceived as the original voice of the person the talking head was representing. This was even the case with two subjects who were confronted with avatars of their own children. Especially if voice and face were felt as matching, the avatar based user interface was preferred.

Test 2: In the second test most subjects completed the task of sorting shapes equally well or badly, independent of the presence of an avatar on the screen. But for the subject with the most progressed state of dementia, it was almost impossible to perform the task without any help when no avatar was shown, whereas with an avatar on the screen the task was performed with success. It was observed that the one subject using an avatar of a relative was more connected to the voice in the first scene when the avatar was shown. In the second scene without the avatar the attention was put more on the text even though voice was played.

Test 3: In the third test it was observed how the participants (n=5) reacted to new scenes (with a known avatar, an unknown avatar and without any avatar) while reading a text. The longest attention was raised by the known avatar. The unknown avatar was not realized as fast as the known one, but in general even after having been focused, the user's attention remained with the screen. Without the avatar the new scene on the screen did not receive as much attention.

Within the test series also other interesting reactions were shown [23]. For one of the subject, where a talking head created by the photo of the daughter was used, the emotional binding toward the system seemed to be rather strong. When finishing one of the tests she passed a kiss by the finger onto the photo of her daughter, which was still displayed on the screen. Also in other cases, where well liked nurses were used, the subjects were pleased about the known face talking to them. One of the subjects started arguing with the avatar (see Fig. 3) when misunderstanding a task message. This pointed out that using photo-realistic avatars can also lead to interpersonal reactions - with positive and negative effects.

4 Discussion and Outlook

As overall interpretation of the results it can be stated that photo realistic avatars can be a method to make user interfaces more accepted by the target group of people with dementia. The effect can be enhanced when familiar faces are used. Furthermore, avatars can be used as eye-catchers in new scenes, in particular when using speech the origin of speech can be located more easily and the attention can be bound to the system. Due to the small number of participants the performed tests do not allow any predications about if the usage of avatars leads to better results in fulfilling a given task.

As human beings have learned to recognize emotions through facial expressions and gesticulation, a high level of realism of virtual characters is needed to achieve acceptance by the human opposite. This does not only include the photo realistic representation, but also facial expression and gesticulation [33].

There might be a limit of how realistic an avatar should become which is discussed within the "Uncanny Valley" phenomenon [34]. Future research could address if this phenomenon is also valid within the target group of mild cognitive impaired (elderly) persons. This question relates to the findings of Morandell [23] suggesting that synthetic speech was perceived as real speech when combined with a photo-realistic avatar of a known person.

In general the findings underline the results presented by the Spanish research activities [27] which show that the presence of an avatar does neither have any positive nor negative effect on recall. These results should be validated by an improved study design, enlarged samples and focusing on single aspects of memory functions. Major benefits could be found in the field of attracting attention and joy of use.

Krämer [11] states that elderly end-users initially face new technological possibilities more sceptically than younger users do. Hence, when initiating the use of new technologies for elderly end-users, in the beginning the end-users should find known interaction patterns and then be familiarized step by step with the new technology (e.g. by tutorials). Furthermore the concrete advantages (whether in a physical, medical or emotional respect [35])of the new technology have to be communicated clearly to avoid that it (in this case the avatar) is just seen as entertainment.

How avatars are perceived by persons with moderate Alzheimer's disease is still an unanswered research question.

The future work of the authors will concentrate on the investigation of differences between elderly people with and without cognitive impairments, the implementation and effects of emotion expression and on the identification of the optimum antropomorphism of avatars. A further interesting research topic might the acceptance and likability of avatars based user interfaces concerning cultural differences. Therefore especially those cultural dimensions are of interest that are connected to communication, information, interaction and dialog design [36].

Acknowledgements

The authors thank the subjects and staff of the Regenbogen Day Care Center and the Seniorenzentrum Dornach for their support and participation.

References

[1] Holzinger, A., Schaupp, K., Eder-Halbedl, W.: An investigation on acceptance of ubiquitous devices for the elderly in a geriatric hospital environment: Using the example of person tracking. In: [37], pp. 22–29

[2] Wu, P., Miller, C.: Results from a Field Study: The Need for an Emotional Relationship between the Elderly and their Assistive Technologies. In: 1st International Conference on Augmented Cognition, Las Vegas (2005)

[3] Fugger, E., Ehrenstrasser, L., Hochgatterer, A., Russ, G.: Proactive assistant for intelligent homecare environments - a development pilot. In: Augusto, J.C., Nugent, C.D. (eds.) Smart Homes and Beyond, Proceedings of the ICOST 2006 (June 2006)

[4] Hochgatterer, A., Prazak, B., Russ, G., Sallans, B.: Behaviour Pattern Based Safety Assistant for the Elderly. In: Eizmendi, G., Azkoitia, J.M., Craddock, G. (eds.) Challenges for Assistive Technology AAATE 2007. Assistive Technology Research, vol. 20, pp. 170–174. IOS Press, Amsterdam (2007)

[5] Holzinger, A.: User-centered interface design for disabled and elderly people: First experiences with designing a patient communication system (pacosy). In: ICCHP 2002: Proceedings of the 8th International Conference on Computers Helping People with Special Needs, London, UK, pp. 33–40. Springer, Heidelberg (2002)

[6] Carrasco, E., Góllner, C.M., Ortiz, A., García, I., Buiza, C., Urdaneta, E., Etxaniz, A., González, M.F., Laskibar, I.: Enhanced TV for the Promotion of Active Ageing. In: Eizmendi, G., Azkoitia, J.M., Craddock, G. (eds.) Challenges for Assistive Technology AAATE 2007. Assistive Technology Research, vol. 20, pp. 159–163. IOS Press, Amsterdam (2007)

[7] Helal, S., Mann, W.C., El-Zabadani, H., King, J., Kaddoura, Y., Jansen, E.: The Gator Tech Smart House: A Programmable Pervasive Space. IEEE Computer 38(3), 50–60 (2005)

[8] Haigh, K.Z., Kiff, L.M., Myers, J., Krichbaum, K.: The Independent LifeStyle Assistant^{TM} (I.L.S.A.): Deployment Lessons Learned. In: AAAI WS on Fielding Applications of AI, San Jose, CA (July 2004)

[9] van Berlo, A., et al.: Design guidelines on smart homes. A COST 219bis Guidebook (October 1999)

[10] Spierling, U.: Der Avatar: Ein Wesen, eine Spielfigur, ein Medium, oder ein UI-Element? In: Umhegt oder abhängig?, pp. 207–220. Springer, Berlin (2006)
[11] Krämer, N.: Freundliche Hilfen für abschreckende Künstlichkeit? Virtuelle Agenten für Senioren. In: Maier, E., Roux, P. (eds.) Seniorengerechte Schnittstellen zur Technik - Zusammenfassung der Beiträge zum Usability Day VI, May 16, 2008, pp. 108–112 (2008)
[12] von Hellberg, P., Maier, E., Kempter, G.: Barrierefreie Informationsplattform für Fernsehgeräte. In: Maier, E., Roux, P. (eds.) Seniorengerechte Schnittstellen zur Technik - Zusammenfassung der Beiträge zum Usability Day VI, May 16, 2008, pp. 96–103 (2008)
[13] Alm, N., Arnott, J.L., Dobinson, L., Massie, P., Hewineso, I.: Cognitive prostheses for elderly people. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 806–810 (2001)
[14] Sumby, W.H., Pollack, I.: Visual contribution to speech intelligibility in noise. Journal of Acoustical Society of America 26(2), 212–215 (1956)
[15] Benoît, C.: On the Production and the Perception of Audio-Visual Speech by Man and Machine. In: Multimedia and Video Coding. Plenum Press, New York (1996)
[16] Fagel, S., Madany, K.: Computeranimierte Sprechbewegungen in realen Anwendungen. Universitätsverlag der Technischen Universität Berlin (2008)
[17] Krahmer, E., Swerts, M.: Hearing and seeing beats: the influence of visual beats on the productin and perception of prominence. In: International Conference on Speech Prosody, Dresden (2006)
[18] Fagel, S.: Emotional McGurk Effect. In: Proceedings of the International Conference on Speech Prosody, Dresden (2006)
[19] Foster, M.E.: Enhancing human-computer interaction with embodied conversational agents. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 828–837. Springer, Heidelberg (2007)
[20] Fagel, S., Clemens, C.: An articulation model for audiovisual speech synthesis - determination, adjustment, evaluation. Speech Communication 44(1-4), 141–154 (2004)
[21] Nass, C., Gong, L.: Maximized modality or constrained consistency? In: Proceedings of AVSP 1999, Santa Cruz, Calif, pp. 1–5 (1999)
[22] Kühnel, C., Weiss, B., Wechsung, I., Fagel, S., Möller, S.: Evaluating talking heads for smart home systems. In: International Conference on Multimodal Interfaces, Chania (accepted, 2008)
[23] Morandell, M.: Day Structuring Assistance for People with Alzheimer's Disease. Master Thesis at the University of Linz, Austria, inaccessible until October 2010 (2007)
[24] Morandell, M., Fugger, E., Prazak, B.: The Alzheimer Avatar - Caregivers' Faces Used as GUI Component. In: Eizmendi, G., Azkoitia, J.M., Craddock, G. (eds.) Challenges for Assistive Technology AAATE 2007. Assistive Technology Research, vol. 20, pp. 180–184. IOS Press, Amsterdam (2007)
[25] Berry, D.C., Butler, L.T., de Rosis, F.: Evaluating a realistic agent in an advice-giving task. International Journal of Human-Computer Studies 63(3), 304–327 (2005)
[26] Carrasco, E., Epelde, G., Moreno, A., Ortiz, A., Garcia, I., Arruti, A., Buiza, C., Urdaneta, E., Etxaniz, A., González, M.F.: Natural interaction between virtual characters and persons with alzheimer's disease. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 38–45. Springer, Heidelberg (2008)

[27] Ortiz, A., del Puy Carretero, M., Oyarzun, D., Yanguas, J.J., Buiza, C., González, M.F., Etxeberria, I.: Elderly users in ambient intelligence: Does an avatar improve the interaction? In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 99–114. Springer, Heidelberg (2007)
[28] Fagel, S.: Merging methods of speech visualization. ZAS Papers in Linguistics 40, 19–32 (2005)
[29] Joassin, F., Mauragea, P., Bruyera, R., Crommelinckb, M., Campanellaa, S.: When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. Neuroscience Letters 369, 132–137 (2004)
[30] Fagel, S.: Auditory-visual integration in the perception of age in speech. In: Proceedings of the ICPhS, Saarbrücken (2007)
[31] Morandell, M., Fuxreiter, T., Hochgatterer, A.: An OSGi based Framework for a Day Structuring Tool in an Assistive Home Environment. In: Eizmendi, G., Azkoitia, J.M., Craddock, G. (eds.) Challenges for Assistive Technology AAATE 2007. Assistive Technology Research, vol. 20, pp. 180–184. IOS Press, Amsterdam (2007)
[32] Mbrola Team: Mbrola Project Homepage [180708], http://tcts.fpms.ac.be/synthesis/mbrola.html
[33] Nett, N., Rüpel, R.: Virtuelle charaktere [180708], http://www.iuk.fraunhofer.de/
[34] Mori, M.: The Uncanny Valley (translated by Karl F. MacDorman and Takashi Minato). In: CogSci-2005 Workshop Towards Social Machansims [250807] (2005), http://androidscience.com/theuncannyvalley/procedding2005/uncannyvalley
[35] Holzinger, A., Searle, G., Kleinberger, T., Seffah, A., Javahery, H.: Investigating usability metrics for the design and development of applications for the elderly. In: [37], pp. 98–105
[36] Heimgärtner, R., Holzinger, A., Adams, R.: From cultural to individual adaptive end-user interfaces: Helping people with special needs. In: [37], pp. 82–89
[37] Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.): ICCHP 2008, Linz, Austria, July 9-11, 2008. LNCS, vol. 5105. Springer, Heidelberg (2008)

Using Clustering Technique for Students' Grouping in Intelligent E-Learning Systems

Danuta Zakrzewska

Institute of Computer Science Technical University of Lodz,Wolczanska 215, 90-924 Lodz, Poland
dzakrz@ics.p.lodz.pl

Abstract. In the paper, application of cluster analysis for students' grouping in intelligent e-learning systems is considered. It is proposed the system architecture, in which teaching paths as well as proper layouts are adjusted to groups of students according to their learning styles and usability preferences. Considered student models are built on the basis of Felder and Silverman model, together with student color choices. It is considered usage of two versions of two-phase hierarchical clustering algorithm for students' grouping. Experimental results for both of the approaches are compared and discussed.

Keywords: cluster analysis, learning styles, color preferences.

1 Introduction

In distance learning, educational software, besides fulfilling pedagogical requirements, should also meet individual learners' preferences, as students' attitude towards using computer program may depend significantly on their acceptance of technology, what in turn, usually depends on the degree that the software and its interface are tailored to comply with users' needs. The contents and organisation of courses as well as the way the information is presented may play significant role in educational software performance, which depends on system's adaptivity and personalisation abilities. As it was stated in [1], students' preferences and needs depend on their personal learning styles.

In the paper, it is considered usage of cluster analysis for students' grouping according to their individual learning styles dimensions and usability preferences at the same time. Application of unsupervised classification allows to build teaching paths and tailor envoironments into requirements of members of dynamically created groups. Connection of two factors: learning styles and usability preferences enables not only to obtain learning outcomes but also build environments that are acceptable for users. It is proposed, application of the two-phase hierarchical algorithm, which, in each of the phases, allows to take into account one of the factors and, what is more, allows tutors to choose parameters corresponding to the courses' characteristic features.

The paper is organised as follows. The related work is depicted in the next section. Then, intelligent e-learning system is proposed, learners' models based

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 403–410, 2008.

on learning styles dimensions and usability preferences - taking color choices as the example, are presented and the two-phase hierarchical clustering algorithm for users' grouping is described. In the next section, some experimental results as well as choice of clustering parameters are discussed. Finally, concluding remarks concerning evaluation of the applied algorithm as well as future research are presented.

2 Related Work

Several authors considered personal learning styles as factors that should be taken into account during adoptation of teaching process into individual needs, as those that may influence learners' individual attitudes towards using the educational software (see [2],[3]). Beaudoin & Dwyer [4] stated that a necessary element for online course efficacy is establishing a collaborative learning environment depending on students' profiling. Graf and Kinshuk [5] showed that students with different learning styles have different needs and preferences. While building an intelligent learning environment, there should be taken into account interface customization to learners' preferences as it was described in [6], where Felder and Silverman model of learning styles was considered. An overview of the research concerning building adaptive systems by taking into consideration learning styles can be found in [7] and [8]. The most popular method, consists in assigning students into predefined groups, without possibility of updating (compare [9],[10],[11]).

In [12], it was presented the architecture of intelligent e-learning system, in which students are grouped according to their learning styles by cluster analysis. As clustering technique, it has been chosen Farthest First Traversal algorithm, applied for data previously cleaned out from outliers by COF method. In that approach tutors had to determine the required number of learners' groups in advance. In the next paper [13], it was considered application of two-phase hierarchical algorithm, which deals with outliers and allows to determine weights for different learning styles dimensions, and with number of clusters determined depending on a required threshold. In the current approach the same algorithm is adopted for grouping according to two factors: learning styles and color preferences.

Test results, presented in [14], showed that students with different dominant learning styles dimensions indicated different features, as important, from the usability point of view. Color preferences are strictly connected with usability needs and are important factors in interface designing. Appropriate usage of colors can make interfaces easier to understand and use [15]. On the other hand colors used in interfaces, usually do not address users' needs and only few authors consider user oriented color models [16]. In the paper [14], there were examined student models based on both factors: dominant learning styles and preferable colors. Investigations, presented there, indicated the possibility of using cluster analysis for building a certain number of users' groups (clusters) with similar requirements.

3 Building Intelligent E-Learning System

3.1 System Architecture

In the paper [12], it was proposed the architecture of e-learning system that enables to adjust learning paths into individual needs of students, according to their preferred learning styles. In the current study, it is proposed to develop the system by including usability preferences as additional attributes used in cluster analysis. The data set of the sample group of students is the basis for forming student clusters by unsupervised classification. Now, teaching materials and layouts may be accommodated to needs of every group and different learning paths together with content organisation may be created. Each new student, after filling the questionnaires is assigned to the proper group. Students who do not fit to any clusters should be considered individually. The overview of the simplified system architecture (without outliers) is presented on Fig. 1. The proposed personalized approach allows to build individual paths and layouts. However, the content is static for each student during the course, it may be changed while starting the new one.

3.2 Students' Models

In the present study, we will focus on color choices as usability preferences and there will be considered student model SM represented by two elements: learning styles and color preferences.

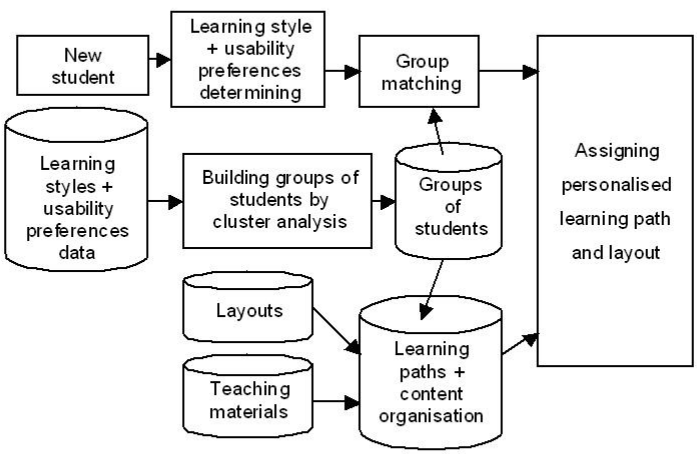


Fig. 1. System architecture

$$SM = (SL, SC) \ , \tag{1}$$

where

$$SL = (sl_1, sl_2, sl_3, sl_4) = (l_{\mathrm{ar}}, l_{\mathrm{si}}, l_{\mathrm{vv}}, l_{\mathrm{sg}}) \ , \tag{2}$$

$$SC = (sc_1, sc_2) \ . \tag{3}$$

SL represents Felder and Silverman model [1],which is often used for providing adaptivity regarding learning styles in e-learning environments [17]. It is based on *Index of Learning Style* (ILS) questionnaire, developed by Felder & Soloman [18]. The results of ILS questionnaire indicate preferences for four dimensions of the Felder & Silverman model: *active* vs *reflective*, *sensing* vs *intuitive*, *visual* vs *verbal*, *sequential* vs *global*. The index obtained by each student has the form of the odd integer from the interval [-11,11], assigned for all of the four dimensions, and is represented as components of SL: l_{ar} means scoring for *active* (if it has negative value) or *reflective* (if it is positive) learning style, and respectively $l_{\mathrm{si}}, l_{\mathrm{vv}}, l_{\mathrm{sg}}$ are points for all the other dimensions, with negative values in cases of *sensing*, *visual* or *sequential* learning styles, and positive values in cases of *intuitive*, *verbal* or *global* learning styles. SC represents numbers of the first (sc_1) and the second (sc_2) choices of preferable colors of each student. Both of the attributes are of categorical type. To obtain the data, students should be asked to choose two the most preferable colors' compositions, from among all that may be used in the system layout, and to indicate the rank for each.

3.3 Two-Phase Hierarchical Clustering Algorithm

The problem statement, which consists of finding groups of students with similar learning styles and usability preferences by unsupervised classification, imposes requirements for characteristic features of a clustering algorithm. In the paper [13], it was presented application of two-phase hierarchical algorithm, that allows to find student clusters according to dominant learning styles and that fulfill tutors' requirements. Technique, proposed there, enables to take into account attributes of different types and it may be used for data represented by SM.

Considered algorithm consists of two stages: the first one is a single layer clustering, during the second phase clusters are formed into hierarchical structure. All the components of vector SM may be used in both phases of the algorithm, but we propose to consider also the second version, in which, elements SL and SC are separated. In that case single layer clustering is applied for components of SL, and then output clusters are merged according to SC components. That version of proposed algorithm takes the following form:

[Input]: A set of N students' SL and SC data, clustering threshold T, maximal number of clusters $KMAX$, minimal clustering threshold $MINT$.
Phase I: Single Layer Clustering
Step 1: Assign the first student SL_1 as the first cluster and its centroid,
Repeat Step 2 and Step 3 for each student $SL_i, i = 2, ...N$:
Step 2: Calculate the similarity by using (4) and (5), between student's SL and the centroid of each existing cluster,

Step 3: Assign the student into the closest cluster, for which the similarity is greater than T, otherwise SL_i initiates a new cluster
Step 4: If there exist clusters containing one element, repeat Step 2 and Step 3 until all clusters stabilize,
[**Output**]**:** Set of clusters *SCM*
Phase II: Hierarchical Agglomerative Clustering
Step 1: Choose all the clusters containing one element and remove them from *SCM*, let they consist the set *SCM1*
Step 2: If number of clusters in *SCM* is less or equal to *KMAX* then go to Step 4,
Step 3: Repeat: Merge the closest clusters according to centroids similarity, by using (4) and (6), until number of clusters *KMAX* is achieved
Step 4: For each cluster from *SCM1*, find the closest cluster from *SCM*, if similarity of their centroids, calculated by (4) and (6), is greater than *MINT* then merge both clusters otherwise indicate clusters containing one element as an outlier.
[**Output**]**:** The set of clusters *SCM* and the set of outliers.

Clustering threshold T depends on considered courses and is connected with sufficiently similar learning styles, while minimal clustering threshold *MINT* concerns color preferences. Maximal number of clusters *KMAX* is connected with number of teaching paths that tutors are able to prepare for the educational process. The first phase and the proper choice of T should guarantee sufficient similarity of students' learning styles in each cluster. The second stage together with suitable *MINT* value should fullfil the requirements of the same color choices among the most possible number of students with similar dominant learning styles preferences. In the algorithm, there is applied similarity function, introduced by Gower [19], that may be used for both types of data: numerical and categorical, determined as follows:

Definition 1. *Let SM_i, SM_j denote two d-dimensional objects. Then the general similarity coefficient sim_{g} is defined as:*

$$sim_{\mathrm{g}}(SL_i, SL_j) = \frac{1}{\sum_{k=1}^{d} w_k} \sum_{k=1}^{d} w_k s_k \quad , \tag{4}$$

where s_k is the similarity of the k-th attribute of SM_i, SM_j and w_k is either one or zero depending on wether or not a comparison is valid for the k-th attribute of the two objects. For numerical attributes s_k is defined as

$$s_k = 1 - \frac{\mid sl_{i_k} - sl_{j_k} \mid}{R_k} \quad , \tag{5}$$

where R_k is the range of k-th attribute. For categorical values s_k has the form:

$$s_k = \begin{cases} 1 & if\ both\ attributes\ have\ the\ same\ value, \\ 0 & otherwise. \end{cases} \tag{6}$$

Similarity measure defined by (4) ensures good data normalization and independence of the algorithm of input data order [20]. Using w_k enables to assign weights for attributes, the ones not valid in the classification process should be equal to 0.

4 Experiment Results

The experiments were done for two data sets: the one of Computer Science students' data and the other artificially generated; their goal was to compare the performance of two versions of the two-phase approach. In the first version all the attributes were used in both phases, while in the second one, they were separated, as it was described in the previous section. To create the first data set, the sample group of 71 students, was chosen from among those, who took part in collaboration online and used educational environment based on Moodle (http://www.moodle.org) . Students examined color compositions of the portal layout, where their activities took place. They were asked to choose 2 from 9 colors as preferable and indicate ranks for them. They filled also ILS questionnaire. The second data, of 73 instances, were generated artificially. The performance of the proposed method was examined by evaluation of the quality of obtained clusters, taking into account two points of view: color choices and learning styles. In the first case, it is measured by number of instances with different color preferences assigned into the same clusters, while in the second one the quality is measured by differences among dominant learning styles dimensions inside clusters. Research was done for different values of $KMAX$ as well as depending on inclusion of color rank into weight coefficients w_k. Thresholds T and $MINT$ were experimentally chosen and constant during all experimets: T equal to 0.92 in the first version and 0.96 in the second one, for the first data set, and respectively 0.78 and 0.93 for artificially generated data; $MINT$ was equal to 0.75 in all the cases.

Table 1. Percentage of cluster objects with different color preferences

Algorithm	Rank	KMAX	Real data	Artificial data
VER.1	No	3	25.35%	30.14%
		4	12.68%	30.14%
		5	14.08%	24.66%
	Yes	3	28.17%	31.51%
		4	19.72%	30.14%
		5	11.27%	26.03%
VER.2	No	3	28.17%	36.99%
		4	26.76%	34.25%
		5	40.84%	34.25%
	Yes	3	28.17%	32.88%
		4	28.17%	32.88%
		5	40.84%	32.88%

Table 1 presents percentage of students, whose color pereferences differ from the majority of the cluster, for two versions of the algorithm, various numbers of clusters, taking or not taking into account color ranks. Result analysis shows that the best quality of clusters, from the point of view of color preferences, was obtained by the first version of the algorithm, what means using all the attributes during the whole clustering process; and for the biggest maximal number of clusters equal to 5, in case of usage of color ranks, but distinguishing the first color choice may have bad influence on cluster quality connected with the other attributes. Comparison of percentage values for both considered versions, and both data sets, shows that using of color preferences attributes in the first phase of the algorithm definitely improve the quality of obtained clusters.

Quality of clusters, from the point of view of learning style dimensions, presents completely disparate image. Detailed analysis of obtained groups, for both of the data sets, shows that, in some of them, asigned objects are characterized by learning style dimensions of significantly different score values. For example, in case of the first version of the algorithm with $KMAX = 5$ and including ranks, in one of the cluster of 13 students, the strongly active ones together with strongly reflective as well as sequential together with global were placed. It may mean that the second version gives better results from the point of view of learning styles, while the first one, if focus is done on color preferences, however, to obtain general conclusions further investigations are necessary.

5 Concluding Remarks

In the paper, it was considered usage of cluster analysis for student grouping according to their dominant learning styles and color preferences. It was proposed intelligent e-learning system, in which teaching materials as well as content organisation are assigned to student groups created by unsupervised classification. The research focused on investigations of two versions of the two-phase hierarchical clustering algorithm. Experiments done for real and artificial data allowed to indicate parameters, which guarantee the best results, but some improvements of the algorithm and further research concerning choice of parameters may be necessary. Investigations showed that different versions of the algorithm should be chosen depending on which part of students' model the emphasis is done, however further research concerning the role of weights may be also useful.

Proposed student model, depends on color preferences as usability needs. In the next step development of models by taking into account other preferences, like navigational paths may be considered. The future research should also consist of further investigations of the examined algorithm and its possible modifications, with the emphasis on the choice of parameters, application of different similarity functions, for different data sets, as well as of comparing its effectiveness with other algorithms. Obtaining students' clusters of good quality is the first stage. Then there should be investigated defining representative features for each group, what will enable tutors to personalize the system. However, more research needs to be done also in that area.

References

1. Felder, R.M., Silverman, L.K.: Learning and Teaching Styles in Engineering Education. Eng. Educ. 78, 674–681 (1988)
2. Lee, M.: Profiling Students Adaptation Styles in Web-Based Learning. Comput. Educ. 36, 121–132 (2001)
3. Lu, J., Yu, C.S., Liu, C.: Learning Style, Learning Patterns and Learning Performance in a WebCT-Based MIS Course. Inform. Manage. 40, 497–507 (2003)
4. Beaudoin, M.F., Dwyer, M.: Learning or Lurking? Tracking the Invisible. Online Student. Internet & Higher Educ. 5, 147–155 (2002)
5. Graf, S., Kinshuk: Considering Learning Styles in Learning Managements Systems: Investigating the Behavior of Students in an Online Course. In: 1st IEEE Int. Workshop on Semantic Media Adaptation and Personalization, Athens (2006)
6. Cha, H.J., Kim, Y.S., Park, S.H., Yoon, T.B., Jung, Y.M., Lee, J.-H.: Learning Styles Diagnosis Based on User Interface Behaviors for Customization of Learning Interfaces in an Intelligent Tutoring System. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 513–524. Springer, Heidelberg (2006)
7. Stash, N., Cristea, A., De Bra, P.: Authoring of Learning Styles in Adaptive Hypermedia: Problems and Solutions. In: WWW Conf., pp. 114–123. New York (2004)
8. Brusilovsky, P., Peylo, C.: Adaptive and Intelligent Web-Based Educational Systems. International J. of Artificial Intelligence in Education. 13, 156–169 (2003)
9. Gilbert, J.E., Han, C.Y.: Adapting Instruction in Search of a Significant Difference. J. Netw. Comput. Appl. 22, 149–160 (1999)
10. Saarikoski, L., Salojärvi, S., Del Corso, D., Ovein, E.: The 3DE: an Environment for the Development of Learner-Oriented Customized Educational Packages. In: ITHET 2001, Kumamoto (2001)
11. Triantafillou, E., Pomportsis, A., Georgiadou, E.: AES-CS: Adaptive Educational System Base on Cognitive Styles. In: AH Workshop, Malaga, pp. 10–20 (2002)
12. Zakrzewska, D.: Cluster Analysis for Building Personalized E-learning System. Pol. J. Environ. Stud. 16, 330–334 (2007)
13. Zakrzewska, D.: Cluster Analysis for Users Modeling in Intelligent E-learning Systems. In: Nguyen, N.T., Borzemski, L., Grzech, A., Ali, M. (eds.) IEA/AIE 2008. LNCS (LNAI), vol. 5027, pp. 209–214. Springer, Heidelberg (2008)
14. Zakrzewska, D., Wojciechowski, A.: Identifying Students Usability Needs in Collaborative Learning Environments. In: 2008 Conference on Human System Interaction, Krakow, pp. 862–867 (2008)
15. Marcus, A.: Designing Graphical Interfaces. Unix World, October (1990)
16. Bauersfeld, P.F., Slater, J.L.: User-Oriented Color Interface Design: Direct Manipulation of Color in Context. In: SIGCHI Conference on Human Factors in Computing Systems: Reaching through Technology, New Orleans, pp. 417–418 (2001)
17. Viola, S.R., Graf, S., Kinshuk, L.T.: Investigating Relationships within the Index of Learning Styles: a Data Driven Approach. Interactive Technology & Smart Education 4, 7–18 (2007)
18. ILS Questionnaire, http://www.engr.ncsu.edu/learningstyles/ilsweb.html
19. Gower, J.: A General Coefficient of Similarity and Some of Its Properties. Biometrics 27, 857–874 (1971)
20. Wang, L.: On Competitive Learning. IEEE T. Neural Networ. 8, 1214–1217 (1997)

Adaptation Criteria for Preparing Learning Material for Adaptive Usage: Structured Content Analysis of Existing Systems

Stefan Thalmann

Innsbruck University - School of Management, Information Systems,
Universitaetsstrasse 15, 6020 Innsbruck, Austria
Stefan.Thalmann@uibk.ac.at

Abstract. Using learning material in an adaptive way requires a preparation beforehand. In order to guide the process of creating learning material, criteria for adaptation and techniques for content preparation have to be known. The aim of this paper is to give first hints for such guidance. Therefore, adaptation criteria and provisions which can be used to realize such a preparation should be identified. For this purpose, adaptation criteria were derived by investigating 30 existing adaptive hypermedia systems. Based on these criteria first hints for preparation of learning material for adaptive use will be given.

Keywords: adaptation criteria, content preparation, structured content analysis.

1 Introduction

Learning is an individual process of gathering and understanding knowledge and learners have different preferences and requirements for learning. This diversity commonly requires a supply of learning material that fits to the learners' needs [1]. In traditional learning scenarios teachers should perform this task. In the case of e-learning, this task could be realized by adaptive hypermedia systems [2]. Thereby, a system is called *adaptable* if the user has the opportunity to change the system characteristics and it is called *adaptive* if the system is able to change its own characteristics automatically according to the users' needs [3]. Before using learning objects (LO) in an adaptable or adaptive way it is necessary to prepare them according to the adaptation purpose. The preparation is normally quite expensive [4], so that a preparation of the whole organizational knowledge base is not realizable. Furthermore, the extreme rise of organizational documents [5] which is additionally boosted by the trend of user-generated content, leads to an unmanageable set of resources which could be used for educational purposes. Therefore, organizations have to decide systematically which knowledge elements of their knowledge base should be prepared for an adaptive usage.

So far, this decision is normally made without a systematic deliberation. Typically, resources which can be reused frequently would be prepared. Surely, frequent reusability reduces the cost per usage (economies of scale) and seems to be a good indicator. But chances for development of a certain topic, importance of LO for specific situations or

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 411–418, 2008.

strategic thoughts are mostly not considered. In times of rapidly growing amounts of organizational contents, this intuitive approach cannot be realized successfully.

This paper focuses on structuring that decision problem by identifying adaptation criteria and giving first hints for the content preparation for an adaptive use. Section 2 describes the content preparation for adaptive hypermedia systems and a study to identify adaptation criteria is presented in section 3. Provisions for content preparation according to the identified adaptation criteria are presented in section 4. Finally, the paper closes with a conclusion and outlook.

2 Content Preparation for Adaptive Hypermedia Systems

The learner´s need for LOs that fit to her preferences and environmental circumstances can be seen as a need for adaptation. In order to fulfil this adaptation need, LOs have to be prepared. A well known distinction of adaptation between adaptation of presentation and adaption of navigation was proposed by Brusilovsky [2]. Adaptation of navigation can be realized by arranging a sequence of content fragments, e.g., LO, in order to adapt the navigation possibilities according to an adaptation need. The presentation can be adapted by presenting different pieces or versions of content to the learner [2]. Generally, versions of content with specific characteristics can be associated to rules or models used in adaptive hypermedia systems [2], [6]. Defining such rules is based on the assumption that different versions of knowledge, like LOs, are available.

Frequently, the realized adaptation is based on topics and thus, existing documents can be assigned to an adaptation need. By contrast, adaptation based on other criteria, like preferences or learning styles, requires different versions of contents for one topic. Adaptation rules are usually not defined for single resources rather than for characteristics of LOs. These characteristics are technically described by metadata [7]. The need for quality metadata in adaptation processes is ample discussed in literature, e.g. [8], [9], [10]. Thus, the most generic applicable step for using LOs adaptively is to describe them by meaningful metadata.

If knowledge is not available in different versions with characteristics useful for adaptation, contents have to be prepared. Thus, different LOs have to be created by modifying the characteristics of one available or created LO targeting one topic. Dimensions of adaptation and thus, the required characteristics of contents have to be known in order to create these different versions and to identify techniques to prepare contents. Identified opportunities for adaptation and related content preparation techniques represent action alternatives of the mentioned decision problem and can be seen as a first structuring approach. In order to investigate dimensions of adaptation and associated content preparation techniques, a study investigating 30 adaptive hypermedia systems was realized.

3 Study to Identify Adaptation Criteria

The main goal of this study is to identify criteria to which contents can be adapted. These general criteria should represent the starting point of an adaptation action and instances of the criteria could be used to define adaptation rules. In the following, these criteria are called *adaptation criteria*. Because of the limited literature in the field of techniques to prepare content for adaptive use a survey of existing systems was conducted.

3.1 Study Design and Procedure

The first step of this study was to identify relevant adaptive hypermedia systems which can be analyzed. One initial classification of adaptive hypermedia systems was found in the very well cited article from Brusilovsky [2]. To structure the identified systems, the author decided to use these three categories: adaptive hypermedia systems in education, adaptive hypermedia systems for serving on-line information, and adaptive hypermedia related to information retrieval (IR) problems. Overall, 158 systems could be found consisting of 56 adaptive educational systems, 52 adaptive on-line information systems and 50 adaptive systems for IR. From the author's point of view, this collection represents a good trade-off between efforts and benefit for searching.

In the second step, a qualitative content analysis based on the approach described in [11] was realized. During the identification and classification of existing systems no detailed investigation of system functionalities has been conducted. The qualitative content analysis uses initial predefined categories derived from literature and forms the starting point for the study [11]. The openness is ensured by allowing a flexible adjustment which is realized by changing or deleting initial categories or creating new categories if necessary [11]. Thus, an existing theoretical foundation can be integrated without losing or not considering the specifics of researched data.

Brusilovsky describes some criteria in [2], to which information can be provided adaptively. These criteria were developed by analyzing existing systems known in 2001. These established criteria were used to inform the study. The criteria are: user interests, user knowledge, individual traits, platform, and location[2]. Brusilovsky describes these criteria in [2] without revealing analysis process and compares his results with those from a paper performed in 1996. This leads to the assumption that these criteria could have changed since 2001 and that they should be examined.

The process of content analysis was quite elaborative and thus, a detailed investigation of all 158 identified systems seemed as too laborious. 10 systems per category seemed as an appropriate sample to identify the criteria for adaptivity. Overall, 30 systems from the 158 identified systems were selected for a detailed investigation. The systems of each category were ordered alphabetically and the first and every 5th system were chosen. For every paper, at least one and at most two academic papers (conference-, journal-paper, technical report or Ph.D – Thesis), have to be identified for detailed investigation. Because of the limited space for this paper, a list of the 30 considered systems with citations can be found at http://www.uibk.ac.at/iwi/adaptation_criteria.html.

For 26 systems two academic papers each could be considered for the structured content analysis. For four systems only one scientific paper could be considered. Overall, 56 academic papers were considered for the detailed investigation of adaptive criteria.

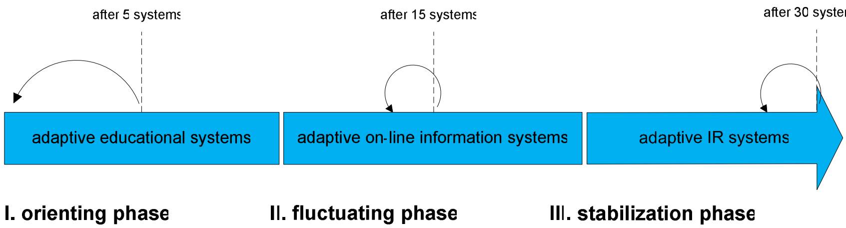


Fig. 1. Coding procedure performed in the study, based on [13]

For identifying the categories for adaptation, a qualitative content analysis approach was chosen. Thereby, the procedure based on the iterative approach proposed by Dey [12]. The phases of content analysis are independent from researched systems.

The content analysis starts with adaptive educational systems and an orienting phase using the initial criteria. New criteria, material can be adapted to, were defined if a description of an adaptation based on these criteria could be found in a scientific paper and these criteria does not fit perfectly to one predefined criteria. Thus, a criterion had to be mentioned in a paper which was used to provide information adaptively. The criteria were formulated in a first step and described more precisely over the time. After the investigation of five systems the first draft of criteria looked reasonable stable to a certain degree. Changes during the development of the first draft of codes demanded a rescan and recoding of the first five systems.

By coding the first five adaptive on-line information systems it turned out that some new criteria occur. To ensure the homogeneity within and between all 15 investigated systems, they were rescanned and recoded. Thereby, few changes in the first 10 adaptive educational systems were necessary. Most changes are related to the new defined criteria, especially relevant for on-line information systems. During analysis of the last 10 systems, no more specific codes for adaptive IR systems were found. Definitions for criteria turned out as suitable and good applicable in this phase. Only few criteria were merged or changed marginally. Overall, the criteria are stable and all identified characteristics could be represented by the final codes.

3.2 Results

The results of the structured content analysis are 13 adaptation criteria which are explained below. Beside the names of the adaptation criteria relative occurrence within the 30 adaptive hypermedia systems are depictured in brackets.

knowledge structure (73,3%): A list of terms or taxonomy of concepts, sub-concepts and their relations in a knowledge domain. The knowledge structure is used to differentiate one topic in several sub-topics and to use different characteristics of the sub-topics for adaptation.

user history (80%): A collection of data describing previous user interactions with a system. Usually, these data are log-data from systems the user has used and data which are useful for future adaptation provisions. Data from the user history are normally analyzed in order to derive values which can be represented by other criteria.

user request (27%): Additional interaction possibilities for user-initiated adaption. They are additional interaction possibilities which are provided to the user in order to change (1) a navigation path or (2) to consider requirements or values entered by the user. In the easiest way these opportunities are additional links and in more advanced cases input boxes are provided. In the context of adaptive hypermedia systems these user requests are used to adapt the systems information supply.

previous knowledge (47%): Knowledge of the user acquired in the past and relevant for using the system which has to be considered for information provision. This knowledge is required in order to gather the provided information successfully. Previous knowledge can be used to transfer information adaptively according different levels of previous knowledge.

content preferences (93%): Goals and interests for presented topics in a certain knowledge domain. Thereby, not the presentation style rather than the presented contents are affected and thus the choice of topics of presented contents is influenced. Different preferences, goals and interests can be used in order to select appropriate topics for the information requester adaptively.

presentation preferences (26,7%): Goals and interests for the style of content delivery. These preferences are used in order to present versions of the same content in different presentation styles, e.g., one figure in different colours or sizes. Thus, users feel more comfortable by consuming information which fit to their presentation preferences.

preferences for media types (7%): Goals and interests for the technical format of content delivery. Users have different preferences according content types, which can be a distinct to preferences of colour settings and usability aspects. Thus, different versions of the same content in the same presentation style can occur in different media types and thus transferred adaptively.

learning style (13%): Preference of a user for proceeding during learning. The suitability of these theories depends on learners and topics and they should increase learner's ability to absorb certain knowledge. Thus, depending on the learner and the topics, contents can be prepared according to different learning styles and pedagogy approaches and provided adaptively.

language (10%): Ability or preference of a user for the language used for content delivery. Information written in a different language can be a significant access barrier for learners. Thus, information should be provided adaptively considering the user's language skills.

device requirements (13,3%): Technical characteristics of the hardware and basic software relevant for accessing the system. Especially, different screen resolutions and colour settings have to be considered in order to ensure a correct presentation. Thus, versions which fit to the different hard- and software requirements have to be created and provided adaptively to ensure a smoothly information provision to the user.

bandwidth (10%): Data transfer rate available during a session with the system. Different quality of internet access influences speed and quality of information provision. The current bandwidth depends not only on the device in use and thus has to be investigated for every initiated session. Unsuitable large data volumes cause in very long loading times or in the worst case the user cannot access the information. In order to ensure a comfortable information access data volume of transferred information should be adapted to the current bandwidth.

location (20%): Physical coordinates relevant for content delivery. The physical location could indicate a certain demand for contents which can be related to a location. By using GPS, IP or cell-based techniques this location is available and could be used for adaptation of information access.

user status (3%): User-related or environmental characteristics describing the users current circumstances. The current status of a user can influence his ability to access information and her preferences. Needs and preferences based on user's status, like stress or loud environment cannot be generalized and has to be entered by the user. Depending on different situational status of the user, information can be provided adaptively.

4 Content Preparation According to Adaptation Criteria

In order to create different versions of LOs according to the identified adaptation criteria, contents have to be prepared. In the following, provisions to prepare contents according to the identified criteria are proposed.

Knowledge structure, user history and user request act as a kind of top level criteria. They are mainly used to structure or to specify other criteria. Furthermore, required content preparation techniques can assign to one of the previous mentioned criteria. For that reason, no specific preparation technique could be identified rather than all other preparation techniques could be applied in principle. As mentioned before, creation of meaningful metadata is necessary for all adaptation criteria and thus, general relatable.

Preparing contents considering previous knowledge should prevent swamping learners with less knowledge and boring learners with high knowledge [14]. The preparation could be achieved by (1) adding information required to understand the offered learning material or (2) shortening information by facts, well-known by people with a certain knowledge level.

Content preferences can be considered by relating LOs to different topics. Thus, different versions, which can be related to different topics, should be created. Thereby, breaking one piece of content into various sub-units and the aggregation of several pieces of content to form larger units, called modularization, can be applied [15]. Contents considering presentation preferences can be created by changing font settings, colours or redesigning the layout of a LO.

Creating different versions of learning material according to learning styles or pedagogy approaches requires a didactical revision. These revision can be realized by adapting the navigation, e.g., modification of rules or creating different versions of learning material that conform with one learning style [16]. Considering different languages can be realized by translating LOs into different languages.

Adapting learning material to a location can be achieved by enriching contents by specifics of the location. That could be realized by adding local context to the learning material, in such way, that the local environment is linked to the learning material. Creating different versions of learning material according to the user's status can be realized by situational preparation. Thus, specifics of a certain status can be considered and contents prepared according to the requirements of that status.

The preparation according device requirements and preferences for media types mainly concentrates on syntactical aspects. Because of the unknown syntax of resources, e.g., wrong formats and resources cannot be processed and have to be converted. Limitations of capacity, like screen resolutions, colour settings, or bandwidth, require transformation in order to readjust the size. That is primary necessary for multimedia contents, like pictures or videos. By converting resources into other formats or readjusting their sizes the appearance can be affected. Bad quality or deformations could appear, but the content itself, its meaning and sense, thus, semantics, is intentional not affected. The process of transferring content from one format or size into another is called transcoding [17]. Transcoding of resources can be realized automatically during build time [18] or during runtime [17].

5 Conclusions

Importance of personalized learning materials increases with the growing popularity of ubiquitous technologies in educational technology. In order to use learning material adaptively a preparation targeting the learners' adaptation needs is necessary. This preparation can be realized by varying the sequence of existing LOs or by preparing or creating specific versions. The need for a systematic procedure for adaptive content preparation increases by the growing demand for adaptive material. This paper introduces 13 adaptation criteria for preparing learning material adaptively and presents first hints of content preparation according these criteria. Both can be used to describe the action alternatives of content preparation. In order to guide the process of selection and preparation of contents for adaptive usage, these criteria can be seen as a first step for the development of such a procedure.

In further steps, these criteria will be sophisticated and together with other criteria a decision model will be developed. This decision model should answer the question: *Which organizational knowledge elements should be prepared for adaptive usage and which techniques should be applied?* In the next step, all authors of papers considered for the detailed investigation in section 3 should be asked within a survey. The main idea is to give developers the opportunity to correct values and to report about planned features and missing factors. Afterwards, all developers of the 158 identified systems should rate the identified criteria and their importance.

References

1. Kareal, F., Klema, J.: Adaptivity in e-Learning. In: IV International Conference on Multimedia and Information and Communication Technologies in Education, Seville, Spain (2006)
2. Brusilovsky, P.: Adaptive Hypermedia. User Modeling and User Adapted Interaction 11(1/2), 87–110 (2001)
3. Oppermann, R.: Adaptively supported Adaptability. International Journal of Human-Computer Studies 40(3), 472–544 (1994)

4. Brusilovsky, P., Knapp, J., Gamper, J.: Supporting teachers as content authors in intelligent educational systems. International Journal of Knowledge and Learning 2(3/4), 191–215 (2006)
5. Lyman, A., Varian, B.: How Much Information? University of California, Berkeley (2003)
6. Cristea, A., Stewart, C.: Authoring of Adaptive Hypermedia. In: Magoulas, G.D., Chen, S.Y. (eds.) Advances in Web-Based Education: Personalized Learning Environments, pp. 225–252. Information Science Publishing (2006)
7. Maier, R., Thalmann, S.: Describing Learning Objects for Situation- Oriented Knowledge Management Applications. In: Gronau, N. (ed.) 4th Conference of Professional Knowledge Management (WM 2007), pp. 343–351. GITO Verlag, Potsdam (2007)
8. Dagger, D., Wade, V., Conlan, O.: Personalisation for All: Making Adaptive Course Composition Easy. Journal of Educational Technology & Society 8(3), 9–15 (2005)
9. Karampiperis, P., Sampson, D.: Adaptive Learning Objects Sequencing for Competence-Based Learning. In: 6th IEEE International Conference on Advanced Learning Technologies. IEEE Computer Society Press, Kerkrade (2006)
10. Motelet, O., Baloian, N.: Taking advantage of LOM Semantics for Supporting Lesson Authoring. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2005. LNCS, vol. 3762, pp. 1159–1168. Springer, Heidelberg (2005)
11. Gläser, J., Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. 2. durchgesehene Auflage ed, Wiesbaden: Verlag für Sozialwissenschaften (2006)
12. Dey, I.: Qualitative Data Analysis. A User friendly Guide for Social Scientists. Routledge, London (1993)
13. Hädrich, T.: Situation-oriented Provision of Knowledge Services, Ph.D. Thesis 2008, University of Halle-Wittenberg, Halle(Saale), Germany (2008)
14. Frohberg, D.: Mobile learning in tomorrow's education for MBA students. In: Proceedings of MLearn 2004: Rom, Italy, pp. 81–84 (2004)
15. Petrides, L., Nguyen, L., Kargliani, A., Jimes, C.: Open Educational Resources: Inquiring into Author reuse Behaviors. In: Dillenbourg, P., Specht, M. (eds.) ECTEL 2008, Maastricht, The Nederlands. LNCS, pp. 344–353. Springer, Heidelberg (2008)
16. Gilbert, J.E., Han, C.Y.: Adapting instruction in search of 'a significant difference'. Journal of Network and Computer Applications 22, 149–160 (1999)
17. Ihde, S.C., Maglio, P.P., Meyer, J., Barrett, R.: Intermediary-based transcoding framework. Ibm Systems Journal 40(1), 179–192 (2001)
18. Lum, W., Lau, F.: On balancing between transcoding overhead and spatial consumption in content adaptation. In: ACM Mobicom., Atlanta, pp. 239–250 (2002)

How Image Based Factors and Human Factors Contribute to Threat Detection Performance in X-Ray Aviation Security Screening

Anton Bolfing, Tobias Halbherr, and Adrian Schwaninger

Department of Psychology, University of Zurich, Switzerland
Max Planck Institute for Biological Cybernetics,
Tübingen, Germany
a.bolfing@psychologie.uzh.ch, t.halbherr@psychologie.uzh.ch,
a.schwaninger@psychologie.uzh.ch

Abstract. The present study examines the relative importance of a series of known and expected factors that highly affect threat detection performance in aviation security X-ray screening. Examined image based factors were threat item view difficulty, threat item superposition, bag complexity and bag size. Further, also the two human/demographic factors X-ray image interpretation training and age were examined. Image measurements and performance statistics for factors estimation are introduced. Three statistical approaches were applied in order to examine the impact of the introduced factors on threat detection performance and revealed consistent results. Bivariate correlations between detection performance and predictors/factors were analysed to estimate the isolated impact of each single factor independently of any other. Multiple linear regression analyses were applied for estimating the overall impact of all image based factors and human/demographic factors respectively. And analyses of covariance were applied in order to check for possible interaction effects between all factors of the models. All analyses were applied separately for the four threat item categories guns, knives, improvised explosive devices and other.

Keywords: aviation security, human factors, image based factors, X-ray imaging.

1 Introduction

Together with the tremendous growth of civil aviation the importance of aviation security and its public perception has dramatically increased in the last few decades [1]. The security checkpoints at the gates for x-raying passenger bags are the key element in aviation security all over the world. Despite great improvements in technical equipment, including high resolution X-ray machines, the decision whether a piece of luggage can enter an airplane or not is still made by human screeners. Therefore aviation security officers and their activity in screening passenger bags are a critical link of utmost importance in aviation security.

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 419–438, 2008.

In 2007 Schwaninger, Michel, and Bolfing [2] contributed an article on X-ray image difficulty estimation based on a set of image based factors. The study revealed that it is possible to predict average detection performance (across a sample of participants) on a single image quite well solely based on computationally accessible image properties. The image based factors used in that model were View Difficulty, Superposition, Clutter and Transparency. All image based factors can be automatically calculated. Multiple linear regression was used for statistical modeling. A comparison between the model based on automatically computed predictors (image based factors) and the same model based on human ratings (of the image based factors) revealed that our image measurements and statistics can predict human performance as well as human raters can. The study was based on a participants sample of 12 undergraduate students and on a X-ray object recognition test consisting of 256 images. The study reported in the following is an extensive amplification of the 2007 article. We were able to replicate the results of the earlier study with professional screeners and additional extensions in terms of data set size, additional factors and additional statistical analyses. The new test consists of 2048 test items and results are based on a participants sample of 90 screeners. Furthermore, three additional factors have been included: Human factors, namely Training and Age as well as the image based factor Bag Size. The number of threat categories was doubled by adding improvised explosive devices (IEDs) and 'other' to guns and knives.

In this study we analyze the effects on detection performance of prohibited items in passenger bags of two different groups of factors: 'Human factors' and 'image based factors'. The concept of image based factors subsumes all properties of the passenger bags' X-ray images that are relevant in mediating performance in detecting prohibited items. The concept of human factors subsumes available properties of the persons performing the screening task relevant in mediating threat detection performance. The aim of the reported study in this article is to investigate the role of image based and human factors on the threat detection performance in passenger bag screening tasks. For this purpose the effects of the different factors on threat detection performance, as well as their interactions will be assessed as comprehensively as possible. Previous work [3,4,2] has identified the following performance relevant image based factors: Threat Object View Difficulty, Superposition by other objects and Bag Complexity (represented in the following by Opacity and Clutter). The experiment is based on an off-line computer based test consisting of 2048 trials. The test is designed with the four image based factors View, Superposition, Bag Complexity and Bag Size systematically varied in order to avoid confounded variables. This design allows analysis of individual and combined effects of the image based factors, as well as analysis of their interactions. Furthermore we will analyze data of the human factors Training and Age [5]. Training was operationalized as the amount of hours spent on training using the 'X-Ray Tutor' computer based training system prior to testing.

1.1 Image Based Factors

Schwaninger (2003) [3] and Schwaninger, Hardmeier, and Hofer (2004) [4] have identified three image based factors which affect human threat detection performance

Fig. 1. Image based factors

significantly: View Difficulty, Superposition, and Bag Complexity. These image based factors have been modeled into mathematical formulae [2,6]. View Difficulty is implemented as an a posteriori calculable value named FTI View Difficulty. The abbreviation FTI represents fictional threat item, X-ray images of threat objects being artificially projected into X-ray images of passenger bags. Superposition and Bag Complexity are implemented as image processing measurements whereby Bag Complexity is split up into Clutter and Opacity. The introduction of the image based factor Bag Size in this study necessitated normalization of earlier implementations of Clutter and Opacity regarding bag size.

FTI View Difficulty. The general formula for FTI View Difficulty can be seen in Equation 1. It is a slight modification of the mean of the inverted detection performance value (DetPerf) over all items (index N_{OV}) containing the same FTI object (subindex O) in the same view (subindex V) as the item in question. 'Inverted' refers to the fact that the measured detection performance is subtracted from a theoretical maximum detection performance (max(DetPerf)), in order to ensure that high values of FTI View Difficulty correspond with high difficulties. The slight modification refers to the exclusion of the item in question from averaging.

$$\text{FtiVD}_{OVj} = \frac{\sum_{i=1,j\neq i}^{N_{OV}} \left(max(\text{DetPerf}) - \text{DetPerf}_{OVi}\right)}{N_{OV} - 1} \tag{1}$$

Superposition. Superposition is modeled as the inverted Euclidean distance in pixel intensity between an SN image (signal plus noise or image containing a threat item) and its corresponding N image (noise or non-threat image).

$$SP = C - \sqrt{\sum_{x,y} \left(I_{SN}(x,y) - I_N(x,y)\right)^2} \tag{2}$$

Please note that in all reported analyses we used logarithmically transformed Superposition values. After inspecting the scatterplots of all our factors with the detection performance d' in order to check for non-linear relationships (a violation of the requirements of multiple linear regressions) we decided to linearize Superposition by this log-transform. This way heteroscedasticity (another violation of the requirements of MLR) can be avoided and the explained variance of the relationship between the factor and detection performance can be increased.

Clutter. Clutter is designed to capture bag image properties like disarrangement, textural noise, chaos or just plain clutter. We modeled the Clutter variable based on the assumption that image properties like the ones mentioned above correspond with larger amounts of high spatial frequencies in the image. Equation 3 represents a convolution of the empty bag image (N for noise) with the convolution kernel derived from a high-pass filter in the Fourier space. I_N denotes the pixel intensities of the harmless bag image. $\mathcal{F}^{-1}$ denotes the inverse Fourier transformation. $hp(f_x, f_y)$ represents a high-pass filter in the Fourier space. BS represents Bag Size (see Equation 5). Cut-off frequency f and transition d (the filter's order) were set to $f = 0.03$ and $d = 11$. The pixel summation on the high-pass filtered image was restricted to the bag's area.

$$\text{CL} = \frac{\sum_{x,y} I_{hp}(x,y)}{BS} \tag{3}$$

$$\begin{aligned} \text{where} \quad I_{hp}(x,y) &= I_N * \mathcal{F}^{-1}(hp(f_x, f_y)) \\ &= \mathcal{F}^{-1}(\mathcal{F}(I_N \cdot hp(f_x, f_y)) \\ \text{and} \quad hp(f_x, f_y) &= 1 - \frac{1}{1 + \left(\frac{\sqrt{f_x^2 + f_y^2}}{f}\right)^d} \end{aligned}$$

Opacity. Opacity represents how well X-rays are able to penetrate an object. High Opacity values represent low penetrability. In X-ray images this property is represented by pixel color and brightness. Opacity represents the total size of areas with pixels being darker than a certain threshold relative to the bag's size. In Equation 4 all pixels being darker than a certain threshold (e.g. 64) are summed up and divided by the bag's size (Bag Size as denominator).

$$\text{OP} = \frac{\sum_{x,y} \left(I_N(x,y) < 64\right)}{\text{BS}} \tag{4}$$

Bag Size. The Bag Size formula below is applicable to grayscale images with pixel brightness values ranging from 0 (black) to 255 (white). All pixels with brightness values lower than 254 (almost white) are considered as part of the bag. Bag Size is then defined as the size of the bag in number of pixels

$$\text{BS} = \sum_{x,y} (I_N(x,y) < 254) \tag{5}$$

2 Methods and Procedures

2.1 Participants

The participants sample consists of 90 professional aviation security X-ray screening officers from two European airports (48 females). On average females are 40.6 and males 35.9 years old with standard deviations of 17.8 and 13.6 years respectively.

2.2 Stimuli

The 2048 test stimuli were created automatically using the image measurement algorithms described above. The number of trial images is determined by the following test design: The test consists of eight threat exemplar pairs per threat category. Given the categories Guns, Knives, IEDs and Other this results in 64 different exemplars of threat items. Each of these threat items is presented with each possible factor combination. Each of the image based factors introduced above is implemented in the design with two dichotomous parameter values representing low and high values. For View Difficulty, Superposition, Bag Complexity and Bag Size this results in 2x2x2x2= 16 factor combinations. The 64 threat exemplars in 16 factor combinations result in 1024 images. In order to apply signal detection theory [7] in the analysis all 1024 bag images containing fictional threat items (FTIs) are also presented in the test not containing any threat items. This results in the total of 2048 images.

The construction process of the test stimuli was partly manual and partly automated. In a first step the 64 threat exemplars were chosen manually such that the diversity of threat items in each of the four categories is well represented. In a second step a set of 1024 bag images was chosen based on the image measurements introduced above. In total 6659 bag images were analyzed regarding Clutter, Opacity and Bag Size. The 1024 bag images that were finally used for the test were all independently checked for credibility by X-ray screening experts. Subsequently we determined the membership of each image regarding high or low parameter values by applying median splits on each of the three image based factor distributions Opacity, Clutter and Bag Size. Opacity and Clutter are very highly intercorrelated. Thus it did not make sense to define and vary high and low parameter values for Opacity and Clutter independently. Instead the dichotomous variable Bag Complexity was defined based on Opacity and Clutter: For Bag Complexity high and low parameter values were defined as bags with both high or low Opacity and Clutter values, respectively. Bags with high Opacity and low Clutter values or vice versa were discarded. For each of the resulting factor combinations - low/high Bag Complexity x small/large Bag Size - 256 images were chosen manually. In the last step fictional threat items were automatically merged with the harmless bags. Each of the 64 threat exemplars exists in two different views. The easy views were depicted in frontal view and the difficult ones were depicted in a 85° rotation relative to the frontal view, either horizontally or vertically. This results in a total of 128 fictional threat items (FTIs). All FTIs were recorded by professional X-ray screening experts for credibility. Each of these 128 FTIs was finally merged with the 256 harmless bags with two different levels of Superposition - low and high. This procedure was applied four times for each combination of harmless bags. As already mentioned, this process was fully automatic. The underlying merging algorithm merges the images and calculates the Superposition value. If the Superposition value lies in the low or high superposition level range it is being saved as such. If not, the process is repeated until the FTI can be merged within the desired superposition value range.

2.3 Procedure

Since the test contains a very large amount of items participants completed it over multiple sessions. The test presentation was implemented within the computer based training system X-Ray Tutor 2.0 which can be run as a testing and a training environment. X-Ray Tutor is a well-established training tool designed to effectively improve and reliably test X-ray image interpretation competency. Customer airports are recommended to advise their security screeners to practice at least 20 minutes per week using X-Ray Tutor. The current test was inserted into the familiar training sequences and continued after each login until completion. After that, normal training continued. The shared basis of training and testing allows extracting training data of each screener prior to testing.

The 2048 images are presented in random order. The participants' task is to decide whether a piece of luggage would be OK or not OK to hypothetically enter an airplane by pressing buttons OK or NOT OK with a computer mouse. Unlike in training mode, in testing mode participants receive no feedback with regards to the correctness of their answers. Data are recorded in a database.

2.4 Statistics

The data are analyzed in three ways. First we present all bivariate correlations between the independent variables (image based factors and human factors) and detection performance d'. This gives a good estimation of how well detection performance can be predicted on the basis of our predictors. The second statistics we present are two separate linear regression models, one for image based factors and one for human factors, respectively. Regression analyses allow estimating the combined impact of the respective predictors together. Regression analyses allow estimating to what extent a certain set of predictors is able to predict the measured values [8], in this case the measured detection performance d'. The regression analysis using the image based factors as predictors is a replication and refinement of an earlier study by Schwaninger et al. [2] which was based on the X-Ray ORT [9]. The present analysis is based on a much larger item- and subject sample since we used a new test for the present study. The most notable differences between the X-Ray ORT and this test are the inclusion of the new image based factor Bag Size and the extension of the threat item categories by IEDs (Improvised Explosive Devices) and "Other". Since linear regression analyses do not take into account any kind of interaction effects between the predictors we report a third type of analysis. The analysis of covariance (ANCOVA), with our image based factors operationalized as repeated measures variables and the human factors as covariates. It is important to be aware of potential interaction effects, since the presence of large interactions would limit the amount of variance explained by the multiple linear regression models.

3 Results

In conformity with the Statistics section the results are reported in the following order: First we report the findings of the bivariate and partial correlations. In a second step we report the results of the multiple linear regression analyses per threat item category as well as for all categories combined. One set of multiple linear regression analyses will be based on image based factors and the other on human factors. Finally we report an ANCOVA with the image based factors as repeated measures and with the human factors Age and Training as covariates.

3.1 Bivariate Correlations

Figure 2 shows the bivariate correlations with d' of each image based factor and the partial correlations with d' of each human factor - with the respective other human factor serving as a control variable. The reason why we decided to treat image based factors and human factors differently is the following: The image based factors have been implemented within the computer based test in order to obtain orthogonal relationships between them. In other words, image based factor values vary independently across test items. Since we could not ensure independence of the human factors Age and Training through test design or sample selection, orthogonal relationships between human factors cannot be assumed. The data reveal that indeed Age and Training are confounded, with people tending to train more the older they are. Therefore we decided to additionally report partial correlations to avoid false conclusions regarding the effect of age and training on visual search tasks. Furthermore we decided to graphically report R^2 values instead of plain Rs. The great advantage of R^2 over R is that it can be directly interpreted as the amount of variance in the dependent variables (d') that can be explained by the independent variable (single factors). The disadvantage is the loss of information on the sign of the relationship due to squaring.

Figure 2 illustrates the relationship of the individual factors on detection performance d' by threat category for image based factors and human factors separately. The graphs clearly reveal that three factors explain a substantially higher amount of variance than all the others, i. e. FTI View Difficulty, Superposition and Training (log-transformed training hours). Age also shows a notable effect which is much smaller, but remains stable across all threat categories. Exact values are reported in Table 1. A detailed discussion on the data patterns is given in the Discussions section at the end of this article.

3.2 Multiple Linear Regression Analysis

Figures 3 - 5 all show scatterplots illustrating the statistical relationship between the observed (empirically measured) detection performance values d' (ordinate) and the

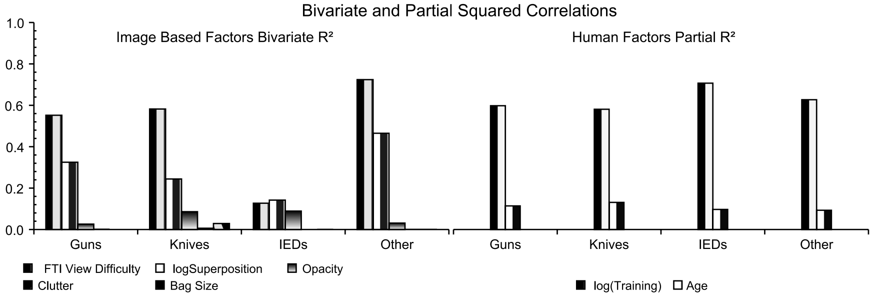


Fig. 2. The bivariate R^2 values for the image based factors and the partial R^2 values for the human factors are estimates for the amount of variance in detection performance d' that can be explained by the single factors

Table 1. Tabulation of the correlations between the single factors (human and image based) and the detection performance measure d' separately for each threat category. For human factors, also partial correlations are given, the respective other human factor taken as the control variable.

	Image based factors bivariate correlations with d'					**Human factors** partial correlations with d'		bivariate correlations with d'	
	FTI-VD	logSP	OP	CL	BS	logTR	Age	logTR	Age
Guns	-.74	-.57	-.16	-.06	-.02	.77	-.34	.75	.19
Knives	-.76	-.49	-.29	-.09	-.17	.76	-.36	.73	.16
IEDs	-.36	-.38	-.30	-.02	-.05	.84	-.31	.84	.28
Other	-.85	-.68	-.18	-.05	-.05	.79	-.31	.78	.24

standardized predicted values estimated by the respective multiple linear regression models (abscissa). For each model R^2 and R values are displayed in the bottom right corner of the scatterplot as a measure for the closeness of the relationship between model prediction and empirical measurements.

Models Across Categories. Figure 3 shows the scatterplots of the multiple linear regression models for the image based factors and human factors respectively. Differences concerning threat categories are not taken into account here. Both models can explain nearly 70% of the observed variance in d'. In the image based factors model 1024 data points are estimated. Each data point represents one signal-noise/noise item pair with its hit rate and false alarm across all 90 screeners. In the human factors model there are only 90 data points because d' values are calculated per subject across all 1024 item pairs.

Table 2 shows the most important statistical values of the multiple linear regression analyses for both models.

Models by Category. Figures 4 and 5 show the corresponding linear regression model scatterplots to Figure 3 but separately for each threat category. This allows us to

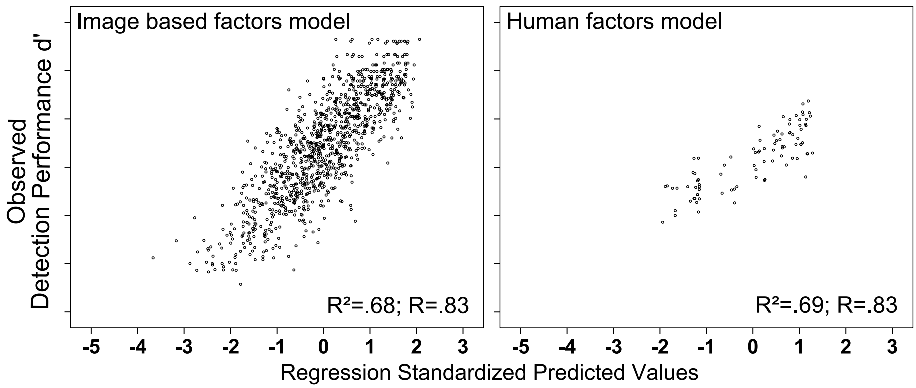


Fig. 3. General regression models across all threat item categories. Scatterplots with standardized predicted values of the image based and human factors multiple linear regression models on the x-axis and observed detection performance d' on the y-axis.

Table 2. Tabular summary of the general multiple linear regression models for all threat item categories. Standardized beta weights and p-values.

Model Summaries (All Categories)			
	Predictors	Beta weights β	Significance p
Image Based Factors	FTI View Difficulty	-.70	.000
	logSuperposition	-.127	.000
	Opacity	-.329	.000
	Clutter	.198	.000
	Bag Size	.021	.288
	$R^2 = .68$, adjusted $R^2 = .68$, $F(5, 1018) = 441$, $p < .000$		
Human Factors	logTrainingHours	.93	.000
	Age	-.26	.000
	$R^2 = .69$, adjusted $R^2 = .68$, $F(2, 87) = 98$, $p < .000$		

compare the relationship between image based factors and threat detection performance with the relationship between human factors and threat detection performance separately for each threat category. Table 3 shows the most important statistical values for each of the reported models.

Figure 4 shows the four scatterplots illustrating the predictive power of the image based factors regression model separately for each threat category.

Figure 5 shows the four scatterplots illustrating the predictive power of the human factors regression model separately for each threat category.

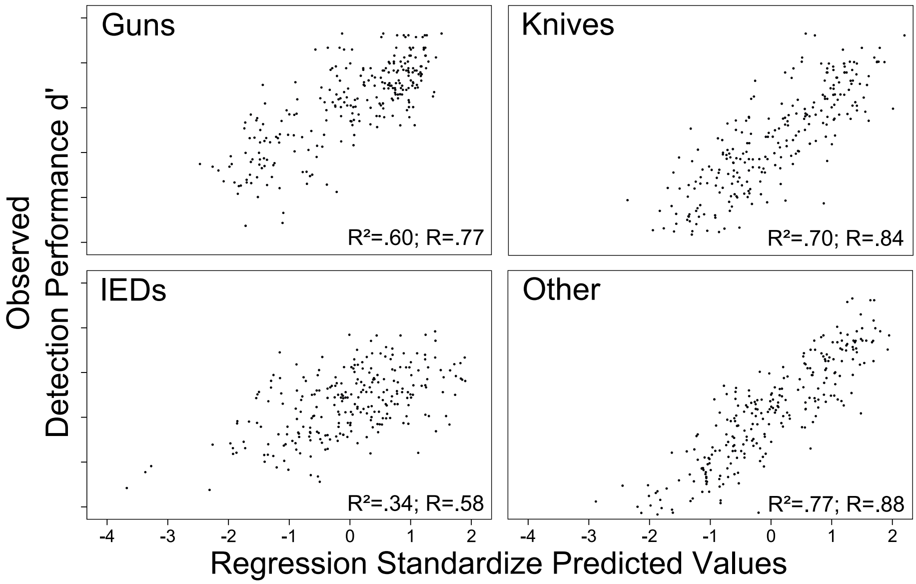


Fig. 4. Separate image based factors regression models for each of the four threat item categories

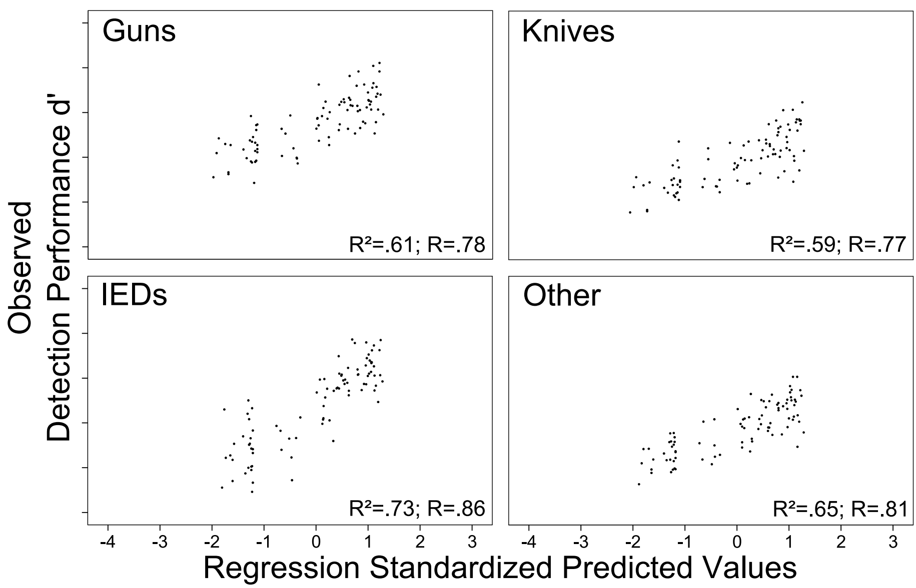


Fig. 5. Separate human factors regression models for each of the four threat item categories

Table 3. Tabular summary of seperate multiple linear regression models for each of the four threat item categories. Standardized beta weights and p-values.

Model Summaries (Per Category)									
	Predictors	Guns		Knives		IEDs		Other	
		β	p	β	p	β	p	β	p
Image Based Factors	FTI View Difficulty	-.67	.000	-.67	.000	-.29	.000	-.79	.000
	logSuperposition	-.12	.025	-.18	.000	-.25	.000	-.09	.061
	Opacity	-.27	.000	-.36	.000	-.51	.000	-.31	.000
	Clutter	.18	.005	.17	.003	.35	.000	.20	.000
	Bag Size	.04	.375	-.05	.256	.09	.105	.05	.187
	R^2	.60		.70		.38		.77	
	$adjusted R^2$	.59		.70		.32		.77	
	$F(5, 250)$	75		117		25		169	
	p	.000		.000		.000		.000	
Human Factors	logTraining	.88	.000	.88	.000	.94	.000	.90	.000
	Age	-.26	.001	-.29	.000	-.20	.003	-.22	.004
	R^2	.61		.59		.73		.65	
	$adjusted R^2$	.60		.58		.72		.64	
	$F(2, 87)$	69		63		118		80	
	p	.000		.000		.000		.000	

3.3 ANCOVA

ANCOVA with Category Treated as a Repeated Measures Factor. Figure 6 shows a short overview of the ANCOVA results. The ANCOVA allows us to integrate human factors as covariates into a repeated measures ANOVA of image based factors (including threat category) and thus allows us to explore interaction effects and dependencies among not just the image based factors, but also between the image based factors in combination with the human factors. On the left, Figure 6 illustrates the importance of the image based factors in terms of their effect size values η^2 (eta square) and their interactions with the human factors. The main effects of each image based factor are reported together with their interaction effects with Training (log-transformed training hours) and Age, respectively. On the right, Figure 6 additionally illustrates the ten largest significant interactions in terms of η^2 values. For details on the data, their patterns and conclusions please consider Table 4 and the Discussions section at the end of this article.

Table 4 shows all η^2 values and the significance levels of the main effects and the interaction effects illustrated in Figure 6.

ANCOVAs by Category. Figure 7 shows the results of the ANCOVAs applied separately for each threat item category. Only image based factors main effects and interactions with the human factors are given here. Refer to the Discussions section for detailed interpretation.

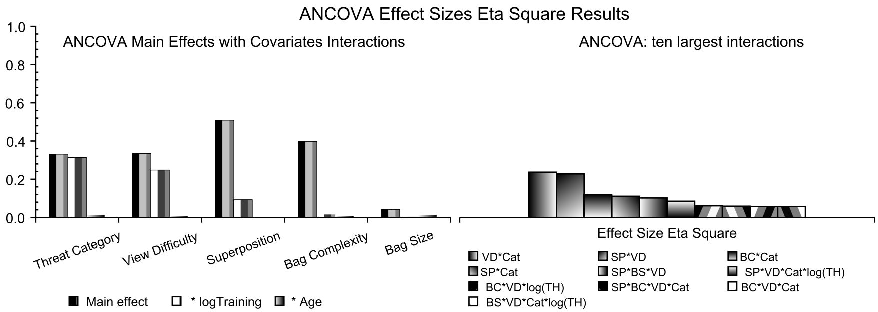


Fig. 6. Summary of ANCOVA main effect and interaction effect sizes. All covariate interactions and the ten largest remaining interactions are reported.

Table 4. Summary of ANCOVA main effects and covariate interactions

ANCOVA effect sizes η^2					
	Category	View Difficulty	Superposition	Bag Complexity	Bag Size
Main effects	$.33^{***}$	$.34^{***}$	$.51^{***}$	$.40^{***}$	.04
* logTraining	$.31^{***}$	$.25^{***}$	$.09^{**}$	.02	.00
* Age	.01	.01	.00	.01	.01

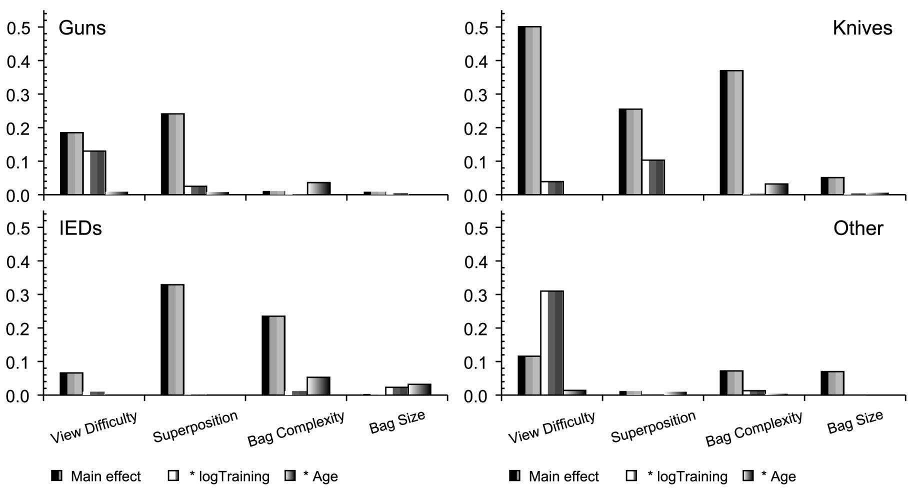


Fig. 7. ANCOVA by Category

Table 5. Tabular presentation of the ANCOVA results by threat item category

ANCOVA effect sizes η^2						
	Category	View Difficulty	Superposition	Bag Complexity	Bag Size
Main effects	$.33^{***}$	$.34^{***}$	$.51^{***}$	$.40^{***}$	.04
* logTraining	$.31^{***}$	$.25^{***}$	$.09^{**}$	.02	.00
* Age	.01	.01	.00	.01	.01
Main effects	Guns	$.19^{***}$	$.24^{***}$	.01	.01
* logTraining		$.13^{**}$	.03	.00	.00
* Age		.01	.01	.04	.00
Main effects	Knives	$.50^{***}$	$.26^{***}$	$.37^{***}$	$.05^{*}$
* logTraining		.04	$.10^{**}$	.00	.00
* Age		.00	.00	.03	.01
Main effects	IEDs	$.07^{*}$	$.33^{***}$	$.24^{***}$	.00
* logTraining		.01	.00	.01	.02
* Age		.00	.00	$.05^{*}$	.03
Main effects	Other	$.12^{**}$	.01	$.07^{*}$	.07
* logTraining		$.31^{***}$	.00	.01	.00
* Age		.01	.01	.00	.00

Table 5 shows all η^2 values and the significance levels of the main effects and the interaction effects illustrated in Figure 7.

4 Discussions

For the discussion of our findings we retain the same presentation order as we did in the Methods and Results sections. Starting with the findings of the bivariate correlations we continue discussing the multiple linear regression models ending with the discussion of ANCOVA results.

4.1 Bivariate Correlations

The correlations between our factors and d' can be interpreted as the observed relationships between our predictors and d' observations. This gives us a first impression of how much explained variance we can expect from a single predictor. Figure 2 and Table 1 reveal that there are close relationships between d' and the three predictors FTI View Difficulty, Superposition (log-transformed) and Training (log-transformed training hours). Also we can report a notable (partial) correlation between Age and d'. The remaining three predictors Opacity, Clutter and Bag Size show poor correlations. Clutter and Bag Size do not even reach the level of statistical significance of $p = .05$, Bag Size in knives being the only exception ($p < .01$). These findings are very surprising since we know from visual search research literature [10] that detection performance should decrease with growing set size. If Bag Size, the size of the area to search in, does not correlate with d' we would expect that Clutter, the amount of distractors in the set, does. Table 1 shows both the bivariate and partial correlations of the human factors with d' in order to allow direct comparisons. There are only very slight changes between bivariate and partial correlations with Training, but note that the signs of the correlations with Age all change from positive to negative when calculating partial correlation. The bivariate correlations reveal that the participants improve their detection performance with increasing age. This finding contradicts earlier studies on visual search tasks that revealed a deterioration of performance with age [11]. The partial correlations put this in perspective: Participants compensated age with more training, and indeed when controlling for training there is a small negative correlation between age and detection performance.

A very interesting aspect of this analysis is the comparison of the correlations' patterns for the different threat categories. The data reveal that different factors are important for being able to identify FTIs belonging to different threat categories. Figure 2 illustrates impressively how View Difficulty has a similar amount of influence on detection performance of guns and knives, but a much smaller amount on IEDs and a notably larger one on Other. Interpreting the pattern of Other is very difficult because Other is just the rest category for all the threat objects that do not fit into any of the other three categories. Thus the category Other includes as diverse objects as throwing stars (shuriken), tasers, hand grenades or gas tanks. Even though the IED stimulus set contains all sorts of hand made bombs, the stimuli are still comparatively homogenous making an interpretation of the image based factors much simpler than with Other: IEDs are generally made up of multiple essential parts (explosive material, fuse, cables, energy source, timer, etc.). Each of these has its own rotation (View) and its own Superposition value. Therefore it is not too surprising that the effects of View Difficulty and Superposition are highly diminished. A very interesting complementary finding is that while for IEDs image based factors in general show the lowest impact on detection performance compared to the other threat categories, human factors - especially Training - show the strongest effects on d' with IEDs.

4.2 Multiple Linear Regression Models

As already anticipated in the Results section we are very happy to report the achieved explained variances of nearly 70% for both the image based factors regression model as well as for the human factors regression model. The fact that we are able to explain such a big portion of variance from two distinct sets of predictors independently makes us very confident to get a grip on the process of X-ray threat detection. We are very confident that the image based factors together with human factors constitute the key aspects to cover for a better general understanding of the cognitive processes involved in this kind of visual search task. Nevertheless we still see some potential to further augment our model fits. This applies to the image based factors model as well as to the human factors model. Particularly with regards to the implementation of Clutter we see great potential to elicit larger predictive power - though to date we have not yet found a

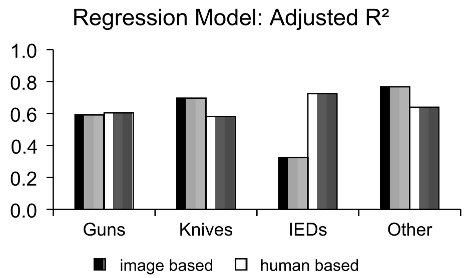


Fig. 8. Comparison of adjusted R^2 values of image based factor and human factor linear regression models for each threat item category

mathematical formula that rendered better results than the one currently in use. As for human factors: Until now we have merely investigated Training and Age. We see great potential in enlarging the set of human factors. Besides the undoubtedly important factor Training we expect visual abilities such as mental rotation, figure-ground segregation or visual search for highly specific patterns to be another important factor that should not be disregarded in a comprehensive model. Unfortunately for this study appropriate data were not available. As another human factor Gender should also be taken into account in future research.

Regarding the differences in terms of the explained variances and how they are made up of in terms of the beta weights among categories some remarkable notes must be made. The first pattern to catch one's eye is the different behavior of IEDs compared to the other threat categories. In the case of IEDs the image based factors' model regarding d' shows a comparatively low amount of explained variance. For all other categories the image based factors have a very high predictive power. A closer look at the beta weights and correlations reveals that it is particularly the effect of FTI View Difficulty which lies far below what would be expected based on the results from the other threat categories. The complementary finding is that compared to the other categories IEDs show the largest amount of explained variance with the human factors model. The beta weights and correlations reveal that this can overwhelmingly be attributed to Training. We can conclude that IEDs depend largely on human factors, especially on Training. We assume that detection performance with IEDs depends on knowledge as opposed to visual abilities to a larger extent than is the case with the other categories. In general the human factor models show much lower differences in predictive power among threat categories than the image based factor models do. We assume that the explanation for the comparatively large variation of the overall predictive power of the image based models as well as the comparatively large amount of variation between the correlations of the individual image based factors with d' lies in detection performance being based on several distinct cognitive processes dealing with the different image based factors, with their relative importance varying between categories.

4.3 ANCOVA/ Interactions

There are some essential differences concerning data types. The factor View Difficulty as it is operationalized for the ANCOVA cannot be directly compared with the variable FTI View Difficulty used in the correlation analyses and the regression models. FTI View Difficulty is a proportionally scaled measure derived statistically from performance data whereas the ANCOVA factor View Difficulty is a dichotomous nominal variable differentiating between easy and difficult views only. As a matter of fact, all image based factors - except of course for the factor 'Threat Category' - are dichotomous in the ANCOVA (refer to the paragraph on stimuli in the Methods section). Bag Complexity replaces the compound of Opacity and Clutter (refer to the Stimuli section).

ANCOVA with Category Treated as a Repeated Measures Factor. The main effects of the ANCOVA on all categories give a similar picture as did the correlations. In analogy to what we found in the correlation analyses, effects of Superposition are larger than effects of Bag Complexity and Bag Size. Furthermore the results show that there are considerable interactions with Training regarding the different threat categories. For example, as discussed above the effect of Training on detection performance is clearly larger in IEDs than in knives.

View Difficulty and Superposition also show interaction effects with Training. The improvement of the detection performance caused by training is clearly larger regarding the difficult views compared to the easy ones. The interaction between Superposition and Training on the other hand is fairly small. This could indicate that dealing with superposition is difficult to improve with training. For Bag Complexity and Bag Size interactions with Training are not significant. No evidence could be provided for interaction effects of any of our image based factors (including threat category) with Age.

In the following itemization we give a short overview of the four largest reported first order interactions and plausible interpretations with examples.

VD * Cat:	reflects the fact that effects of View Difficulty differ between Threat Categories e.g.: compare correlations of View Difficulty and d' between IEDs and Other
VD * SP:	difficult views are more affected by high Superpositions that easy ones and vice versa
BC * Cat:	reflects the fact that effects of Bag Complexity differ between Threat Categories
SP * Cat:	reflects the fact that effects of Superposition differ between Threat Categories

ANCOVAs by Category. Performing the ANCOVA individually per category makes several interesting effects visible. According to ANCOVA and in correspondence with the results of the correlation analysis, detection performance with guns depends on View Difficulty and Superposition. The interaction of View Difficulty with Training indicates that for guns training is particularly effective with difficult views. If we compare the relative effect size of View Difficulty in the ANCOVA with the relative effect size (R^2) of FTI-View-Difficulty in the correlation analysis we can see that the effect of View Difficulty in the ANCOVA is comparatively smaller. Since FTI-View-Difficulty also measures effects of properties of the threat item itself - i.e. the specific gun - we can conclude that for guns it is not only the View Difficulty that is important in mediating detection performance but also the kind of gun in question. For Knives View Difficulty, Superposition and Bag Complexity have large main effects. In contrast to the results for Guns, View Difficulty does not have a comparatively smaller effect size in the ANCOVA than in the correlation analysis, indicating that for Knives View Difficulty is indeed important in mediating detection performance, whereas the type of knife is not. Furthermore, for Knives the interaction between Training and View Difficulty does not reach statistical significance. In other words: dealing with View Difficulty in Knives is not improved through training in Knives. There is however a significant interaction between Superposition and Training. Finally, Knives are the only category with a (barely) significant main effect of Bag Size. The ANCOVA results for IEDs confirm the negligible influence of View Difficulty on detection performance for this threat category. Performance is mediated by Superposition and Bag Complexity alone. There is a small but statistically significant interaction between Bag Complexity and Age with IEDs. The ANCOVA for the category Other shows only two significant and still small main effects, namely the effects of View Difficulty and Bag Complexity. However there is a large interaction between View Difficulty and Training. The ANCOVA results for Other are in stark contrast to the results of the correlation analyses - especially with regards to Superposition. Superposition had a large effect size (R^2) in the correlation analysis whereas in the ANCOVA the main effect of Superposition does not even reach statistical significance. As has been mentioned above, the category Other is very heterogeneous which makes any interpretation of these seemingly contradictory results very difficult and speculative.

5 Conclusions and Recommendations for Improving Human-Machine Interaction in X-Ray Screening

5.1 FTI View Difficulty and Superposition

To date, X-ray screening technology delivering only one image per bag have been common in aviation security. More recent technology is capable of providing multiple views of a bag. Our research has shown that the image based factor View Difficulty can be addressed effectively with computer based training. The ANCOVA analysis has supported earlier findings showing that training improves detection performance particularly for difficult views [12]. The results for Superposition were not as promising, however. Figure 9 illustrates how new multi-view systems might be able to reduce the detection problems due to View Difficulty and Superposition. Objects that are very much superimposed by other objects from one perspective may be clearly visible from another.

5.2 Opacity

Opacity refers to the extent to which a bag is penetrated by X-rays. X-ray systems with higher penetration have the potential to reduce detection problems due to Opacity. Even

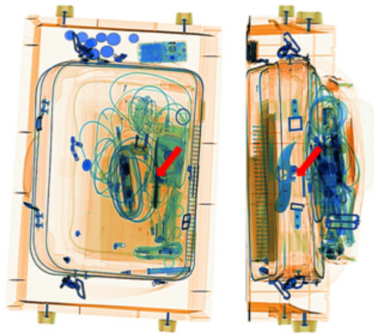


Fig. 9. Illustrative example of how multi-view systems can help improving detection performance in spite of undesirable View Difficulty and Superposition effects

so, the problem of dealing with baggage that render too opaque images remains. A solution is the implementation of a 'dark alarm' in X-ray equipments. This dark alarm would warn the human operator if a maximum amount of opaque areas in a bag were exceeded. Manual search would follow when a dark alarm was indicated.

5.3 Screener Selection and Training

A very important approach in improving X-ray screening efficiency and increasing threat detection performance is screener selection and screener training. Psychological literature provides evidence that figure-ground segregation (related to Superposition) as well as mental rotation (related to View Difficulty) are visual abilities which are fairly stable within a person. Hofer, Hardmeier, and Schwaninger [13] and Hardmeier, Hofer, and Schwaninger [9] have shown that pre-employment assessment procedures based on computer based object recognition tests can help increase detection performance substantially.

The present study shows, that in addition to stable abilities there are also trainable skills that play a very important role in mediating detection performance. Knowledge based factors such as knowing which objects are prohibited and what they look like in X-ray images are learned by screeners in the computer based training. Computer-based training can be a powerful tool to improve X-ray image interpretation competency of screeners [12,14,15].

The list of human factors analysed in this study is far from being exhaustive. Additional studies investigating human and contextual factors such as vigilance, stress, heat and time pressure are in preparation but would go far beyond the scope of the study at this point.

5.4 Educational and Usability Aspects

Despite massive advances in the development of technological equipment it is still the human operator who decides whether a bag can enter an aircraft or not. As can be seen from the results regarding the importance of X-ray image interpretation training one must not forget that all technological equipment is of limited value if its operators are not trained thoroughly in their task. Therefore it is very important for engineers to keep in mind the human operators behind the technological equipments. The present study gives a series of key aspects that must be taken into account when developing training systems such as the image based factors introduced. On the other hand some of these aspects can indeed be directly addressed by technological means. For example the use of multi-view systems for dealing with view difficulty and superposition.

Acknowledgments

This research was supported by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/F/TH-80403). Special thanks go to the two airports that supported this study by supplying screeners and data.

References

1. Annual review of civil aviation 2005. International Civil Aviation Organization 61, 9 (2006)
2. Schwaninger, A., Michel, S., Bolfing, A.: A statistical approach for image difficulty estimationin x-ray screening using image measurements. In: Proceedings of the 4th Symposiumon Applied Perception in Graphics and Visualization, pp. 123–130. ACM Press, New York (2007)
3. Schwaninger, A.: Evaluation and selection of airport security screeners. Airport 2, 14–15 (2003b)
4. Schwaninger, A., Hardmeier, D., Hofer, F.: Aviation security screeners visual abilities & visual knowledge measurement. IEEE Aerospace and Electronic Systems 20(6), 29–35 (2005)
5. Riegelnig, J., Schwaninger, A.: The influence of age and gender on detection performance and the criterion in x-ray screening. In: Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia, Montenegro, June 24-28,2006, pp. 403–407 (2006)
6. Bolfing, A., Schwaninger, A.: Measurement formulae for image-based factors in x-ray imagery. In: VICOREG Technical Report, November 26 (2007)
7. Green, D.M., Swets, J.A.: Signal detection theory and psychophysics, pp. 187–194. Wiley, New York (1966)
8. Coolican, H.: Research Methods and Statistics in Psychology, 4th edn., Hooder & Stoughton, London (2004)
9. Hardmeier, D., Hofer, F., Schwaninger, A.: Increased detection performance in airport security screening using the x-ray ort as pre-employment assessment tool. In: Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAT 2006, Belgrade, Serbia, Montenegro, June 24-28, pp. 393–397 (2006b)
10. Wolfe, J.: Guided search 2.0 a revised model of visual search. Psychonomic Bulletin & Review 1(2), 202–238 (1994)
11. Madden, D.J., Gottlob, L.R., Allen, P.A.: Adult age differences in visual search accuracy: Attentional guidance and target detectability. In: Psychology and Aging, pp. 683–694 (1999)
12. Koller, S., Hardmeier, D., Hofer, F., Schwaninger, A.: Investigating training, transfer, and viewpoint effects resulting from recurrent cbt of x-ray image interpretation. Journal of Transportation Security (2008)

13. Hofer, F., Hardmeier, D., Schwaninger, A.: Increasing airport security using the x-ray ort as effective pre-employment assessment tool. In: Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA, November 27 - December 1, pp. 303–308 (2006)
14. Schwaninger, A., Hofer, F., Wetter, O.E.: Adaptive computer-based training increases on the job performance of x-ray screeners. In: Proceedings of the 41st Carnahan Conference on Security Technology, Ottawa, October 8-11 (2007)
15. Ghylin, K.M., Drury, C.G., Schwaninger, A.: Two-component model of security inspection: application and findings. In: 16th World Congress of Ergonomics, IEA 2006, Maastricht, The Netherlands, July, pp. 10–14 (2006)

A Process for Human Centered Modelling of Incident Scenarios

Claudia V.S. Guerrero[1], M. Fatima Q. Vieira[2], Jean-Marc Mercantini[1], and Charles Santoni[1]

[1] Laboratoire des Sciences de l'Information et des Systèmes, Université Paul Cézanne – Aix-Marseille III, Av. Escadrille Normandie-Niemen 13397 Marseille CEDEX 20 – France
{claudia.serey,jean-marc.mercantini,charles.santoni}@lsis.org
[2] Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, CP 10053 CEP 58.109-970 Campina Grande, PB, Brazil
fatima@dee.ufcg.edu.br

Abstract. This paper presents the process of building a conceptual model of incident scenarios caused by human error. The process was conceived to be human centered and was validated through a case study in the context of the electricity generation sector. The corpus of study was extracted from an electricity distribution and transmission company in Brazil. The resulting conceptual model forms the basis of an incident scenario simulator built to study user behavior when dealing with safety critical situations. This work is within the context of conceiving ergonomic human interfaces for industrial systems. The objective is to conceive interfaces that facilitate the operation of these systems, particularly during critical situations when the human error is considered one of the main causes of accidents.

Keywords: Cognitive science, control system human factors, error analysis, user interface human factors, user centered design.

1 Introduction

From the viewpoint of operation safety, the higher degree of perfection and performance of industrial machinery and tools has transferred to the human operator the responsibility for almost every failure that happens during an interaction between them. This is the result of differences in work pace and in representation languages, which lead to misunderstandings that are responsible for the majority of the accidents reported in this sector. On the other hand, the operators risk increasing their fatigue and stress levels when attempting to manage their own abilities and error rate [2].

According to studies on the impact of error in complex man-machine systems, there is a general consensus that the majority of the serious accidents are due to human errors [25], [26], [22]. According to Johnson [9] 80% of aviation incidents are caused by human errors. This problem reaches higher proportions in industrial applications supported by complex systems that are considered safety critical from the viewpoint of the consequences of errors and faults. Beyond precision and functionality, it is essential for the

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 439–458, 2008.

users of these systems to have built in safety, adaptability to different degrees of expertise and work situations, and the support for easy learning.

The research context of the Human Interface Group (GIHM) at the UFCG is to improve the quality of the interaction between operators and systems. This is to be achieved through the conception of more ergonomic interfaces to support this interaction during normal operation and particularly under critical situations. A research currently under development at the GIHM aims to adapt the interface conception and evaluation method - MCIE [24] to the context of critical industrial systems. Throughout this research, models, techniques and tools have been integrated to MCIE. However, it does not yet account for human error, needing information on the situations where these occur and on the user behaviour under those circumstances.

Typically, to consider the user during the conception of user interfaces means to account for information obtained from the ergonomic analysis of the work, such as: age, gender, specific knowledge of the task and on the strategies to perform it and also on information linked to the user's cognitive abilities. In order to do so, this work proposes to integrate a new set of information into the method MCIE, along with the knowledge on the task, the user profile and the context of use - in other words, the knowledge on situations which lead into accidents and on the user behavior under those circumstances. The objective is to conceive user interfaces that are more ergonomic and thus contribute to reduce accidents due to human error. To reach this objective, it is proposed to integrate a behavioral model of human operators, faced with critical situations, into the method MCIE. To obtain this behavioral model, the approach proposed is to place the operator in working critical situations by means of an incident scenario simulator.

The purpose of this paper is to present the elaboration process of a conceptual model of accident scenarios. In essence it presents how to identify, extract and represent the knowledge about accidents caused by human errors during the interaction with safety-critical systems. This model will be used to build an incident scenario simulator with which it will be possible to study the user behavior when dealing with safety critical situations. The study will consist in recording the operator's behaviour as well as the reasoning process in order to build the behavioural model based on this knowledge.

The current work lies in the following research fields: Accidents (the terms accident and incident will be used indistinctively throughout the text) and Human error modelling, Human behaviour under critical situations and Human Interface Design. In respect to human error modelling this work concerns the representation of incident scenarios that involve human errors. Along the scenarios of interest the situations are considered critical because they demand from the operator an adequate and fast response to avoid the escalation of gravity.

This paper is organized as follows. Section 2 presents a bibliographical review on accident analysis and modelling. Section 3 presents the process conceived to build the Incident Scenario Conceptual Model - MCCA. In this, each process phase is described by its objective, inputs and outputs, and exemplified through a case study related to the operating sector of an electricity company. Section 4 discusses the process resulting model and how it will be used by the MCIE for user interface conception and when building the incident scenario simulator. Section 5 presents the conclusions and proposes directions for future work.

2 Analysis and Modelling of Errors and Incidents

The human error concept has many definitions. Reason [19] characterizes the operator error through the intention of the result. Thus, the error exists if, after a planned sequence of mental or physical activities, the desired objective will not be reached. According to Keyser [10] errors only can occur when the operator has certain degrees of freedom. If the situation does not allow choices, there will not be the possibility of error, or at least, the error responsibility will not be the operator's. Among these definitions a common idea is that the human error [11], [16], [18], [25] is a malfunction of the human operator (mental, psychomotor, sensorial or physical activities) expressing an abnormal shunting line in relation to a norm or an established standard. For many authors, the error is not seen isolated but as a relation between operator, environment and task.

The main motivation for the study of human error has been the need to improve system safety by reducing or even eliminating accidents and their consequences. Geared by this objective, many studies have been published that try to identify the causes of accidents. The concern about human errors has raised research interests and concentrated efforts in approaches such as error management, building taxonomies of error types and applying theories of cognitive processing. These approaches differ as they concentrate on different elements associated to the error: the operator, the user interface, the organization structure and the installations. Nonetheless, they all share interest in finding the causes for the human error in order to mitigate it.

Amidst the error-mitigating approaches there are those, which are based on error report analysis, such as the one presented in this paper, but with clear distinct purposes. These tend to offer support to the incident analysis process, providing information systems (databases). One example of an error mitigating approach is the British Airways Human Factors Reporting Program (HFR) [17] which records and analyses data on "what", "why" and "how" the crew dealt with a problem. Also based on report analysis, the analytic process (reading and interpreting reports) aims to completely understand the sequence of cause and effect throughout an event, in order to get a clearer picture of the underlying causal factors.

In contrast, the human factors studies tend to concentrate on the human behavior under normal situations - such as the study conducted by IRSN, and reported in [4]. In this study the analysis method consisted of the systematic and structured reading of incident reports. Despite the similarity with the incident analysis approach, the objective of that work is the storage of pertinent data into a database which will help to highlight the relation between: incident causes, recovery factors and mechanisms related to the progression of the incident. Amongst the many factors considered in the RECUPERARE model there is the quality of the human interface, which is in contrast is the focus of this work presented here.

According to O'Leary [17], system designers use incident analysis data to examine how their design works in the real world. However, he calls the attention to the incident analysis impact on the apparent system reliability of operator's problem solving and adaptability to improve the design of equipment and procedures. It is also the authors' understanding that, although the design of human interfaces acknowledges the importance of human error, the application of error models into the interface design method and its impact on the quality of the interaction is still to be investigated.

Svedung & Rasmussen [23] present the AcciMapping technique that is a set of graphic representations of accidents to help analysing the structure of hazardous work systems and to identify the interactions within a socio-technical system. Those shape an overview of how accidents can occur. Similarly to our works, these authors analyse a set of accidents to obtain a generic representation of accidents. Whereas we look for scenarios-types their objective is to understand the relationship between the different levels of the system and the decision makers in order to understand the causes of the accidents. According to these authors, accidents result from decisions and actions which originate from all levels, not being restricted to the process controllers. Unlike us, they are interested in the causes of accidents that are investigated with the help of a tool developed to support the communication between the accident analysis experts. In contrast, this work does not aim to investigate accident causes but to identify and represent accident scenario types.

Other error mitigating approaches are based on the knowledge of the task and its context such as those proposed by [14], [27]. Vergara analyses and models the task and the user, accounting for the potential errors and related accidents. The analysis is performed through the direct observation of the user in its working environment and the knowledge extraction is based upon the verbalization technique.

In this research, as in [1], [13], [14], the knowledge extraction and representation method KOD (Knowledge Oriented Design) [28] was adopted. As in these, the process is based on the analysis of accident reports and other textual documents, a corpus for which KOD has shown to be adequate and efficient. In [1] and [13] the objective was to build systems to support the process of diagnosing accident causes through knowledge extraction from structured databases on accident reports. However, in [14] the objective was to simulate accidents in order to support the proposal of preventive safety measures in industrial sites located in urban areas. Regardless of their specific differences, all those studies have one common objective - to raise operating safety levels, and to reduce errors and the related accidents.

Although some authors cite the user interface as a potential source of problems which can lead into human error [5], [11], these works do not show how to consider it nor how to include the human error perspective into the user interface conception. In [3], techniques are used to analyse incidents and accidents and the results are fed into the system model aiming to conceive safer interfaces for critical systems, thus making these more robust. Initially the authors model the system using Petri nets and ICO and proceed to consider the events identified as dangerous as the result of the incident analysis using Goal Structuring Notation (GSN). Through a simulation of an accident scenario the authors illustrate that the simulated event cannot happen since it had been identified and blocked.

It is this work's premise that the human error must be considered as a factor during the design of safer industrial systems. It is important to realize that the human error cannot be eradicated. Designers must admit its existence and anticipate it in order to prepare for it and to avoid it if possible. The interface design must account for the human error, embedding the knowledge on when the operator is more likely to commit an error and thus preparing for it, avoiding or eliminating all those situations.

An accident rarely will be explained through a single cause. To understand it, it is necessary to consider the context in which it occurred. To understand the causes of an accident it is necessary to understand how it happened [9]. This work's objective is not to investigate who is responsible for the error but to understand the context on which it occurs in order to remove the interface ergonomic flaws.

3 The MCCA Building Process

The MCCA conceptual model describes and represents the main situations that lead an operator into errors during the interaction with the industrial system. The MCCA described here is specific to the case study. These errors are considered accident triggers.

Following is the description of process conceived to build the MCCA, with its phases and the resulting products, or artifacts. This process is composed of five phases: (1) Corpus definition, (2) Knowledge extraction, (3) Knowledge analysis and abstraction (ontology construction), (4) Ontology validation, (5) Scenario-types identification. Figure 1 illustrates the method and its phases and products. In this figure the phases are represented as grey ellipses, the related products as rectangles and the resources needed for each phase are represented as traced ellipses. The highlighted rectangles represent the two main products that compose the MCCA:

1. The validated ontology consists of a common and coherent vocabulary with terms and actions related to accident situations. It is organized in such a way as to be meaningful to system designers and users. It must also be representative of the error situations that happen during operator interaction with the system. This ontology also includes the operator errors classification.

2. A set of Scenarios-Types. A scenario-type is a generalization that corresponds to a group of accidents that present similarities in specific aspects [4]. It links an accident to conditions that led to his emergence. The underlying hypothesis is that incident scenario-types have effects upon operator's behavior. The analysis of those situations will enable us to determine the nature of the existing relations between: the operator's behavior types, the user interface characteristics, the state of the industrial installation and the critical context. In this sense, the simulator will be conceived on the basis of these incident scenario-types in order to favor the emergence of typical factors associated to the operator's behavior.

The MCCA building process is based on the KOD method, which belongs to the family of Knowledge Engineering (KE) methods. KOD is based on an inductive approach, which requires to explicitly express the cognitive model (also known as the conceptual model) based on a corpus of documents, comments and experts' statements.

The main features of this method are based on linguistics and anthropological principles. Its linguistics basis makes it well suited for the acquisition of knowledge expressed in natural language. Thus, it proposes a methodological framework to guide the collection of terms and to organize these based on a terminological analysis (linguistic capacity). Through its anthropological basis, KOD provides a methodological framework, facilitating the semantic analysis of the terminology used to produce a cognitive model (conceptualisation capacity). It guides the work of the knowledge engineer from the extraction of knowledge to the development of the conceptual model.

The employment of the KOD method is based on the conception of three types of successive models: the practical models, the cognitive model and the software model, as represented in Table 1. Each of these models is conceived according to the paradigms: <Representation, Action, Interpretation>. The Representation paradigm allows to model the universe such as an expert represents it. This universe is made of related concrete or abstract objects. The Action paradigm allows modelling the behaviour of active objects that activate procedures upon the receipt of messages. Therefore, action plans devised by human operators as well as by artificial operators will be modelled in the same format. The Interpretation / Intention paradigm allows modelling the reasoning employed by the experts in order to interpret situations and to elaborate action plans related to their intentions (reasoning capacity).

The practical model (PMi) is the representation of a speech or a document belonging to the corpus, that is expressed in the terms of the domain by means of taxemes (static representation of objects), actemes (object activity representation) and inferences (at the basis of the task cognitive structure). The cognitive model is built by abstracting the practical models. It is composed of taxonomies, actinomies and reasoning patterns. The software model result from the formalization of a cognitive model expressed in a formal language, and is independent of programming languages.

Table 1. KOD, the three modelling levels

Paradigms / Models	**Representation**	**Action**	**Interpretation**
Practical	Taxeme	Acteme	Inferences
Cognitive	Taxinomy	Actinomy	Reasoning Pattern
Software	Classes	Methods	Rules

To illustrate the MCCA building process there follows a description of its application to the case study developed for an industry of the electricity sector. This case study was considered to be particularly interesting due to the nature of the system and the availability of a corpus of reports on accidents. The accident reports are a legal imposition from the Brazilian industry's regulations.

As mentioned previously, this MCCA will support the building of an accident simulator. The purpose of the simulator is to serve as a training tool for the company's operators and, from the research point of view, to support the observation and the recording of the user behavioral model. For this case study, the user observation could not be performed in the work environment due to safety reasons and also because an observation in loco would have interfered with the user behavior and with the task. The user behavior model will be integrated into the user interface conception method MCIE.

3.1 Corpus Definition

This phase's objectives are to identify, within the problem domain, the relevant knowledge for the MCCA. It requires a well-defined and well-delimited problem domain.

In the case study, the electricity company made available 386 accident reports corresponding to the period 1998 to 2005. Since this study's interest is focused on human

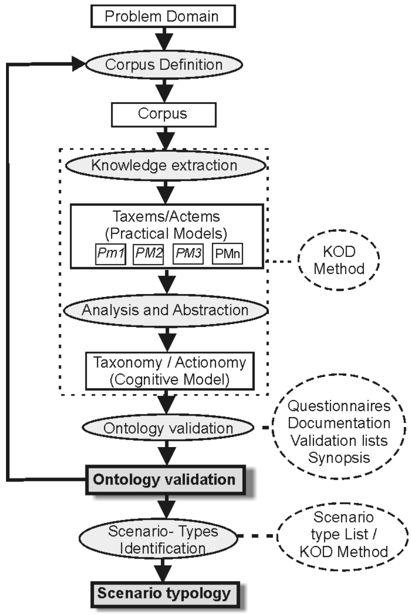


Fig. 1. MCCA building process

errors during system operation, it was adopted the following set of criteria was adopted, originally proposed in [15]: a) Reports of accidents related to the distribution and transmission of electricity; b) Reports of accidents that happened during routine procedures performed during system supervision and control; c) Reports of accidents related to human errors that occurred during operator interaction with the supervision and control system; d) Reports where the human error initiated the accident (considered the main cause); e) Reports of accidents that happened in the control room or in the patio where the equipment is installed.

After applying the above criteria the corpus of study incorporated: a) 21 Accident Reports – of which, according to the company's classification, eighteen were related to human errors and three, were related to operating failures; b) Company's operating manuals: a set of documents that prescribe the operator's tasks, and describe the related equipment and devices; c) Operator interviews: registry of interviews with the operators to clarify the contents of the accident reports; d) Technical visits to the company's installations (substations) in order to get acquainted with the equipment, devices and work environment terminology.

3.2 Knowledge Extraction

This phase consists in extracting from each accident report belonging to the corpus, all the elements (objects, actions, and concepts) that are relevant to the accident scenario representation. It is important to notice that knowledge concerning the interpretation paradigm was not available on the analysed reports, so, this dimension will not be treated on this study. The lack of this knowledge is the main reason to develop the incident scenario simulator.

Extracting taxems. To obtain the taxems, the linguistic analysis is performed in two steps: the verbalization and the modelling into taxems. The verbalization step consists in paraphrasing the corpus documents in order to obtain simple phrases, which allow qualification of the terms employed during document analysis. Thus, some terms appear as objects, others appear as properties, and yet others appear as relations between objects and values. The modelling step consists of representing the phrases in the format of taxem: <object, attribute, value>.

The taxem characterizes an object from the real world by means of a relation (attribute), which links the object to a value. There are five types of relations: classifying (is-a, type-of), identifying (is), descriptive (position, failure mode, error mode, cause…), structural (composed-of) and situational (is-in, is-below, is-above…).

The example that follows illustrates the process employed to obtain the taxems from one phrase extracted from the report in the case study.

"...switch breaker 14T1 unduly opened in place of switch breaker 14D1..."

Paraphrases

14T1 is a Switch breaker

14T1 unduly opened in place of 14D1

Taxems

<14T1, is a, Switch breaker>

<Open 14T1 in place of 14D1, is a, Action on wrong object>

The extent of this analysis on the set of documents, associated to operators interviews and guided tours, have allowed obtaining the set of taxems needed for the representation of incident scenarios. As example, it follows an extract of the set of taxems which model the Switch CLT:

<Switch CLT, is a type of, Switch>

<Switch CLT, is composed of, switch handle>

<Switch CLT, failure modes, switch handle broken >

<Switch CLT-14D1, is a, Switch CLT>

Extracting actems. In order to obtain the actems, the linguistic analysis consists on identifying the verbs which represent the activities performed by the operator during an incident. In general terms, an activity is performed by an action manager, by means of one or more instruments, in order to modify the state (physical or knowledge) of an addressee who temporarily takes control by means of the instruments. The action manager can also be the addressee. Occasionally the action manager can be the one who directs the activity and at the same time is also subjected to the change of state (knowledge acquisition). The following example illustrates how to extract the actems from two phrases taken from the reports in the Corpus.

"...placed the switch CLT- 14D1in the LOC position...

…placed the switches CLT-14V1 and CLT-14D1 in the TEL position..."

Paraphrase

Place Switch CLT-14D1 is equivalent to toggle Switch CLT 14D1.

Once identified, the actems are translated into a 7-tuple: <Action Manager, Action, Addressee, Properties, State1, State2, Instruments>, where: the Action Manager performs the action; the Action causes the change; the Addressee undergoes the action (this could also be the Action Manager); the Properties represent the way the action is performed; State 1 is the state of the addressee before the change; State 2 is the state of the addressee after the change; Instruments, is one or a set of instruments representing the means used to cause the change.

<*Control room Operator, Toggle, Switch CLT-14D1, missing information, position TEL, position LOC, hands and switch handle of Switch CLT-14D1*>. Figure 2 illustrates an actem from the model - 'TO TOGGLE'.

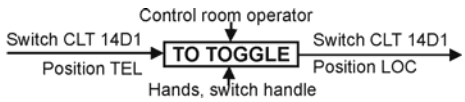


Fig. 2. Representation of the Actem 'TO TOGGLE'

Actems model the task activity. In this case, the actem is composed of textual items extracted from the reports, which describe the change of state of an object or concept as used by the domain experts. Each actem is also represented as shown in Table 2.

Table 2. Actem "To Toggle"

TO TOGGLE SWITCH = TO POSITION = TO PLACE	
Components	**Values**
Action Performer	[auxiliary operator, control room operator]
Recipient	
State 1 (recipient) Position (Switch CLT or LOC TEL)	[position LOC; position TEL]
State 2 (recipient) Position (Switch CLT or LOC TEL)	[position LOC; position TEL]
Instruments	[Hands, switch handle, telecommand function]

3.3 Analysis and Abstraction

This phase consists on the analysis and abstraction of all the Practical Models built in the previous phase. The objective is to build the domain ontology in the context domain. In other words, the aim is to classify the used terminology and thus obtain the KOD Cognitive Model.

Building the Taxonomies

Term Analysis: The terms are analyzed, with homonyms and synonyms, thus building a common terminology for the context of the case study.

Concept Identification: This step is based on the analysis of the taxems and consists in highlighting the nature of the attributes, which characterize each object. The attribute nature is the basis for the construction of the taxonomies (relations 'type-of' and 'is-a') or other tree type structures (relations: 'is, composed-of', 'position', 'failure mode', 'is-in', 'is-below', 'is-above', etc.). Analyzing the taxems list it was found that the term "switch" is meaningful and thus it deserves the status of a concept. The same happens to the terms: Action Error and Action on wrong object. On the other hand, each component of the Action on wrong object is considered an instance.

<Action error, type of, Operator Error> <Sequence error, type of, Action Error>
<Goal error, type of, Action Error> <Amplitude error, type of, Action Error>
<Destination error, type of, Action Error>

<Action on wrong object, type of, Destination Error>
<to open switch breaker 14T1 instead of 14D1, is a, Action on wrong object>

Each concept represents an Action error category (sequence, goal, amplitude, destination) belonging to a higher-level category in the hierarchy. The term Operator Error will be equally considered as a concept whereas the attribute 'type' links concepts. This is a relation (type-of), which allows building the taxonomies. Figure 3 illustrates part of the operator error taxonomy. Each concept (represented underlined on the tree) must be defined, as it is exemplified below for the concepts 'Action on the wrong object'.

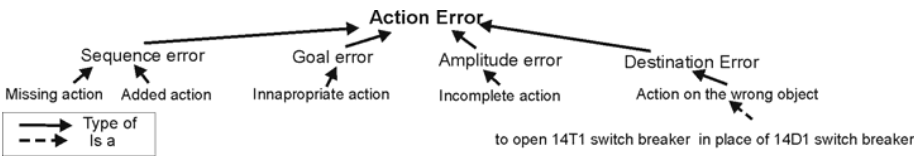


Fig. 3. Action Error Taxonomy

The concepts are defined through their attributes or relations.

Name: Action on the wrong object
Reference: to open the switch breaker 14T1 instead of the 14D1 switch breaker;
Caused by: {inattention; stress, installation configuration}
Location: {Patio, SE Control room}
Caused Incident: {Supply cut off; Wrong configuration}
Executed Task: {To release equipment; to restore equipment}
Executed Task attribute: {simple; complex, usual, rare}

Building the Actionomies. In this project, the *actionomies* are the result of the *actem* organization, according to incident scenarios. Thus, an actionomy must contain the actors, the action sequence, the instruments used to perform each action, and the initial and final state. Each incident has been modeled by means of an actionomy. To illustrate the building process of an actionomy there follows the extract of a textual description of an incident relative to one report taken from the corpus.

"During the maneuver procedure, related to the normalization of the switch break 14V1 at the SE NTD (Substation Natal II), having finished the preventive maintenance, the switch break 14T1 was unduly opened (when it should have been opened the switch break 14D1), keeping the transformer energized in open (no load)…"

The context of this incident scenario specifies that the operator was in a typical situation performing programmed and routine maneuvers, when performing a prescribed action acts on the wrong device. The corresponding actionomy is:

"NTD Receive 14V1 free to be operated on CROL Authorize NTD to normalize 14V1
NTD deactivate surge schema related to LT 04V3 CGD/NTD
NTD place switch CLT-14D1 in the LOC position
NTD close 34V1-4 and 34V1-5
NTD place switch 43-14V1 in the ET position

NTD close 14V1
NTD Open 14T1 must have been open 14D
NTD Place switch 43-14V1 in the position "N"
NTD open 34V1-6
NTD place switches CLT-14V1 and CLT-14D1 in the TEL position

Figure 4 illustrates the graphical representation for the actionomy described above.

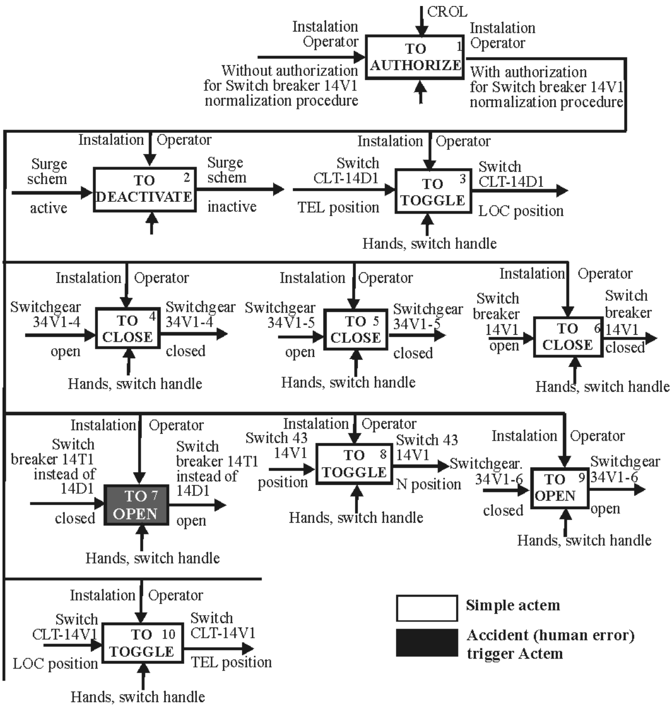


Fig. 4. Actionomy extract

The representation of incident in the form of actionomies allows for performing a detailed analysis, which aims at proposing the scenario-types. The underlying hypothesis to this analysis is that the incident scenarios reveal the operator's typical behavior when piloting under critical situations. The actionomy also allows to focus on each incident, its attributes and context. This knowledge can be represented and organized in different taxonomies. The analysis of those situations will allow determining the nature of the existing relations between: the operator's behavior types, the user interface characteristics; the state of the industrial installation and the critical aspects of the context. Therefore, the simulator will be conceived on the basis of the identified scenario-types, to expose the typical factors associated to the operator's behavior.

3.4 Ontology Validation

This phase consists on validating the terminology and the term classification. There are two aspects to be considered during this ontology validation: the correctness and the completeness.

Validating the Correctness. During the construction of the ontology some terms presented ambiguity (concepts, taxems and actems) in respect to definitions and classification. In some cases there were various definitions for the same term (homonyms) whereas in others there were many terms sharing the same definition (synonymous). Another aspect that motivated this validation was the knowledge that the reports in the corpus were written by engineers and psychologists based on interviews with the operators, adopting a terminology different from that used by the operators. It was intended to investigate if the ontology extracted from the corpus was well known and also used by the operators. The ontology would be considered correct if the terminology and its classification were correct, understood and also adopted by the company's operators.

Validating the Completeness. The conceptual model completeness is the property that ensures that every scenario that fulfils the requisites of the corpus of study can be represented by the elements represented in the model ontology (actems, taxems, concepts, taxonomies and actionomies). To validate the conceptual model completeness two approaches were used: (1) Representation of new scenarios which consisted on modelling scenarios outside the corpus of study (building actionomies); and (2) Operators' role playing during which the operators were given a scenario related to their installation, describing the first steps performed during a maneuver. The operators were then asked to complete it, in the basis of their previous knowledge of the objects and situation involved. The objective was to investigate how well the operators knew the ontology employed during the scenario description (actions, objects, and terms). A validation table was conceived to support the ontology validation process from the viewpoint of its completeness. During this process, when faced with a new term (actem or taxem), the person in charge would consult this table and decide if the new term was already represented in the ontology, and if not whether to include it or to ignore it.

The validation phase resulted in the insertion of new terms, and the correction and complementing of existing concepts. It became clear from the new accident cases that new terms can be added to the ontology without altering its structure.

3.5 Scenario-Type Identification

In this work, an "incident scenario" consists of a set of elements required to represent a specific incident case. Thus, an incident scenario is composed of: action sequence, actors, instruments, context elements, objectives, error, error causes, and error consequences. This definition was conceived based on the Go & Carroll scenario definition [7].

On the other hand an "incident scenario-type" consists of a generalization of a group of accidents with common characteristics. From the analysis of each incident scenario, it is possible to establish the relationship between the scenario elements and to identify the characteristics of an incident scenario-type. In this work an incident scenario type is represented by two components: one static and one dynamic. Those two representations must be compatible and also complimentary.

The static representation of an incident scenario-type is defined by the “scheme of incident scenario”. Figure 5 describes the generic scheme of incident scenarios. It is composed of the following elements: a) the task: objectives to be reached by the operator; b) The incident: consequences of operator’s errors; c) the error initiating elements (the internal and external error causes); d) the installation: different equipment and devices applied in the scenario; e) the context: elements which can modify, influence or characterize the meaning of the scenario components.

This scheme represents the most important elements and the relations between them. The “operator” concept encompasses concepts relative to the different kinds of operator’s errors (action error, cognition error, perception error), as well as their causes such as ‘Internal Error Cause’. Nonetheless, in Figure 5 the relation between those concepts is not represented, since this is not found, explicitly, in the analyzed corpus. It is not found either the relation between the concepts and the external causes. The missing information will be investigated as a hypothesis to be tested during the user interaction with the operator training simulator.

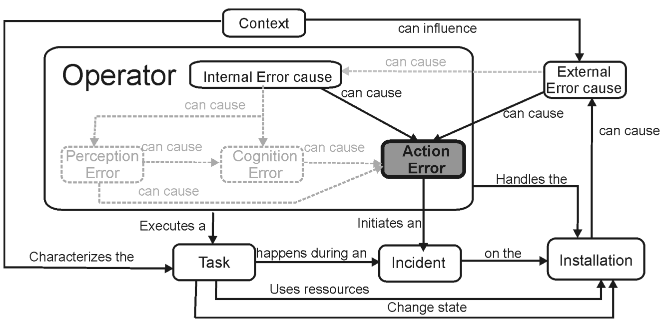


Fig. 5. Scenario type generic scheme

The context has an important role when describing the scenario type. As mentioned before, context elements have the characteristic of modifying and influencing the meaning of other elements. The schedule is an example of context element. If the operator task is being executed during a period of high demand of electricity, the task is no longer a routine task and acquires the status of emergency task raising the operator stress level (internal error causes). The working team composition, the environment, the operator knowledge about the risks and the communication (speech, dialogue) are other examples of context elements.

The scenario type dynamical representation is a chain of situations induced by the event of an action (an operator action on the installation, a system action, or an operator change of state). It is consists of an actionomy composed of: an action sequence, the actor, the instruments used to execute each action, and the instruments initial and final state.

In order to derive the scenario types many different approaches could have been employed. For instance, from the generic scheme it is possible to focus on each aspect and

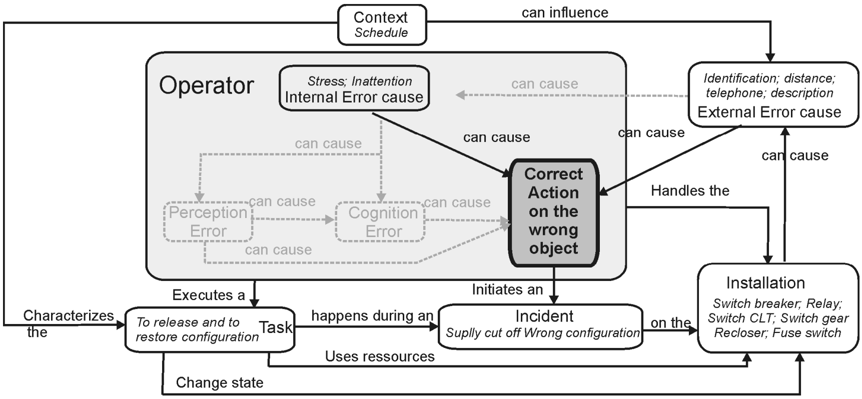


Fig. 6. Scenario Type 1: Correct Action on the Wrong Object

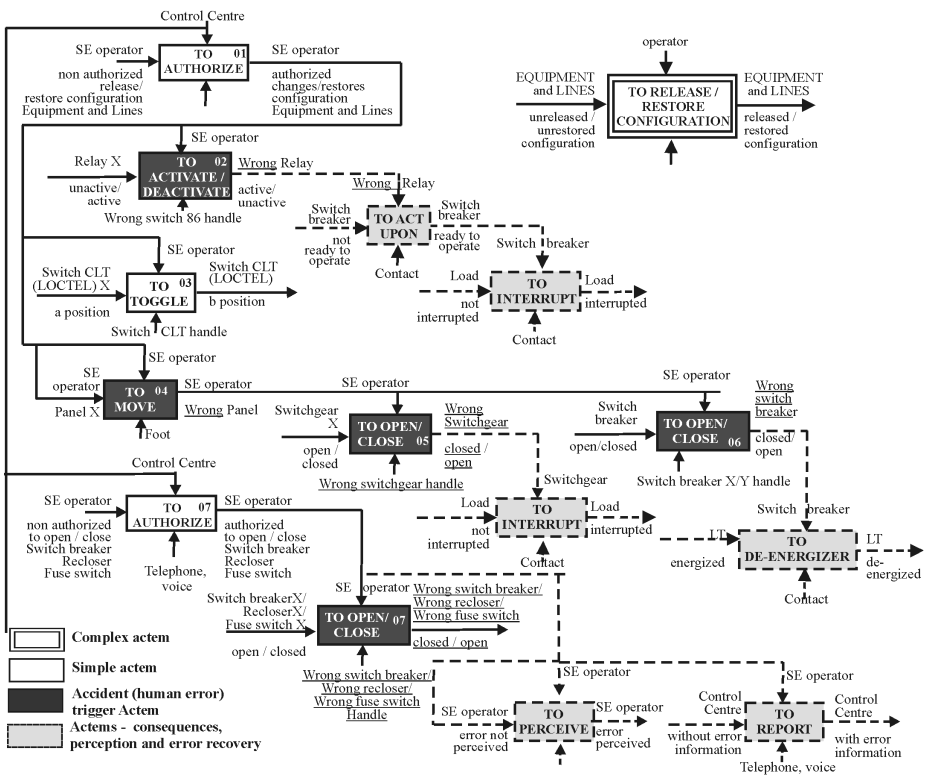


Fig. 7. Scenario Type 1: actionomy

thus compose a different scenario type. However, given the focus of this work, the chosen approach was to derive them from the different kind of error (destination error). These errors were evidenced during the knowledge extraction, since they represent the event that trigger the incident and the action performed on the human interface.

In the case study, five incidents' scenarios-types were identified: correct action on the wrong object, missing action, added action, inappropriate action, and incomplete action. These were named after the initiating action error. There follows, in Figure 6, the description of the most frequent scenario type: Correct Action on the Wrong Object, in the studied corpus. Also in this figure it is represented the instances of each concept for this particular scenario type.

The scenario type dynamical representation (Figure 7) consists of an actionomy as a meta-representation of all the scenarios of the same type. Analyzing this actionomy it was found that this kind of accident usually occurs during the tasks: release equipment and restore configuration. Both tasks consist of manipulating equipment and device switches, inside and outside the control room, during operator displacement. Some actems are highlighted to represent the actions that can lead into human error thus initiating an accident.

4 Results

The Process: The process here proposed reached its purpose as it produced the incident scenario conceptual model to be used in the construction of the incident scenario simulator. It is based on the extraction, analysis and modelling of the knowledge on a corpus of accidents reported in the electricity industry, as the result of human errors. Since it was conceived to support the improvement of the human interface of those systems, all its phases are focused in the human errors related to the interaction with the studied system. In spite of being specific to the case study, this process was generalized to be used in other contexts. Each of its phases was described by its objectives, inputs and outputs, resources and method.

The corpus definition phase provided the boundaries for the study and had as a requisite the availability of accident reports in the form of written documents, interview registrations and direct observations of work. During the knowledge extraction the KOD method was adopted which has proven to be adequate for the extraction of row knowledge represented in natural language format.

During the process presentation, the method KOD was also detailed to complement its original description in its source material. The ontology validation was considered a very important step in this research and was therefore performed with a sample of operators of real system. From the validation it was concluded that the error taxonomy was consistent with the literature [20] [25].

The identified scenario types were represented in a structure, which allows to relate the definition concepts with the corresponding actionomy that is employed to describe the various actions performed by the operator before, during and after the accident.

The Model: With the MCCA model it is expected that the incident scenario simulator would be capable of reproducing realistic critical situations during system operation, becoming an effective support tool for their training, particularly improving their reactions during urgency.

As mentioned by Johnson in [9] the accident reports explain the causes of human errors and system failure, but they do not give information on the operator's reasoning during the events. This is why it was decided to build the simulator to analyse operator's behaviour during similar situations.

Before building the accident simulator it is necessary to formalize the cognitive model obtained with the KOD method and then translate it into a software model. This step has been initiated and resulted in the proposal of a method [12][21] to translate the cognitive model obtained with KOD into a DEVS model [29].

The MCCA model is also the basis of an experimental protocol to observe the operators interacting with the scenario simulator. This protocol prescribes how to immerge the operator in the simulated environment bringing to a cognitive state as close as possible to the one produced by the real environment. It also prescribes how to keep the operator in such a state and how to observe and register the relevant parameters. The parameters will also be extracted from a post-session interview to acquire the inferences required by the KOD method and which of these inferences will evidence how the operator reasons during the interaction with the system.

The scenario types obtained through the process application give an insight into the accident situations, since these simplify the analysis. The simulation process is presented in the figure 8.

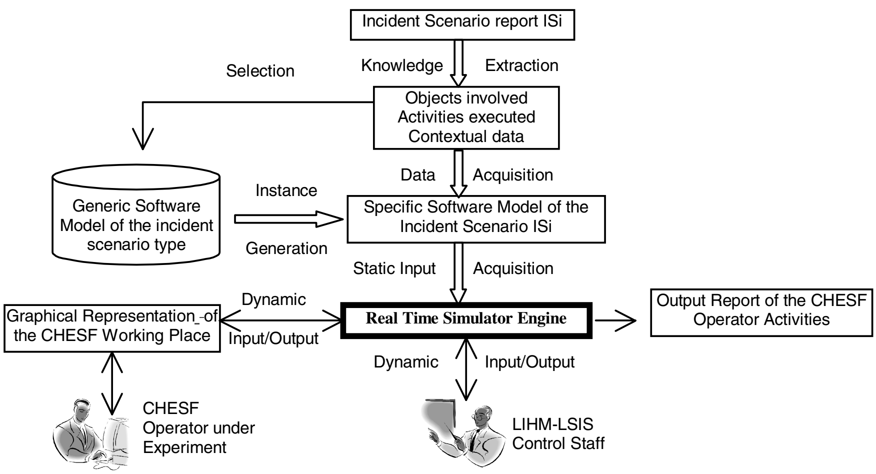


Fig. 8. The Simulation process

Along the process application, a situation is characterized by the task performed; the operator state and the installation configuration, at a particular time. This situation concept enables to diagnosis do be done of which actions performed by the operator (and which objects used to perform those actions) were directly related with the error. It also allows for analyzing the operator's behavior and perception of the situation dynamics during the task. This analysis, in turn, can lead into the diagnosis of accident causes related to the human error, and thus propose changes in the system's human interface design that could reduce their occurrence for the relevant scenario types.

The MCCA also represents knowledge which can be used in the initial process of the human interface conception. In this study, as in [3] it was possible to identify and conceptually map three categories of MCCA attributes and the attributes of the MCIE

Table 3. Task representation in MCCA and in MCIE

Task attributes in MCCA	Task attributes in MCIE (MAD*)
Terminology / Action definition (tasks)	Task Body (name, objective)
Task type (simple, complex, usual, rare)	Task centrality (frequency, importance)
Task sequence / parallelism	Temporal relations (SEQ, PAR, SIM)
Action performer	Task body (actor)
Objects employed (terminology and definition)	Task input and output conditions (explicit amidst other s the objects used to perform the task). Observation: does not include object definition.
Initial state (objects subjected to the action)	Task input conditions (initial state, pre-conditions, triggering conditions)
Final state (objects subjected to the action)	Task output conditions (final state, post-conditions, results)

(method employed in the conception of human interfaces). Table 3 also illustrates the relation between task-related attributes. Analyzing the scenario-types it is possible to investigate which tasks among those represented in the task model employed in the MCIE are critical, and the related operator's erroneous actions that might result in an accident.

The user aspects considered in the MCIE are represented In Table 4. It allows the conclusion that all of the user attributes accounted for in the MCCA are also considered in the MCIE.

Table 4. User representation in MCCA and in MCIE

User attributes in the MCCA	User attributes in the MCIE
Personal information	Personal profile
Professional information (job-description, goals with the system, level of experience)	Professional profile (job-description, experience, qualifications)
Behavior	Profile psycho-mental (problem solving)
Error causes (internal to the operator)	Profile psycho-mental / Clinic profile

It was also found that most of the attributes related to the context in the MCIE method (Table 5) can also be mapped to the MCCA.

Table 5. Context representation in MCCA and in MCIE

Context attribute MCCA	Attributes Context no MCIE
Environment (place category and definition)	Physical environment / organizational environment
Equipment (description, faults, layout)	Equipment description (product identification, product description, main application area, main function)

5 Conclusions and Perspectives

This paper presents the results of the first step towards the extension of the method MCIE to conceive human interfaces more robust to human error under critical situations. This step concerns the obtaining of a conceptual model of incident scenarios relative to the operators' errors. This model was obtained through a cognitive approach, which consisted in applying the KOD method, which has proven to be adequate. With the MCCA conceptual model, the incident scenarios, which constitute the corpus, could all be modeled.

This in itself is considered a contribution to the design of human interfaces for critical systems in the studied domain. However, a cognitive model of the operator's behavior when dealing with incidents is still needed to understand the reasons that lead into wrong actions. This understanding will allow the conception of systems that empower the user, avoiding the reoccurrence of these errors. As a result of the modelling process, there is ontology of the analyzed context, a categorization of the human error and an accident typology.

The resulting MCCA model will allow the integration of the accidents situations knowledge into the process of human interface conception, for the studied context. The specific model for the case study will support the construction of an accident simulator for the electricity company. This work is in the research context of the joint work of the human interface group GIHM – UFCG and that of the LSIS, which aims to extend the Method for Conception of Ergonomic Interfaces - MCIE, to adjust it for the context of critical industrial systems. As future work we propose the following actions.

Represent the time in the MCCA model. In the current version of the actem representation there is no mention of the time taken to perform specific actions. This information is available in the corpus, in the accident reports, which refer to the action duration directly or indirectly. With this information it will be possible to enrich the MCCA model making it possible to refine the design of the simulator as well as to support the conception of more ergonomic interfaces through more refined task modelling.

Integrate the knowledge acquired through the MCCA process into the method MCIE. This integration demands the formalization of the extracted knowledge on human error in order to include it into the process of user interface conception.

Represent the accident scenarios as task models. when trying to integrate the different models there are the difficulties, which result from the many phases of this work (engineering, psychology, sociology and computing). Each of these domains has its own paradigms and tools, which also make the communication difficult between experts [23]. When building the actionomies it was found that these are very similar to the task model representation. Since one of the research goals is to integrate the model knowledge into the conception method MCIE, and since this method already applies task modelling, it is proposed to study how to represent actionomies using task model formalism. This representation will allow to integrate the information on the errors with that of critical tasks.

Represent the time in the MCCA model. In the current version of the actem representation there is no mention of the time taken to perform specific actions. This information is available in the corpus, in the accident reports, which refer to the action duration directly or indirectly. With this information it will be possible to enrich the MCCA model making it possible to refine the design of the simulator as well as to support the conception of more ergonomic interfaces through more refined task modelling.

Integrate the knowledge acquired through the MCCA process into the method MCIE. This integration demands the formalization of the extracted knowledge on human error in order to include it into the process of user interface conception.

Verify the applicability of the proposed process to other industrial contexts, in particular the applicability of the knowledge to other electricity companies.

References

1. Ahmed, W.B., Bigand, M., Mekhilef, M., Page, Y.: MRC: un Modèle de Représentation des Connaissances en Accidentologie. In: Conférence de Ingénierie des Connaissances. Association Française d'Intelligence Artificielle. AFIA – IC 2003 (2003)
2. Amalberti, R.: The Paradoxes of Almost Totally Safe Transportation Systems. Safety Science 37, 109–126 (2001)
3. Basnyat, S., Chozos, N., Johnson, C., Palanque, P.: Incident and Accident Investigation Techniques to Inform Model-Based Design of Safety-Critical Interactive Systems. In: Gilroy, S.W., Harrison, M.D. (eds.) DSV-IS 2005. LNCS, vol. 3941, pp. 51–66. Springer, Heidelberg (2006)
4. Baumont, G., Matahri, N., Holbe, C.: Modèle d'Analyse des Incidents RECUPERARE Intégrant les Facteurs Techniques, Humains et Organisationnels. Rapport Scientifique et Technique (2002)
5. Brenac, T., Fleury, D.: Le Concept de Scénario Type d'Accident de la Circulation et ses Applications. Revue Recherche Transport Sécurité, vol. 63, pp. 63–74. Elsevier, Amsterdam (1999)
6. Di Giulio, A., Trucco, P., Pedrali, M.: Analysis of an Accident in a Co-Generation Plant: Cognitive and Organizational Issues Involved. In: ESREL 2001, Torino, Italy (2001)
7. Go, K., Carroll, J.-M.: Scenario-Based Task Analysis. In: Diaper, D., Stanton, N. (eds.) Handbook of Task Analysis for Human-Computer Interaction (2003)
8. Johnson, C.W.: A First Step Towards the Integration of Accident Reports and Constructive Design Documents. In: Felici, M., Kanoun, K., Pasquini, A. (eds.) SAFECOMP 1999. LNCS, vol. 1698, pp. 286–296. Springer, Heidelberg (1999)
9. Johnson, C.W.: Failure in Safety-Critical Systems: A Handbook of Accident and Incident Reporting. University of Glasgow Press, Glasgow (2003)
10. Keyser, V.: L'Erreur Humaine. La Recherche 20(216), 1444–1455 (1989)
11. Laprie, J.-C.: Guide de la Sûreté de Fonctionnement, Cépaduès edn., Toulouse, p. 324 (1995)
12. Mercantini, J.-M., Santoni, C., Turnell, M.F.Q.V., Guerrero, C.V.S.: Elaboration of an Incident Scenario Simulator to Analyze the Man-Machine Interaction during Critical Situations. In: Barros, F., Bruzzione, A., Frydman, C., Giambiasi, N. (eds.) I3M-Conceptual Modelling and Simulation Conference, Marseille, France, pp. 217–222 (October 2005) ISBN : 2-9520712-2-5
13. Mercantini, J.-M., Chouraqui, E., Belanger, N.: Etude d'un Système d'Aide au Diagnostic des Accidents de la Circulation Routière. In: Ingénierie des Connaissances 1999, IC 1999, Palaiseau, France (1999)
14. Mercantini, J.-M., Loschmann, R., Chouraqui, E.: Modélisation et Simulation d'un Système à Risques Multiples suivant une Approche Cognitive. In: Atelier Risque et Intelligence Artificielle - Plateforme AFIA, Laval, France (2003)

15. Mercantini, J.-M., Chouraqui, E., Turnell, M.F.Q.V., Guerrero, C.V.S., Vieira, F.A.Q., Pereira, M.R.B.: Human Centred Modelling of Incident Scenarios. In: International Conference on Systems, Man and Cybernetics. The Hague - The Netherlands. Proceedings of IEEE SMC (2004)
16. Norman, D.A.: Categorization of Action Slips. Psychological Review 88, 1–15 (1981)
17. O'Leary, M.: The British Airways Human Factors Reporting Program. Reliability Engineering and System Safety 75 (2002)
18. Rasmussen, J.: Human Errors: a Taxonomy for Describing Human Malfunctions in Industrial Installations. Journal of Occupational Accidents 4, 311–335 (1982)
19. Reason, J.: L'Erreur Humaine. Paris: PUF (1990)
20. Rouse, W.B., Rouse, S.H.: Analysis and Classification of Human Error. IEEE Transaction on System, Man, and Cybernetics SMC-13(3), 539–549 (1986)
21. Santoni, C., Mercantini, J.-M., Turnell, M.F.Q.V., Guerrero, C.V.S., Scandolo, S.: Elaboration of a Critical Situation Simulator based on a Cognitive Approach. In: Augmented Cognition, HCI International 2005 (July 2005)
22. Senders, J., Moray, N.: Human Error: Cause, Prediction, and Reduction. Lawrence Erlbaum Associates, Mahwah (1991)
23. Svedung, I., Rasmussen, J.: Graphic Representation of Accident Scenarios: Mapping System Structure and the Causation of Accidents. Safety Science 40, 397–417 (2002)
24. Turnell, M.F.Q.V.: Accounting for Human Errors in a Method for the Conception of User Interface. In: International Mediterranean Modelling MultiConference - I3M 2004, Genova, Italy (2004)
25. Van Elslande, P., Alberton, L.: L'Accident de la Route: Chercher l'Erreur. Securité et Cognition, HERMES edn. (1999)
26. Vanderhaegen, F.: Analyse et contrôle de l'Erreur Humaine. Lavoisier Hermes Science publications, Paris (2003)
27. Vergara, W.R.H.: O Uso da Experiência no Processo de Resolução de Problemas: um Estudo de Caso. Revista Produto & Produção, UFRGS 3(3), 25–34 (1999)
28. Vogel, C.: Génie Cognitif. Masson (1988)
29. Zeigler, B.P.: Theory of Modelling and Simulation. Krieger Publishing Co., Inc. (1984)

Design Patterns Applied in a User Interface Design (UID) Process for Safety Critical Environments (SCEs)

Thomas Grill and Margit Blauhut

Frequentis Austria, Innovationsstrasse 1, 1100 Vienna

Abstract. When designing applications and interfaces within the area of safety critical environments the user plays an important role. During the design phase of such applications and their interfaces, companies argue that they regard the user and comply with their needs when defining user interfaces for safety critical systems. Nevertheless the usual process that is went through in this domain is that users are rarely consulted during the requirements engineering and design process. Usability requirements are often defined by the customer himself without consulting the user and defining the user's and usage requirements. When regarding the whole spectrum of safety critical applications that are used in safety critical environments it is not possible to simply define safety critical parameters and requirements that are taken in general into account when defining such user interfaces. In order to accomplish with this we use the approach of using and reusing patterns that serve as already proven examples for design problems to address the issue of regarding the user within a user interface design process. Having the threat for people's life in focus the appropriate design solutions are evaluated against this problem and requirement. Through building a pattern repository that reflects the design issues for user interfaces within such areas we address the manifold issues of safety critical environments and propose a solution-oriented approach of addressing the user and his requirements within safety critical environments.

Keywords: Safety critical environments, patterns, HCI.

1 Introduction

A decade has elapsed since Levenson wrote the book: "Safeware: System Safety and Computers" [1]. During this time, the use of computer and digital communication systems has strongly increased in safety critical environments.

Many spectacular system failures are caused by human and user interface design errors rather than failure in software functioning, e.g. the much publicized London Ambulance Service and Therac-25 accidents [2,1] were partly attributable to poor and unusable UI design.

In this paper we examine definitions and terms related to the field of safety critical systems. We elaborate on the user interface design process applied by Frequentis. Frequentis has the unique setting that it provides communication and

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 459–474, 2008.

information systems for a broad range of safety critical domains. The knowledge and experience gathered in this area requires to specify and formalize the findings gathered over generations of user interface designers and over a long history of user interface design projects at Frequentis. We built a user interface design repository which provides a mean of holding the quality of the user interfaces designed at Frequentis stable and to implement improvements and adaptations through time. Design patterns that we identified and that are valid in safety critical environments describe archetypical design solutions in the area of safety critical environments and are gathered in a pattern library that acts as a repository of design solutions for our user interface designers [3] .

In order to be able to use and access the patterns as well as to have a concurrent approval of the existing patterns and thus a validation of the defined pattern library we establish a framework of how to apply and use patterns within the Frequentis user interface design process for applications in safety critical environment and adress tools like a pattern map to make the patterns accessible for the designer.

2 Related Work

When analysing the area of safety critical environments regarding the usage of the systems found in these areas we identified major areas that play an important role for the usability engineering process in safety critical environments.

2.1 Safety Critical Systems and Environments

> "A system whose failure may have very serious consequences such as loss of human life, severe injuries, large-scale environmental damage, or considerable economical penalties." – Isaksen [4].

This traditional definition of safety critical systems is given by Isaksen and identifies the main issue of safety critical systems that is that an error may lead to threat of human life.

The range of safety critical systems is quite broad - from "invisible" computer systems in cars to complex weapon delivery systems. Frequentis mainly focuses on communication systems in safety critical environments. Such communication systems consist of failsafe communication infrastructure providing communication facilities and end user devices (like e.g. touchscreen terminals) enabling the user to communicate with telephony and radio services. Furthermore Frequentis provides third party information through one device acting as an integrator between multiple legacy services like weather data or flight plan data in an air traffic control environment.

Isaksen [4] defines safety critical areas by identifying the usage of safety-critical systems. We expand the concept by categorizing safety critical environments according to the context of use of our products.

This context of use defines the workflow the user applies when using our systems and further directly influences the way of using the applications and also the effect of using our applications regarding to the level of safety criticality.

We define these levels as follows

– *Safety Relevant System issues* are occurrences where a failure has influence on the applied workflow and may influence the degree of safety during operation. Such a system might be a failure of a hardware device if the current workflow implies the availability of a spare device.
– *Safety Critical Systems issues* are occurrences where a failure has direct impact on the safety of people. Such safety critical issues might e.g. be a wrong indication on a radar terminal that directly leads the user to an erroneous decision.

Frequentis offers control center solutions in the following safety critical and safety relevant environments that are based on the corresponding business units like air traffic control (ATC) in civil and military environments, public safety (police, firebrigade, ambulance), public transport (train, metro) and maritime.

The term safety related sytems ("A safety-related system is one by which the safety of equipment or plant is assured." – Storey [5]) is mostly used as a synonym for safety critical sytems [5]. In this paper we will not differ between the terms safety critical and safety relevant areas in this paper as this will be reflected in the usage of the defined patterns in the framework.

2.2 Applying Usability Engineering within Safety-Critical Environments

Palanque et al. [6] discussed formal methods applied throughout the design of interactive systems with a special focus on health care in order to be able to regard safety-related issues during the design phase of such systems. They elaborated about the design space with respect to the different types of interactions and interactive devices that are applied. Regarding the safety-related issues they used the approach of incident and accident investigation [6,7].

Fourney and Carter [8] proposed guidelines for designing tactile and haptic interactions while elaborating the categories of the the respective functionality that constitute the subject of design when targeting to design for haptic and/or tactile interactions through the usage e.g. touchscreens or external navigation devices while realising functionalities through tactile/haptic interfaces. Such functionalities may be of input and/or output type.

Holzinger et al. [9] applied a user centered user interface design approach in a development process for a medical application with safety-relevant aspects reflected through the integration of the users already in the requirements phase and the available domain knowledge of the designers and developers.

Reuss et al. [10] cover the area of analyzing workflows in safety-critical environments during the requirements engineering process in order to regard the human aspect to achieve an acceptance of a proper tool.

We extend these approaches of applying existing domain knowledge by making the domain knowledge as well as already proven design solutions available and consistent through the usage of design patterns. These patterns are made accessible through a common design library and applied through an integration of the usage of patterns within the user interface design process established at our company.

2.3 Design Patterns for Safety Critical Environments

Christopher Alexander introduced the idea of describing solutions for design problems as a kind of structured patterns in order to make proper design solutions categorizable, findable and reusable (see [11,12]). According to Mahemoff usability patterns show how recurring problems in user interface design can be solved according to high level principles [13] which suit their individual capabilities and needs.

Interaction design patterns describe interaction scenarios and outline methods that are evaluated and proved to solve particular design problems.[14,15,16]

User interface design patterns define how to solve particular recurring design problems in screen- and webdesign. [15,16,12]

Nowadays the terms interaction design patterns and user interface design patterns are often used equivalently although the term itself is defined in a different way. This results from the fact that it is hard to separate interaction design from user interface design and design decisions in each category usually influence each other.

Based on the work of Dearden [17] we extend the concept of categorizing patterns by expanding the range of pattern categories through the category of information visualization in order to cover the need of distinguishing between different approaches of information architecture and visualization when categorising patterns for safety critical environments. We identified patterns in three different categories - interaction design patterns, user interface design patterns and information visualisation design patterns.

Each category depicts a different point of view in the user interface design process (see section 3) where depending on the particular user interface design task the patterns available in a pattern repository may be preselected accordingly. Hussey [18] defined a basic pattern language for safety critical applications and showed in a case study that such patterns are applicable. Connelly et.al [3] base their work on this definition from Hussey and introduce safety considerations when identifying patterns patterns for safety critical applications. We use this criteria for defining patterns in safety critical environments accordingly to elicit if a design solution fits into the area of safety relevant design approaches or not.

Table 1. Pattern categorization

Category	Description
Interaction design patterns	IxD patterns are describing an interaction flow and interaction design on a more generic level.
User interface design patterns	UID patterns provide concrete design solutions. Such solutions are normally broken down to e.g. graphical design recommendations.
Information visualization patterns	InfoViz patterns describe visualization of complex behaviour or data. Such solutions are normally broken down to e.g. graphical design recommendations. InfoViz patterns have a "partOf" relation to UID patterns (see section 4.5)

3 The Frequentis User Interface Design Process

Usability engineering at Frequentis is depicted in a complex research & development (R&D) process covering the analysis of users working in safety critical environments. The user interface design process thus is a subprocess within the R&D process of Frequentis. Other usability engineering tasks like elaborating on human factor requirements and task requirement analysis are depicted in other subprocesses within this process.

The Frequentis user interface design process (see figure 1) as a subprocess is embedded in the research and development process of Frequentis, which depicts the whole hard- and software development of Frequentis. It is splitted into three basic parts.

– *The requirements analysis and definition* steps of the process are based on requirements gathered in other subprocesses and providing information and requirements about the system that shall be defined as well as the restrictions and requirements of the users working within a particular safety critical working place.

  When understanding and defining the user interface requirements usually a workshop with the user is carried out in order to avoid misunderstandings based on unclear requirements before the start of the design phase. Further the technical system requirements and limitations are discussed and clarified with the system engineers in order to be able to regard to the requirements affecting the design of the user interface in the following design phase.

– *The early design phase* itself usually starts with the development of a first draft and prototypes in order to develop first ideas and/or to be able to evaluate already existing designs for their reuse potential.

  Later in this step the users are involved to discuss and clarify the first ideas and involve them in the design process during user interface design workshops. The output of this phase is usually a draft of a user interface specification together with prototypes of the particular design approaches.

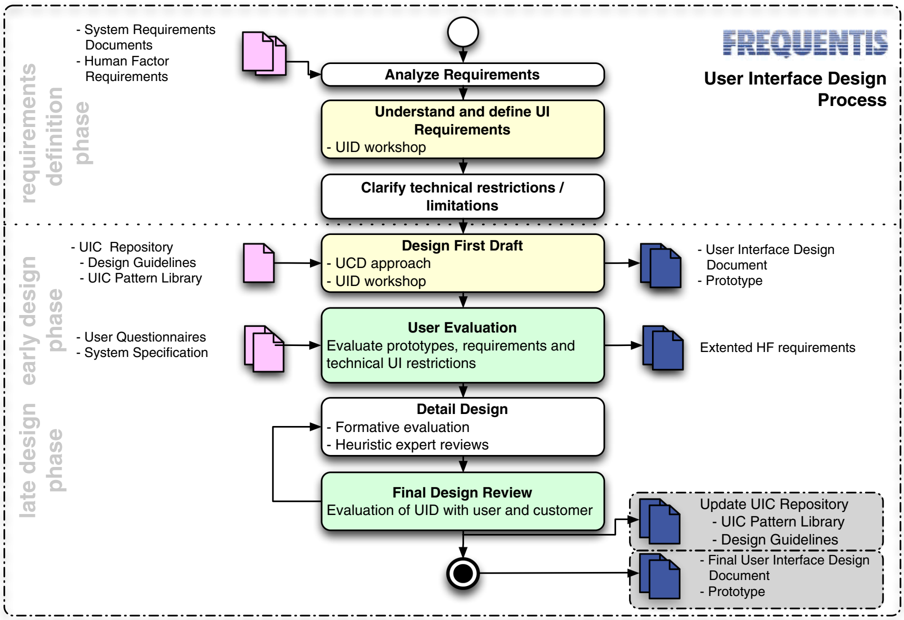


Fig. 1. Frequentis User Interface Design process

When a primary draft is developed and prototypes are available the results are evaluated together with the user in the form of user evaluations. If no direct access to the users is possible in form of user questionnaires. This enables us to identify problems of prototypes and first designs already in an early stage . Further we gather user requirements that are collected and extend our picture of the user working in an safety critical environment. Such requirements are usually depicted in personas [19,20] that describe the targeted user of the product.

– *The late design phase* is based on a detailed user interface design where designs are frequently evaluated during the design phase in the form of formative usability evaluations and expert reviews.

When the detailed design is finished a further evaluation of the final design is done with the user and customer. If no open issues need to be solved for the release of the design this results in a design freeze.

In case of identifying open issues during the evaluation phase we go back to the detail design phase and include the needed changes that are agreed with the customer. This ensures an iterative process and defines the evaluation and agreement with the customer as an exit condition for the user interface design process.

The final output of this process is a reviewed and evaluated user interface design document that in combination with the developed prototypes is the basis for the development of the product.

The last process step before finishing the project is the update of the Frequentis User Interface Center (UIC) repository containing information about design guidelines, styleguides and patterns in safety critical environments.

4 A Framework for Using Patterns in Safety Critical Environments

At Frequentis we use a repository (see section 3) that provides us with historical information, guidelines and lessons learned from former projects (UIC repository). This repository also holds a pattern library. Archetypical design solutions are depicted in the form of patterns that are made accessible via the pattern library and are described by a pattern map that makes it possible to find and make a preselection of the appropriate patterns.

Figure 2 shows a model how we define, maintain and use patterns within the Frequentis user interface design process. It further shows how the patterns are evaluated and how the validity of the pattern library is guaranteed.

4.1 Pattern Repository

Deng [21] analyzed numerous existing pattern tools that have the goal to support the designer during his task. He identified that basically these existing tools provide a collection of patterns presented in pattern libraries. Some of these tools are presented in workshops, published on the web or in books.

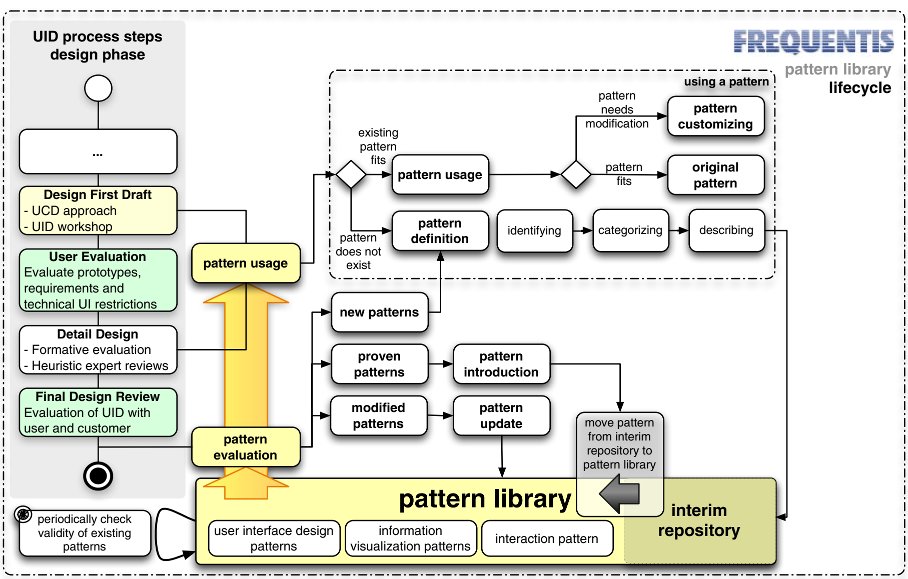


Fig. 2. Using patterns with the frequentis user interface design process

Our approach is to use an intranet based solution for providing our designers with the appropriate resources while supporting the knowledgement approach within the Frequentis User Interface Center by actualizing and filling our UIC repository where the pattern library used builds a part of it.

Gaffar et. al. presented a model of how to integrate existing patterns and also new patterns of different domains into one pattern repository. They defined the seven "C"s [22] that define a stepwise approach to an integrated pattern repository:

- **Collect:** Place different research work on patterns in one central data repository
- **Cleanup:** Change from different formats/presentations into one style is done through the definition of one common pattern style.
- **Certify:** The domain for out pattern library is based on the requirements of safety critical environments
- **Contribute:** The patterns are evaluated internally and exchanged between the products and different safety critical areas.
- **Categorize:** The categorization for the patterns is based on different approaches for the different design areas.
- **Connect:** The second level of complexity is implemented via a pattern map used for depicting the current state of the pattern repository.
- **Control:** Machine readable format for future tools

Based on this process we built our repository that integrates already existing patterns from other domains and depicts them in our repository for safety critical environments reflecting the possibilities and restrictions imposed through that domain. In order to visualize and relate the patterns we use a pattern map which is currently manually administered (see section 4.5).

For future use it is planned to use an ontology based pattern map that describes the relations between the patterns as well as it provides a machine readable format for working with the patterns and keeping the pattern library valid and consistent.

4.2 UID Process Steps

Within the Frequentis user interface design process we use two defined types of process steps where we target the usage of patterns within the design process.

The first step that incorporates an evaluation of our pattern library for the existence of a fitting pattern to a design problem is defined as the "Design First Draft" process step as well as the "Detail Design" step that might be visited iteratively within our design process.

During this steps we evaluate design problems that exist based on the requirements defined in earlier stages by evaluating the patterns already defined in our pattern library (see 4.6).

The second process step when patterns are regarded within our design process is the clean-up phase of our process. This phase includes the task of evaluating the project and revising the project for new findings that might be included in our UIC repository. If new patterns that are valid pattern candidates are identified the pattern is being defined through the *pattern definition* task (see 4.3).

If patterns already defined, but not placed in the pattern library are used they will be revised and if they proved as valid recurring design solutions they are moved from the *interim repository* to the pattern library.

If patterns are used but need adoption we evaluate whether the pattern has changed that much so that it needs to revised and the pattern in the pattern library needs to be updated.

4.3 Defining a Pattern

Dearden [17] defines characteristics for models in HCI design as operationality, expressiveness and re-usability. These characteristics exactly describe the pattern as a model and depicts the important requirements for a pattern to be fulfilled.

According to this, *operationality* defines the degree to which the knowledge encapsuled in the pattern is useful for the specific task of designing user interfaces. *Expressiveness* defines the quality of how a pattern describes the particular design problem and how clear the solution is given and applicable for designing interfaces. This is also shown by Coad who stated that in software engineering, it is usually agreed that patterns must be discovered by reference to design solutions, rather than being constructed from first principles [23,24], which we also apply to the area of usability engineering. *Re-usability* is the property of a pattern that defines weather a problem described qualifies to be a pattern useful within the task of designing userinterfaces.

In order to be able to describe how patterns are generated and considered as a pattern we split the process of defining a pattern into three subtasks.

Identifying a Pattern. As already shown by Welie [25] it is not difficult to identify patterns in user interfaces but it is hard to identify those patterns that are a real benefit for the user, and explain the usability aspects.

Almost every design may include patterns and every design solution shows a way of how to get from a problem to a proper solution. But what makes a pattern a pattern? When identifying a pattern we should be aware of a certain problem and be able to define the problem in a way so that this problem may be *more* generalized. This degree of generalization defines the re-usability of a pattern and thus the usefulness of a pattern. Defining a pattern does not make sense for a distinct solution that will probably never occur again.

A pattern needs to have a proven solution. Thus the solution should have been evaluated by testing the usability of the design solution described in the pattern.

According to Connely et.al. [3] an iterative process is needed to identify, define and revise the patterns. This approach is integrated into the framework of how to apply patterns within a user interface design process (see 4) and iterations are implemented in the pattern evaluation process (see 4.6).

After identifying a pattern the following step is to elaborate on the categorization of the pattern which is an essential step in order to make the pattern re-usable, findable and accessible for further use.

Categorizing the Pattern. In order to make a pattern reusable and be able to identify patterns needed during the design process a categorization of the patterns is introduced and used within the pattern description (see 4.3 – Describing the pattern).

There are two levels of categorization applied.

- **Pattern type** defines the type of the pattern according to the relevant areas for designing patterns. Currently we use the following categorization for our pattern library: interaction design, user interface design, information visualization.
- **Subcategory** is a keyword based sub-categorization describing the particular areas the content targets. An example for such a categorization may be for "interaction design" the subcategory of *navigation* or *direct access*. In order to facilitate the keyword selection a list of predefined keywords is given. This list is extendible but it has to be evaluated that the terms used are unambiguous. The current approach is based on the user problem categorization of Welie [25] (see table 2)

Table 2. Welie – user problem categories

Visibility	Command Area, Wizard, Contextual Menu
Affordance	Mode Cursor, Like in the real world, Setting Attributes
Natural Mapping	Like in the real world, List Browser, Continuous Filter, Navigating between spaces, Container Navigation
Constraints	Unambiguous Format, Focus
Conceptual Models	Grid Layout, Like in the real world
Feedback	Progress, Hinting, Contextual Menu
Safety	Warning, Shield
Flexibility	Preferences, Favourites

Describing the Pattern. In order to describe patterns we need a pattern form that holds the structure of the pattern and guarantees that patterns can be described and structured in a proper way.

The pattern form used at Frequentis is based on the pattern forms used by Tidwell ([15]) and Welie [16]. Tidwell introduced her pattern form in one of the first important HCI pattern collections "Common Ground". The form shown in table 3 is a modified form of the original "alexandrinian form" that built the basis for the pattern forms of Tidwell and Welie.

Table 3. Pattern Form based on Tidwell and Welie

Name	A meaningful "conceptual handle" for discussion identifying the pattern
Pattern Category	One or more categories defining the application area of the pattern
Illustration	A picture visualizing the problem
Problem	Statement of the problem and intent of the solution
Context	Tells how and where the problem occurs and when the solution works
Forces	Forces are described through design tensions, trade-offs, goals and constraints, motivating factors and concerns and tell why the problem is difficult.
Solution	The solution tells how to generate the solution. It structures its participants and collaborations
Rationale (optional)	The rationale defines underlying principles and heuristics justifying the solution. It explains underpinnings of why the solution works out.
Related Patterns	Patterns which are similar, or which may precede/follow this one
Known Uses/ Examples (optional)	3 or more independent instances of "real world" success
Notes (optional)	Additional notes that are important for the usage of the pattern

4.4 Usage of Patterns within the Frequentis UID Process

Case-based reasoning is the process of solving new problems by referencing to previous examples [26]. When using a pattern library to identify whether we already have existing patterns that may be used, we apply this method by first identifying the problem. We analyze the problem according to its categorization and thus restrict the search within the pattern library according to the previous found restrictions.

Through scanning the subcategorization done by keywords we are able to identify patterns that might fit. Another decision-making aid is given by the pattern name which helps to identify if a pattern may fit to the required problem case or not.

After identifying a pattern that may fit we analyse the forces and the proposed solution. The forces described in the pattern provide information if the pattern was defined for the desired problem areas or not. If this is true and the the solution is a valid solution to our problem we consider the pattern to be used.

If a pattern fits only in some aspects or the problem areas vary from the problem areas identified for our problems it needs to be evaluated whether the solution still is valid or may be valid in a slight modified way which leads to a possible modification of the pattern.

If a pattern does not provide a valid solution for the requirements identified for the problem we exclude the pattern from our considerations.

If a pattern was found and successfully applied this will be noted in the pattern evaluation phase and the problem case solved shall be added to the "known uses" of the pattern in the pattern evaluation phase.

4.5 Frequentis Pattern Map

The Frequentis pattern map provides a mean of accessing patterns within the Frequentis pattern library. It depicts the relations between the patterns and associates the patterns with the particular categories. This helps finding the appropriate pattern easily and gives information about the relations of the patterns in the pattern library.

Currently we are developing the pattern map as a quite static map using Microsoft Visio as a tool. For future work we intend to build an ontology of the patterns using Protege as a tool and describing the relations between the contained patterns, categories and the appropriate hierarchy as rules within the ontology.

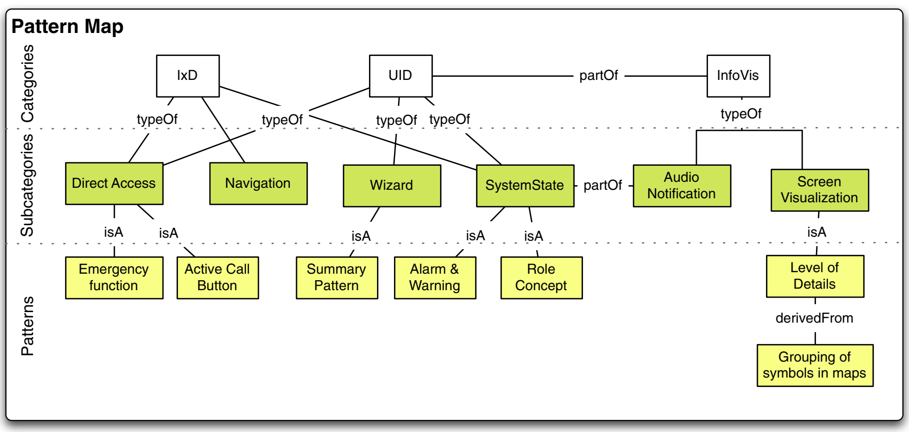


Fig. 3. Frequentis pattern map

Figure 3 shows an aperture of our internal pattern map which is currently under development. It contains the abstract categories of interaction design (IxD), user interface design (UID) and information visualisation (InfoVis). The subcategories used in the pattern map define a possibility to group the patterns according to the previously defined categorization (see section 4.3).

By defining our pattern map we use the following relationships in order to describe the rules within the targeted pattern ontology.

- *typeOf* defines the main category our pattern(s) or pattern group(s)
- *partOf* defines the relationship between categories, pattern group(s) or pattern(s). This relation between the nodes is always located on the same hierarchical level
- *isA* defines the relationship between the pattern(s) and the categorization.
- *derivedFrom* defines a hierarchical relationship between patterns.

4.6 Evaluating a Pattern

After the project is finished we perform a so-called clean-up of the project where we discuss problems and new solutions we used and adopted within the project.

If a new design solution is identified and evaluated to be considerable as a pattern we define the pattern and introduce the new pattern into an interim pattern repository which is part of the pattern library. The difference between the interim repository and the pattern library is that it holds patterns that are defined but still not proven as a pattern and thus their re-usability still is not tested. If a pattern which is located in the interim repository is used more often ($>$ three times) it is considered to be introduced into the regular pattern library and thus receive the state of being an official design pattern to be used in safety critical environments.

Further the pattern is categorized and depicted in a pattern map so that the user interface designers have access to the available patterns and find out how the patterns are related. Such a pattern map further gives information about the hierarchy of patterns and shows how patterns are related (see section 4.5).

When we identified a pattern that is used but modified it is evaluated whether the pattern is still valid or needs to be revised. This may happen due to new possibilities in technology or new findings within in the area of interaction or interface design that can be applied and introduced in the area of safety critical environments. If just a slight modification of such a pattern occurred but the pattern itself did not change it will only be adopted if the changes are introduced due to an update of the pattern. If the changed pattern is used only once it will remain as it is but a note will be added to the pattern giving a hint of possible uses.

If a used pattern shows grave relevant changes that are important to be depicted, a new pattern may result and be defined by implementing the pattern within a hierarchy based on the pattern type and described in the pattern map (see figure 3).

Another quality measure for patterns is expressed through the quality of the pattern description itself. This gives information about the reusability, the general applicability, the exactness of the categorization, etc. This quality criteria is usually defined by domain experts who conduct heuristic evaluations on the pattern in order to define a quality index.

By combining the usage count and the quality of a pattern description we are able to define a quality index that allows us to revise the particular patterns

Quality indicator	Description
***	a regularly used and evaluated IxD pattern
**	a pattern proven to describe a good interaction design approach
*	marking a pattern as relevant for the categories defined
No asterisks	a new identified interaction design pattern

Fig. 4. Pattern categories

regarding to their quality. We express such a quality index through categorizing the patterns appropriately and depicting the quality indication in the pattern itself by means of the numbers of asterisks given (see figure 4).

4.7 Householding of Patterns

A big problem when it comes to repositories containing historical information is the actuality of the data contained within such a repository.

Our approach is that at Frequentis there is a concurrent task of checking the pattern repository for actuality. Patterns are marked and the oldest patterns are identified to be rechecked again first.

Frequently the patterns in the libraries are checked and evaluated concerning validity and actuality of the pattern (see 4.6). If a pattern is outdated it is updated or if it is not valid anymore it is sorted out.

Also if a designer is confronted with out-dated patterns through his concurrent design tasks he marks these patterns to be checked or he immediately evaluates the pattern and performs the appropriate actions. This ensures an actuality of the patterns in the pattern repository.

5 Conclusion and Future Work

In literature and on the internet there already exist numerous pattern repositories that are useful and quite comprehensive. These pattern repositories describe patterns for general interaction and user interface design purposes and due to their generality they are not sufficient for the needs of safety critical environments. Further the overwhelming amount makes it difficult to use these pattern libraries as they are. We targeted this issue and introduced a special restricted pattern library that uses existing patterns as well as patterns defined for this area. We extended our user interface design process for using a pattern library located in a repository. As technology is evolving and existing pattern libraries are often not up-to-date we introduced a process that periodically evaluates the validity of the patterns. If patterns are outdated they are reworked and adopted to the actual needs of the user interface designer. The Frequentis pattern library is currently under development and we are gathering and defining the patterns relevant for safety critical environments. Future work includes building a tool that enhances the possibility to access the pattern library by using ontologies and search functionalities. Further plans are to link from user interface/interaction design/information visualization patterns to software design patterns which will help us making the product development more stable and the outcome more consistent in look and feel.

References

1. Leveson, N.: Safeware: System Safety and Computers. Addison-Wesley, Reading (1995)
2. Leveson, N.G., Turner, C.S.: Investigation of the therac-25 accidents. IEEE Computer 26(7), 18–41 (1993)
3. Connelly, S., Burmeister, J., MacDonald, A., Hussey, A.: Extending and evaluating a pattern language for safety-critical user interfaces. technical report 01-01, University of Queensland, Queensland 4072, Australia (2001)
4. Isaksen, U., Bowen, J.P., Nissanke, N.: System and software safety in critical systems. Technical Report RUCS/97/TR/062/A (1997)
5. Storey, N.R.: Safety Critical Computer Systems. Addison-Wesley Longman Publishing Co., Inc., Boston (1996)
6. Palanque, P., Basnyat, S., Navarre, D.: Improving interactive systems usability using formal description techniques: Application to healthcare. In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 21–40. Springer, Heidelberg (2007)
7. Johnson, C.W.: Failure in Safety Critical Systems: A Handbook of Accident and Incident Reporting. University of Glasgow, Glasgow (2003)
8. Fourney, D., Carter, J. (eds.): Proc. of GOTHI 2005 Guidelines On Tactile and Haptic Interactions, Saskatchewan, CANADA. Univ. of Saskatchewan, USERLab, Univ. of Saskatchewan (October 2005)
9. Holzinger, A., Geierhofer, R., Ackerl, S., Searle, G.: The user centered development of a new medical image viewer. In: Zara, J.S., Sloup, J. (eds.) Proc. of Central European Multimedia and Virtual Reality Conference 2005, pp. 63–68. Eurographics Library (2005)
10. Reuss, E., Keller, R., Naef, R., Hunziker, S., Furler, L.: Nurses' working practices: What can we learn for designing computerised patient record systems? In: Holzinger, A. (ed.) USAB 2007. LNCS, vol. 4799, pp. 55–68. Springer, Heidelberg (2007)
11. Alexander, C.: A Pattern Language: Towns, Buildings, Construction. Center for Environmental Structure Series. Oxford University Press, Oxford (1977)
12. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design patterns: elements of reusable object-oriented software. Addison-Wesley Professional, Reading (1995)
13. Mahemoff, M.J., Johnston, L.J.: Principles for a usability-oriented pattern language. In: OZCHI 1998: Proceedings of the Australasian Conference on Computer Human Interaction, Washington, DC, USA, p. 132. IEEE Computer Society, Los Alamitos (1998)
14. Borchers, J.: A Pattern Approach To Interaction Design. Wiley, Chichester (2001)
15. Tidwell, J.: Designing Interfaces, 1st edn. O'Reilly, Sebastopol (2005)
16. van Welie, M., van der Veer, G.C.: Pattern languages in interaction design. In: INTERACT (2003)

17. Dearden, A.M., Harrison, M.D.: Abstract models for hci. International Journal of Human-Computer Studies 46(1), 151–177 (1997)
18. Hussey, A.: Patterns for safety and usability in human-computer interfaces. technical report 99-05, University of Queensland (1999)
19. Grudin, J., Pruitt, J.: Personas, participatory design and product development: An infrastructure for engagement. In: Proceedings of PDA 2002, Malmö, pp. 144–161 (2002)
20. Pruitt, J., Grudin, J.: Personas: practice and theory. In: DUX 2003: Proceedings of the 2003 conference on Designing for user experiences, pp. 1–15. ACM Press, New York (2003)
21. Deng, J., Kemp, E., Todd, E.G.: Managing ui pattern collections. In: CHINZ 2005: Proceedings of the 6th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction, pp. 31–38. ACM Press, New York (2005)
22. Gaffar, A., Javahery, H., Seffah, A., Sinnig, D.: A pattern framework for eliciting and delivering ucd knowledge and practices a pattern framework for eliciting and delivering ucd knowledge and practices pattern framework for eliciting and delivering ucd knowledge and practices. In: Proceedings of HCI International 2003, Greece, pp. 108–112 (2003)
23. Coad, P.: Object-oriented patterns. Commun. ACM 35(9), 152–159 (1992)
24. Dearden, A., Finlay, J.: Pattern languages in hci: A critical review. Human-Computer Interaction 1, 49–102 (2006)
25. Welie, M.V., Trætteberg, H.: Interaction patterns in user interfaces. In: Proc. Seventh Pattern Languages of Programs Conference: PLoP 2000, pp. 13–16 (2000)
26. Kolodner, J.L.: Case-Based Reasoning. Morgan Kaufmann Publishers, San Francisco (1993)

Under Watch and Ward at Night: Design and Evaluation of a Remote Monitoring System for Dementia Care

Yvonne Schikhof[1,2,3] and Ingrid Mulder[3,4]

[1] School of Health Care Studies
[2] School of Engineering and Applied Sciences
[3] School of Communication, Media, & Information Technology
Rotterdam University
P.O. Box 25035
3001 HA Rotterdam, the Netherlands
[4] Telematica Instituut
P.O. Box 589
7500 AN Enschede, the Netherlands
y.schikhof@hro.nl, mulderi@acm.org

Abstract. The ageing population as well as the tight labor market put pressure on future health care. In this article, we explore the role of monitoring systems in health care. By incorporating principles of Value-Sensitive Design in a Human Centered Design process we developed a system for remote monitoring at night in dementia care. The performance of the working system was evaluated in the real-life context of a nursing home. Next to reporting the design and evaluation of the monitoring system, we reflect upon the approach taken.

Keywords: Value Sensitive design, Human centered design, Assistive technology, Acceptance, Monitoring people with dementia.

1 Introduction

Due to demographic changes the number of people with dementia worldwide is predicted to rise from 25 million in 2000 to 63 million by 2030 and to 114 million by 2050 [13]. In the Netherlands estimations start from around 175,000 in 2002 to approximately 207,000 by 2010, leading to a rough estimate of 412,000 dementia patients in 2050. The rise of other chronic conditions in elderly people, together with the tight labor market, jeopardizes the quality of care. On top of this the potential working population (age 15 to 64 years) is set to undergo a further substantial decline. Whereas there were still 63 potential workers for every dementia patient in 2000, this number is set to fall to 55 in 2010 and to be just 27 in 2050 [6]. These figures stress the importance of developing laborsaving technology.

For these reasons, we explored the minimal complement of staff and the role of assistive technology in a nursing home at night, taking into account the requirement established by the Netherlands Health Care Inspectorate to have adequate monitoring

A. Holzinger (Ed.): USAB 2008, LNCS 5298, pp. 475–486, 2008.

and permanent alertness during day and night. For this purpose, monitoring aids are welcomed. Infrared sensors are often used to alert staff in case a psycho-geriatric patient who might need assistance, for example because of a high risk of falling, gets out of bed. The frequency of false positive alarms is an often heard complaint. Responding to every alarm will cost valuable time of staff due to the large working area and may cause alarm fatigue. This problem will even get larger due to the increasing interest for small-scale housing for dementia patients. Long term care for people with dementia is changing in The Netherlands. The focus is increasingly on quality of life and delaying institutionalization by offering substitutes to the traditional nursing home. Small-scale units allow for optimal supervision during the day and are considered to provide a better environment. A small group of older people with dementia, approximately six, live together in a home-like environment. In order to keep daily life for the residents as normal as possible, the required personal care is integrated in daily routines. For the nursing staff, it implies that they need to perform care tasks as well as domestic tasks, such as cooking and cleaning. Furthermore, the concept of group living care means that residents lead a normal daily life and can therefore be managed by just one or two staff members each day. This is a major difference to traditional nursing homes, where more staff members are usually present. According to recent research, elderly people suffering from dementia are better off in small-scale care accommodation than in a large-scale nursing home. It seems that the professional caregivers also prefer to work in smaller scale environments. These group living homes provide a more attractive psychosocial working environment, resulting in a higher level of well-being of the nursing staff in comparison to traditional nursing homes. Staff members are happier with their work and suffer fewer burnt-out symptoms. They enjoy greater job satisfaction. The relatives of residents have also expressed that the staff members are less rushed and believe that more respect is shown for the experiential world of the residents and the feeling of family members [2,11]. During daytime one staff member is always present in the unit. At night, however, one nursing-assistant is responsible for four units and thus the time to respond to an alarm will increase. Other than in a large-scale nursing home, the units are not connected to each other and have their own entrance door. Consequently, the potential of assistive technology for the planning of new residential care facilities can hardly be ignored. How to select and implement such assistive technologies, however, is not obvious; no best practices of integrated monitoring systems for dementia care are yet commercially available. In a recent state-subsidized project an integrated system was developed, using standard infrared sensors in bedrooms. No cameras were used in the bedrooms as it was assumed that this would violate privacy issues [8]. Given the current problems with false positive alarms and inadequate staffing, health care institutions are looking for a proper solution.

2 Case in Dementia Care

The management of a health care organization in Rotterdam, the Netherlands, put the problem of false positives alarm and the plans for small-scale housing on the agenda. In this discussion the role of assistive technology in monitoring dementia residents was emphasized, in particular, the possibility of using an integrated system with sensors and cameras in bedrooms to monitor the residents with problem behavior at night. Currently, one staff member on night duty has the responsibility for 30 residents with advanced dementia living in one unit. But with small-scale housing one staff member will be responsible for about 27 residents living in four separate units. It is hardly surprising that adequate monitoring is becoming increasingly difficult. With the help of the management and a nurse team leader a first set of requirements for a monitoring system was formulated, including issues of privacy and other values. The premise was that privacy and other values could be respected through a conscientious design of the system and guidelines. It was, for instance, clear from the beginning that the cameras should start recording only when activated by a sensor. In addition the staff member on duty should be able to see the life images (and hear the sounds) to make a proper decision whether to respond to the alarm and go to the bedroom. In other words, the integrated system should assist the staff member on duty. However, privacy and other important values had to be guaranteed. These values, which still needed further specification, had to be incorporated in the designing process. The first sketch of the system was made to find a company who could develop the system and to discuss the system in the context of use, see Figure 1.

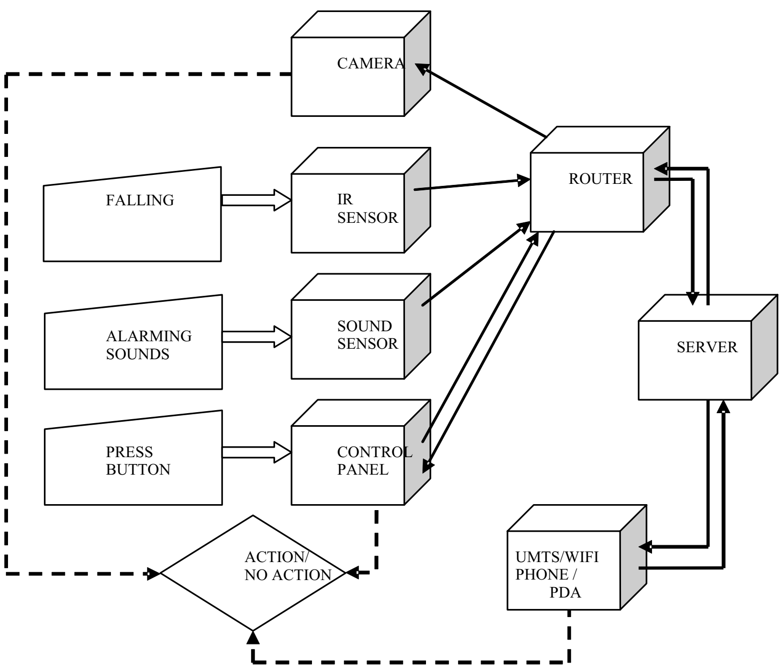


Fig. 1. Sketch to illustrate the requirements of the proposed system

In close consultation with the health care organization and the company, based on this sketch the following first level requirements for the current nursing home were specified: infrared sensor; door switch; microphone; camera; server; individualized alarm setting; no inconvenience for residents; live images and sound to be sent to a personal digital assistant (PDA) carried by the staff member on duty in case of alarm; assistive technology; informed consent by representative of resident; and information to residents' relatives and staff. In addition, the acceptance by both proxies and staff was seen as crucial for the continuation and implementation of the proposed remote monitoring system.

3 Designing Remote Monitoring System

As the initial discussions emphasized the human values rather than the potential technological opportunities, principles of both value sensitive design and human centered design were leading in our design trajectory.

3.1 Value Sensitive Design

Value sensitive design (VSD) is a methodological design approach that aims to build in moral values into technological design. The assumption is that human values, norms and moral considerations can be imparted in things we make and use. VSD is proactive and influences the design of technology in all stages of the design process. It includes all values, especially those with moral import. VSD builds on an iterative methodology that integrates conceptual, empirical and technical investigations [4].

Conceptual investigations provide philosophically informed analyses of the central constructs and issues relevant to the system under development. Technical investigations identify how existing technical designs and mechanisms engender value suitabilities and, conversely, how the identification of specific values can lead to new technical designs and mechanisms to support better those values. Empirical investigations draw on social science methodologies to understand the value-oriented perceptions and experiences of the direct and indirect stakeholders of a given system. The Value-Sensitive Design investigations are employed in consort with other successful technical methods [3].

In this case the Value-Sensitive Design investigations needed to be incorporated in a design process which was suited for a human centered approach. This is in keeping with Harper et al. [5] who suggest to focus on human values and to pinpoint those that we wish to design for and to research as a new initial stage in the Research and Design Cycle.

3.2 Human Centered Approach

In the current study, the need for a remote monitoring system was not driven by technology but rather, by the practical problems of health care staff and management in dementia care. Consequently, the need for human centered design was identified and the human centered approach was chosen as a leading method.

User information is an important source of human centered design; a design that is based around the real and actual requirements of users, involving the end-users from the beginning, and typically involves task analysis, prototype development with users, evaluation, and iterative design [7,10]. Nursing-assistants, nurse team leaders and residents were identified as current users. The aim was to inform and consult the nursing-assistants about the design of the remote monitoring system after students of

Health Care Technology of the Rotterdam University had tested the working prototype and before it was made ready to be tested in practice. Patient involvement in the design was a difficult issue. Dementia makes independent human beings dependent on decision-making by others. Information provided by representatives, usually family members, can be used in the design. One should be careful, however, not to confuse the needs of these relatives with the needs of the residents. Introducing technology in dementia care always involves the risk of dehumanizing the person with dementia [1]. At the start of designing the remote monitoring system the question 'What is driving the development?' had to be answered. As has been said, in our case, practical problems in dementia care and not the push for technology started the process. By choosing human centered design as the leading method and the fact that important values were at stake, VSD was incorporated in the standard ISO 13407 for a human-centered design process [12]. As a result, the cycle of work, as illustrated in Figure 2, was complemented with extra activities from the VSD methodology within step one and two.

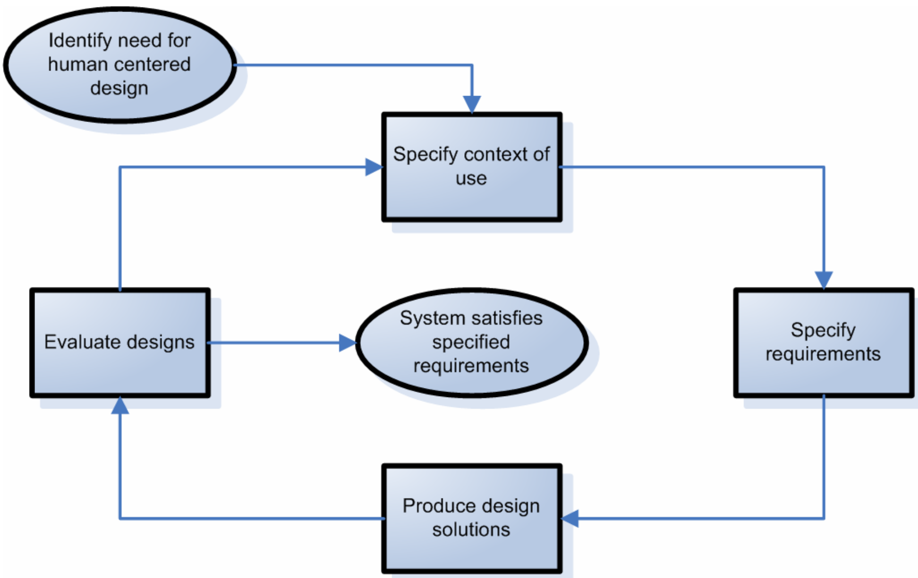


Fig. 2. Cycle after ISO 13407 (1999)

These activities that structure the remainder of the article as well are defined as followed:

Activity 1. Specify the context of use, including conceptual investigations; i.e., analyzing the context by observation and flow charts, deliberation with staff and identifying stakeholders and values;
Activity 2. Specify user and organizational requirements, including empirical and technical investigations; i.e., laboratory test, survey on acceptance, qualitative research and group discussions;

Activity 3. Produce design solutions; newly specified requirements led to design decisions;
Activity 4. Evaluate design; i.e., performance test in real life context, survey and discussion.

3.3 Context of Use and Conceptual Investigations

For each of three problem behaviors (falling, panicking and wandering) of a resident during the night all possible scenarios were summarized in flow charts. These flow charts embodied the monitoring system, the possible actions of the staff member on duty and the time needed to decide on the response. The flow charts made clear what information was needed and what sensors should be used. A staff member on night duty (the user) is responsible for the surveillance of the 30 residents in the ward. In general the staff members make rounds and provide care and assistance to those residents who need this. Staff members must be able to move freely and have to prioritize the follow-up of alarms, the needs of residents, et cetera.

The context of use was further specified by the management and successively by the nurse team leaders. The team leaders also selected the ward that they thought to be best suited to test the system in practice. After that stakeholders and values concerned were identified in a conceptual investigation.

Direct stakeholders - that is, the residents with dementia and their family (representatives) - are people who are affected by the system. Indirect stakeholders - that is, the staff members and the management of the health care organization - are going to use the system. The values displayed in Table 1 were identified by the stakeholders.

Table 1. Identified values by the different stakeholders

Residents and Family members	***Staff members***	***Management***
Privacy	Privacy	Privacy
Consent	Consent	Consent
Respect	Respect	Quality of Care
Individuality	Autonomy	Feasibility
Dignity		
Warmth		
Safety		
Well-being		

3.4 Specify Requirements, Including Empirical and Technical Investigations

A laboratory test was done to determine whether the proactive design supported the identified values and to sharpen the user requirements. Besides it was also introduced to make sure that the users in the nursing home could test a system without asking too much of their patience by unveiling all kinds of problems. Students tested a working prototype in an improvised bedroom. The management and a team leader gave their comments and feedback. Teething troubles, like easy to solve problems, which are annoying and possibly demotivating to users while doing their jobs, were detected.

Also the requirements were accentuated. For instance, the camera used in the laboratory test had a zoom function. Students thought this to be helpful at night to recognize the facial expression as the night vision gave a somewhat blurry image. The team leader thought that the zoom function was not necessary because staff was familiar with the residents. It was decided that the simpler (and cheaper) camera was going to be used. The usability evaluation should pay particular attention to the need for extra information, such as facial expressions. Steering the camera by the PDA to be able to view the entire bedroom was added to the design. In the laboratory-bedroom, different sensors for bed-exit detection were tested. All sensors had their disadvantages. The optimal situation was provided by a combination of the commonly used infrared sensor beside the bed and an infrared sensor above the bed.

After the laboratory test the acceptance of the designed system was investigated. Two informational meetings were organized, one for the nursing assistants and one for the family members of the chosen ward. A survey was carried out before and after the informational meetings. The pre-test as well as the post-test included identical questions and statements about the acceptance of camera surveillance in the bedrooms of the residents with dementia. The questions were ranked on a 4-point Likert scale (from highly acceptable till not acceptable at all). We opt for a 4-point Likert scale to enforce people to choice and not to give a neutral answer.

Both staff members (n=8) and family members (n=13) attended the informational meetings. 18 persons participated in the survey. Before the information was given and questions could be asked about testing and using the system in practice, six more staff members and family members found limited camera surveillance in the bedroom of a resident with dementia not or not acceptable at all compared with the post-test. After the informational meetings, all staff members and family members found it acceptable or highly acceptable. Table 2 displays the exact numbers.

Table 2. Acceptance of limited camera surveillance in the bedrooms of residents with dementia

N=18	**Highly acceptable**	**Acceptable**	**Not acceptable**	**Not acceptable at all**
PRE	1	11	4	2
POST	4	14		

In addition, qualitative data were collected by interviewing three family members of the residents who had participated in the survey in order to distinguish the needs of the family member from the needs of the resident when favoring the system. In semi-structured interviews the family members referred to safety and better surveillance as grounds for the acceptance of limited camera surveillance in the bedroom. Terms for acceptance were informed consent and that only staff members (nursing-assistants, nurses and doctors were mentioned) were allowed to see the live images. Obviously, 'trust' (user value) seemed to be important.

During the informational meetings staff and family could ask questions, and give comments and feedback for a more detailed specification of the requirements. Some family members were worried that the performance test of the current system eventually would lead to the complete replacement of staff at night by a remote system. Some staff members were concerned that the system would be used to control their activities. In both meetings there was a discussion about whether and how long the recorded images should be stored.

In sum, the following requirements were specified: infrared sensor above and beside the bed; door switch; microphone; camera with motor drive but without zoom; camera starts when activated by a sensor; server; individualized alarm setting; no inconvenience for residents; in case of alarm live images and sound to be sent to PDA carried by a staff member on duty, good signal reception anywhere in the selected ward; duration of live images is adjustable; no talking back facility; assistive technology for staff members only with simple use of the PDA; written instructions; informed consent by representative of resident; no storage of images; system cannot be switched off by staff member on duty; system is in operation from 8 PM till 7 AM.

3.5 Design Solution

The newly specified requirements were used to improve the design for the remote monitoring system that would be installed in the bedrooms of the selected residents in the selected ward. The following design decisions were made:

- An infrared sensor above the bed should be installed for early bed-exit detection with residents at high risk of falling when leaving the bed without assistance.
- The infrared sensor beside the bed was suited for residents who tend to come out of bed and start disturbing other residents, wandering, et cetera.
- The noise threshold of the microphone could be individually set to detect restlessness, panic reactions, et cetera.
- The door switch could detect if a resident left the bedroom or if someone (other than the staff member on duty) entered the bedroom.
- An infrared light accommodated the night position of the camera.
- The computer and server were placed in the office of the team leader where the individual settings could be changed.
- Wifi access points were installed to make reception possible within the department.
- An instruction manual for the use of the PDA by the nursing assistants was written. A short version of the manual providing a quick overview of how to activate the monitoring system with the PDA was placed above the two PDAs in their charger.
- The use of the PDA was kept as simple as possible.
- The nursing assistants also received guidelines how to incorporate the monitoring system in their work programme.
- Two telephone numbers were available for user support during the test.

The structure of the system is displayed in Figure 3. Figure 4 shows the resulting PDA used by the nursing assistants.

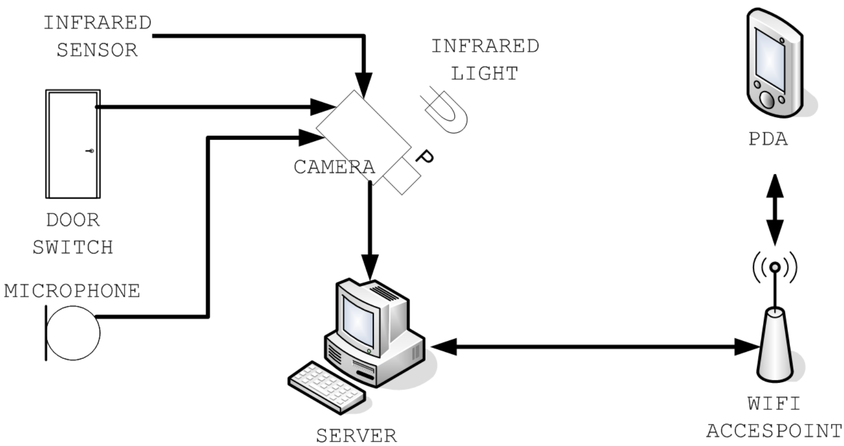


Fig. 3. Design structure for remote monitoring system

Fig. 4. PDAVOS (Video Observation System)

4 Evaluate Design

The usability evaluation of the working system took place in a performance test in actual practice at the nursing home. The remote monitoring system was tested with four residents in the selected ward during the evening and night shifts in a 4-week period. On the PDA as shown in Figure 4 a life video from the bedroom of a resident would show, accompanied by an alarm sound whenever a sensor alarm was activated the camera. The nursing assistant could steer the camera by placing a finger on the screen. By pressing OK the video would disappear and the image as displayed in Figure 4 appears again or would be followed by another alarm and video.

Major usability problems were identified that had not been revealed by the early testing in the laboratory phase, like the problems due to a power failure and the bad signal reception near a reinforced wall. Some new problems could already be solved during the performance test. A body armour for the PDA with adjustable carrying strap was introduced to prevent damage and switching off the PDA. The timeframe of the system was extended by one hour in the morning.

The nursing-assistants on duty were asked to complete a questionnaire for every shift in which the system was used. Questions concerned the users' effectiveness and satisfaction and the usability and effectiveness of the system in practice. We learned that the nursing-assistants could comfortably work with the system within the first shift with the PDA. They used the remote steering of the camera, but did not feel the need for an extra zoom facility on the camera. They rated the system as satisfactory during the test and were disappointed that it had to be removed after the test phase.

The results were discussed with the staff member who had worked most frequently with the system, with the team leader and with the management. Any difference in environment and working process between the current situation and the future small-scale housing were also discussed with the management. Advice was given on the monitoring of the residents in small-scale housing. Interestingly, the system was referred to as 'an extension of the eyes and ears' of the staff member.

After having presented the results to the staff and the management, we also presented the project to outsiders. We asked a peer health organization with experience in introducing technology in a residential environment to comment on the presentation. A professor Integrated Care for the Elderly acted as an opponent and staff from the university as well as from other health care organizations was present. The audience could ask questions or react to the opponents. Overall, the evaluation was very positive. Especially the role of the stakeholders and the test in the real life context of the nursing home was appreciated. The fact that an early user test could detect problems that can be solved before implementing a system was considered as a very positive contribution in accepting new technology. The peer healthcare organization with experience in this field concluded that this system would suit the users more then the technology they worked with. Additionally, the focus on values like privacy gained applause. Moreover, the boards of the current health care organization and the peer organization decided to implement a remote monitoring system in a small-scale housing project next year.

5 Conclusions and Discussion

In this study, we have designed and tested a new system for monitoring people with dementia at night using human centered design and value sensitive design methodology. It can be concluded that combining the methodological design approaches of Human Centered Design and Value Sensitive Design in an iterative process proved to be beneficial. Values were discussed very early in the process, and the new working system of assistive technology performed very well. The idea that using cameras in bedrooms of residents with dementia is not acceptable because of intrusion on privacy was not upheld in this case. In the care relationship between staff and residents seems trust to be an important value for acceptance both for residents ad proxies.

The chosen approach resulted in a system that was workable according to the experts (users). Incorporating values into the design is different from discussing ethics when introducing innovative technology in geriatric care, as has been suggested by Sävenstedt and colleagues [9]. A technology driven process involves greater risks. It will be more difficult to meet the human needs of staff as well as the needs of the frail elderly people when a development life-cycle is not used. Innovative technology can be seen as a promoter of both humane and inhumane care [9]. We learned that aiming at humane care is promoted by using a human centered approach.

Acknowledgments. This project was financially supported by the Ministry of Health, Welfare and Sports. The remote monitoring system was tuned to the design in cooperation with Anneloes Cordia and EEG Technology inc., who also provided an excellent help desk during the test phase. Jacomine de Lange kindly gave her advice. The project could not have taken place without the kind cooperation of the management, staff, residents and family members of the nursing home Smeetsland from the 'De Stromen Opmaat Groep'. We are also grateful to the students involved in the project, and last but not least research assistant Jan-Willem Bullee who helped to achieve the results. Tischa van der Cammen and Hanny Groenewoud are thanked for their comments on the manuscript.

References

1. Astell, A.J.: Technology and personhood in dementia care. Quality in ageing 7(1), 15–25 (2006)
2. te Boekhorst, S., Willemse, B., Depla, M.F.I.A., Eefsting, J.A., Pot, A.M.: Working in group living homes for older people with dementia: the effects on job satisfaction and burnout and the role of job characteristics. International Psychogeriatrics 20(5), 927–940 (2008)
3. Friedman, B., Howe, D.C., Felten, E.: Informed Consent in the Mozilla Browser: Implementing Value-Sensitive Design. In: Proceedings of the 35th Hawaii Conference on System Sciences. IEEE, Los Alamitos (2002)
4. Friedman, B., Kahn, P.H., Borning, A.: Value Sensitive Design: Theory and Methods. UW CSE Technical Report 02-12-01, University of Washington, Dept. of Computer Science & Engineering, Seattle, WA (2002)
5. Harper, R., Rodden, T., Rogers, Y., Sellen, A. (eds.): Being Human: Human-Computer Interaction in the Year 2020. Microsoft Research Ltd., Cambridge (2008)
6. Health Council of the Netherlands. Dementia. Publication no. 2002/04E. The Hague: Health Council of the Netherlands (2002)
7. Memmel, T., Reiterer, H., Holzinger, A.: Agile Methods and Visual Specification in Software Development: A Chance to Ensure Universal Access. In: Stephanidis, C. (ed.) Universal Access in HCI, Part I, HCII 2007. LNCS, vol. 4554, pp. 453–462. Springer, Berlin (2007)
8. Nouws, H., Sanders, L., Heuvelink, J.: Domotica voor dementerenden [Home automation for people with dementia], De Vijfde Dimensie, Amersfoort, The Netherlands (2006)
9. Sävenstedt, S., Sandman, P.O., Zingmark, K.: The duality in using information and communication technology in elder care. Journal of Advanced Nursing 56(1), 17–25 (2006)

10. Thimbleby, H.: Understanding User Centered Design (UCD) for People with Special Needs. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1–17. Springer, Heidelberg (2008)
11. Trimbos Institute. Elderly with dementia better off in small-scale nursing homes (29-01-2007). Press sheet Trimbos Institute, Netherlands Institute of Mental Health and Addiction, http://www.trimbos.nl/default33.html?printId=22199
12. UPA. What is User-Centered Design? Retrieved from: June, 25 2008, http://www.upassoc.org/usability_resources/about_usability/what_is_ucd.html
13. Wimo, A., Winblad, B., Aguero-Torres, H., von Strauss, E.: The magnitude of dementia occurrence in the world. Alzheimer disease and associated disorders 17(2), 63–67 (2003)

Author Index