IFIP AICT 332

Peter Forbrig
Fabio Paternó
Annelise Mark Pejtersen
(Eds.)

Human-Computer Interaction

Second IFIP TC 13 Symposium, HCIS 2010
Held as Part of WCC 2010
Brisbane, Australia, September 2010
Proceedings

Springer

World Computer Congress
Brisbane 2010
International Federation for Information Processing

IFIP Advances in Information and Communication Technology 332

IFIP – The International Federation for Information Processing

IFIP was founded in 1960 under the auspices of UNESCO, following the First World Computer Congress held in Paris the previous year. An umbrella organization for societies working in information processing, IFIP's aim is two-fold: to support information processing within its member countries and to encourage technology transfer to developing nations. As its mission statement clearly states,

> *IFIP's mission is to be the leading, truly international, apolitical organization which encourages and assists in the development, exploitation and application of information technology for the benefit of all people.*

IFIP is a non-profitmaking organization, run almost solely by 2500 volunteers. It operates through a number of technical committees, which organize events and publications. IFIP's events range from an international congress to local seminars, but the most important are:

- The IFIP World Computer Congress, held every second year;
- Open conferences;
- Working conferences.

The flagship event is the IFIP World Computer Congress, at which both invited and contributed papers are presented. Contributed papers are rigorously refereed and the rejection rate is high.

As with the Congress, participation in the open conferences is open to all and papers may be invited or submitted. Again, submitted papers are stringently refereed.

The working conferences are structured differently. They are usually run by a working group and attendance is small and by invitation only. Their purpose is to create an atmosphere conducive to innovation and development. Refereeing is less rigorous and papers are subjected to extensive group discussion.

Publications arising from IFIP events vary. The papers presented at the IFIP World Computer Congress and at open conferences are published as conference proceedings, while the results of the working conferences are often published as collections of selected and edited papers.

Any national society whose primary activity is in information may apply to become a full member of IFIP, although full membership is restricted to one society per country. Full members are entitled to vote at the annual General Assembly, National societies preferring a less committed involvement may apply for associate or corresponding membership. Associate members enjoy the same benefits as full members, but without voting rights. Corresponding members are not represented in IFIP bodies. Affiliated membership is open to non-national societies, and individual and honorary membership schemes are also offered.

Peter Forbrig Fabio Paternó
Annelise Mark Pejtersen (Eds.)

Human-Computer Interaction

Second IFIP TC 13 Symposium, HCIS 2010
Held as Part of WCC 2010
Brisbane, Australia, September 20-23, 2010
Proceedings

Springer

Volume Editors

Peter Forbrig
University of Rostock, Department of Computer Science
Albert-Einstein-Str. 21, 18051 Rostock, Germany
E-mail: peter.forbrig@uni-rostock.de

Fabio Paternó
ISTI-CNR
Via G.Moruzzi 1, 56124 Pisa, Italy
E-mail: fabio.paterno@isti.cnr.it

Annelise Mark Pejtersen
Risø National Laboratory, Centre for Cognitive Systems Engineering (CSE)
4000 Roskilde, Denmark
E-mail: annelise.m.pejtersen@risoe.dk

Library of Congress Control Number: 2010932427

CR Subject Classification (1998): H.4, C.2, H.5, I.2, K.4, J.4

ISSN	1868-4238
ISBN-10	3-642-15230-9 Springer Berlin Heidelberg New York
ISBN-13	978-3-642-15230-6 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

IFIP World Computer Congress 2010 (WCC 2010)

Message from the Chairs

Every two years, the International Federation for Information Processing (IFIP) hosts a major event which showcases the scientific endeavors of its over one hundred technical committees and working groups. On the occasion of IFIP's 50th anniversary, 2010 saw the 21st IFIP World Computer Congress (WCC 2010) take place in Australia for the third time, at the Brisbane Convention and Exhibition Centre, Brisbane, Queensland, September 20–23, 2010.

The congress was hosted by the Australian Computer Society, ACS. It was run as a federation of co-located conferences offered by the different IFIP technical committees, working groups and special interest groups, under the coordination of the International Program Committee.

The event was larger than ever before, consisting of 17 parallel conferences, focusing on topics ranging from artificial intelligence to entertainment computing, human choice and computers, security, networks of the future and theoretical computer science. The conference History of Computing was a valuable contribution to IFIPs 50th anniversary, as it specifically addressed IT developments during those years. The conference e-Health was organized jointly with the International Medical Informatics Association (IMIA), which evolved from IFIP Technical Committee TC-4 "Medical Informatics".

Some of these were established conferences that run at regular intervals, e.g., annually, and some represented new, groundbreaking areas of computing. Each conference had a call for papers, an International Program Committee of experts and a thorough peer reviewing process of full papers. The congress received 642 papers for the 17 conferences, and selected 319 from those, representing an acceptance rate of 49.69% (averaged over all conferences). To support interoperation between events, conferences were grouped into 8 areas: Deliver IT, Govern IT, Learn IT, Play IT, Sustain IT, Treat IT, Trust IT, and Value IT.

This volume is one of 13 volumes associated with the 17 scientific conferences. Each volume covers a specific topic and separately or together they form a valuable record of the state of computing research in the world in 2010. Each volume was prepared for publication in the Springer IFIP Advances in Information and Communication Technology series by the conference's volume editors. The overall Publications Chair for all volumes published for this congress is Mike Hinchey.

For full details of the World Computer Congress, please refer to the webpage at http://www.ifip.org.

June 2010

Augusto Casaca, Portugal, Chair, International Program Committee
Phillip Nyssen, Australia, Co-chair, International Program Committee
Nick Tate, Australia, Chair, Organizing Committee
Mike Hinchey, Ireland, Publications Chair
Klaus Brunnstein, Germany, General Congress Chair

Preface

The IFIP World Computer Congress (WCC) is one of the most important conferences in the area of computer science at the worldwide level and it has a federated structure, which takes into account the rapidly growing and expanding interests in this area. Informatics is rapidly changing and becoming more and more connected to a number of human and social science disciplines. Human–computer interaction is now a mature and still dynamically evolving part of this area, which is represented in IFIP by the Technical Committee 13 on HCI. In this WCC edition it was interesting and useful to have again a Symposium on Human–Computer Interaction in order to present and discuss a number of contributions in this field.

There has been increasing awareness among designers of interactive systems of the importance of designing for usability, but we are still far from having products that are really usable, and usability can mean different things depending on the application domain. We are all aware that too many users of current technology often feel frustrated because computer systems are not compatible with their abilities and needs in existing work practices. As designers of tomorrow's technology, we have the responsibility of creating computer artifacts that would permit better user experience with the various computing devices, so that users may enjoy more satisfying experiences with information and communications technologies. This has raised new research areas, such as ambient intelligence, natural interaction, end user development, work analysis and cultural and social interaction.

The interest in the conference was positive in terms of submissions and participation. We received 56 contributions, 19 submission were accepted as long presentations and 7 as short ones. Additionally, five submissions were accepted as poster presentations. The selection was carried out carefully by the International Program Committee. The result is a set of interesting and stimulating papers that address such important issues as specification and modeling, user-centered development, usable systems, social and cultural problems, mobile and Web applications and interaction. The interest shown in the conference was truly world-wide: if we consider both full and short papers, we had authors from Europe, South America, Canada, Japan, Singapore, and South Africa. The final program of the symposium included one technical invited speaker: Penelope Sanderson from The University of Queensland on "The Power and the Puzzles of Auditory Interfaces." This is a very interesting topic for many application domains.

We hope that the IFIP Symposium on Human–Computer Interaction contributed to the World Computer Congress in terms of many interesting discussions and that the second symposium was even more successful than its first edition in Milan in 2008.

June 2010

Peter Forbrig
Fabio Paternò
Annelise Mark Pejtersen

HCIS 2010 Conference Organization

The Second IFIP Human–Computer Interaction Symposium
(HCIS 2010)

was a co-located conference organized under the auspices of the
IFIP World Computer Congress (WCC) 2010
in Brisbane, Australia

Peter Forbrig
HCIS 2010 Co-chair
Peter.forbrig@uni-rostock.de

Fabio Paternò
HCIS 2010 Co-chair
fabio.paterno@isti.cnr.it

Annelise Mark Pejtersen
HCIS 2010 Co-chair
ampcse@mail.dk

HCIS 2010 Program Committee

Julio Abascal, Spain
Liam J. Bannon, Ireland
Simone Diniz Junqueira Barbosa, Brazil
Torkil Clemmensen, Denmark
Andy Dearden, UK
Anke Dittmar, Germany
Henry B.L. Duh, Singapore
Achim Ebert, Germany
Nahum Gershon, USA
Giuseppe Ghiani, Italy
Nicholas Graham, Canada
Tom Gross, Germany
Jan Gulliksen, Sweden
Judy Hammond, Australia
John Karat, USA
Joaquim A. Jorge, Portugal

Ute Klotz, Switzerland
Paula Kotze, South Africa
J. L. Kulikowski, Poland
Masaaki Kurosu, Japan
Cecilia Sik Lanyi, Hungary
Marta Kristin Larusdottir, Iceland
Barbara Leporini, Italy
Gitte Lindgaard, Canada
Zhengjie Liu, China
Vaclav Matousek, Czech Republic
Monique Noirhomme, Belgium
Philippe Palanque, France
Kari-Jouko Räihä, Finland
Carmen Santoro, Italy
Lucio Davide Spano, Italy
Gerhard Weber, Germany
Janet L. Wesson, South Africa
Chui Yin WONG, Malaysia

Table of Contents

Usable Systems

Social and Cultural Problems

Mobile and Web Applications

Epistemology, Emotions and Personalization

Posters

The Power and the Puzzles of Auditory Interfaces

Penelope M. Sanderson

The University of Queensland
St Lucia, Queensland, Australia 4072
psanderson@itee.uq.edu.au

Abstract. Auditory interfaces are increasingly prevalent in work and everyday environments. I survey recent uses of non-speech auditory interfaces and advances in knowledge about them, highlighting research from The University of Queensland.

Keywords: Auditory interface design, auditory displays, distraction, multimodal displays, multisensory integration, individual differences, music.

1 Overview

Auditory interfaces are present in work and everyday environments as diverse as stockmarket trading, web traffic monitoring, patient monitoring, phones and personal assistants, vehicles, defence contexts, and assistive technologies for the visually impaired—I provide examples of recent designs. For the last 20 years in particular, auditory interface design has been stimulated by a multidisciplinary community drawn from computer science, psychology, physics, music, performance, acoustics, and other areas. During that time there has been greater clarity in how to design different forms of auditory interfaces, such as audifications, parameter-mapping sonifications, model-based sonifications, auditory icons, and earcons.

Successful design of auditory interfaces is challenging and requires attention to many issues, some of which are supported with a long theoretical and empirical tradition, and some which are not. Principles of auditory perception and psychoacoustics play a crucial role. Because auditory interfaces can inform or distract, or both, understanding how auditory attention works is also crucial if a display is to work as intended. Auditory interfaces work alongside information from the other senses such as vision, touch, and the vestibular system, so understanding how multimodal information is processed is increasingly important. Because auditory information is naturally heard by all within a certain range, understanding factors underlying people's tolerance of auditory displays is important, as is determining the appropriate scope of social participation in auditory information. Finally, individual differences can affect how well people can use auditory interfaces, yet such differences are seldom tested

In the address I will provide examples from our research on auditory displays that address some of the above questions, supplemented with work from other laboratories. Our main focus at The University of Queensland has been displays for medical

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 1–2, 2010.

contexts, but the principles we have uncovered extend well beyond healthcare. In one branch of our work we have addressed the problem of confusion between alarms. Recently we performed a controlled study of some controversial alarm sounds proposed by Patterson and Edworthy over 20 years ago—the results have important lessons for designers of auditory interfaces. In another branch of our work we have explored the question of how people combine auditory information with visual information from a head-mounted display while they are walking around and performing physical tasks. The results suggest that there is a subtle interplay of vestibular, optic, haptic, and auditory factors, and leave open many puzzles still to solve.

In much of our research, individual differences figure quite strongly. For example, we often find distinct performance differences between listeners with and without musical training, yet research and evaluations of auditory interfaces seldom classify participants in terms of musicality or music training. We have also uncovered differences in how easy and useful listeners at different points on Eyesenck's emphathy/psychoticism scale find it to work with auditory information. If such findings prove to be generalisable, they may have important implications for design of auditory interfaces and for how we deliver auditory information to listeners.

Improving the Search for User Interface Design Patterns through Typed Relationships

Jordan Janeiro[1,*], Simone D.J. Barbosa[2], Thomas Springer[1], and Alexander Schill[1]

[1] Technische Universität Dresden (TU Dresden), Department of Computer Science, Institute of System Architecture, Dresden, Germany
{jordan.janeiro,thomas.springer,alexander.schill}@tu-dresden.de
[2] Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Informatics Department, Rio de Janeiro, Brazil

Abstract. Despite being a set of proven, well-documented, contextualized recommendations for solving frequently occurring user interface design problems, user interface design patterns are still not widely used. We believe this is due to the lack of tools to help designers find patterns and identify how they can be combined to solve user interface design problems. This paper proposes to classify and make explicit the relationships between user interface design patterns. We conducted a small-scale study that indicated that this proposal is more efficient and better accepted by the participants than browsing through a user interface design library.

Keywords: user interface design patterns, semantic relationships, design pattern libraries, user interface design.

1 Introduction

Design patterns are a set of proven, well-documented, contextualized recommendations for solving frequently occurring problems. According to Alexander [1], each pattern describes a recurrent problem in a certain context and “the core of the solution to that problem (...) required to solve the stated problem, in the stated context”. Design patterns have been brought from Architecture [1] to Computer Science, first to the field of Software Engineering [2] as general software design patterns, and then to Human-Computer Interaction [3][4][5][6] as user interface design patterns.

User interface design patterns (UIDP) emerge from successful experiences with the construction and evaluation of user interface and interaction design solutions to certain real-world problems in certain domains and contexts of use. Having been recognized as successful patterns, they are usually collected and organized in pattern libraries, such as Jenifer Tidwell’s pattern collection [7] or the Yahoo! Design Pattern Library [4].

[*] Please note that the LNCS Editorial assumes that all authors have used the western naming convention, with given names preceding surnames. This determines the structure of the names in the running heads and the author index.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 3–14, 2010.

Despite their reputation, UIDPs are still not widespread among software developers. One of the key reasons for that is the lack of tools to support their usage, like for example finding a pattern which solves a certain problem, or even identifying a set of patterns which can be used together to solve a more complex user interface problem.

We believe that a first step towards such a tool would be to provide explicit typed relationships between patterns, allowing designers to quickly identify which of them may or may not be used together, and how to combine them to address more complex problems. In typical UIDP libraries, the types of relationships between patterns are usually left implicit in the UIDP description.

An analysis of the library of van Welie [6], for example, illustrated by Figure 1, shows a high number of relationships between the patterns; each vertex represents a design pattern and the edges represent the relationships between two UIDPs. Such number of relationships demonstrates the possibilities to form many combinations of compositions with UIDPs. However, the creation of the compositions is still not possible, because the current libraries only express the relationships as links between the patterns, and no further information about the meaning of the relationship is described. Thus, whenever a user needs to identify the type of relationship between UIDPs, she needs to read the complete textual descriptions and interpret the type of the relationships that are only implicitly expressed in the description.

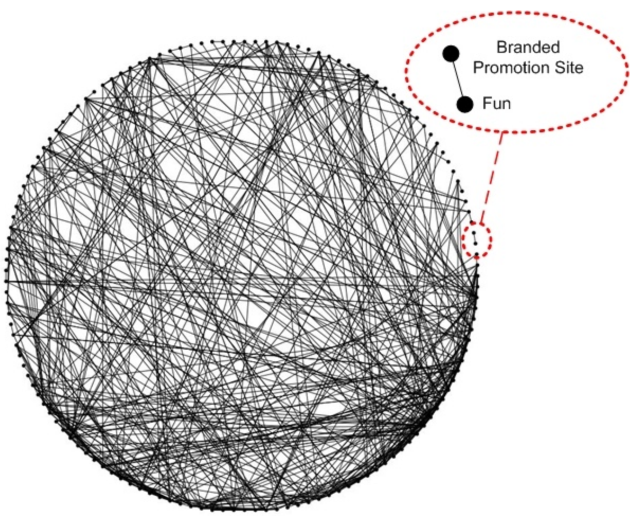


Fig. 1. Graphical Representation of the Relationships between User Interface Design Patterns in a Typical Library

In this paper, we propose some classifications to help identify the relationships between UIDPs (Section 4), making it possible, in future work, to create a tool that uses them to support the user in identifying not only individual patterns, but how they may be combined to handle complex user interface problems.

We have validated the kinds of relationship types we propose (Section 5) asking a group of users to classify pattern relationships found in a UIDP library in the proposed relationship types. Finally, we evaluated the usefulness of such classification (Section 6), requesting a group of users to report their experiences using the typed relationships in searching for UIDPs which should solve a given problem.

2 User Interface Design Patterns

User interface design patterns (UIDPs) are well-documented user interface and user-system interaction solutions for known and frequently occurring user interface problems. These solutions have been shown to benefit end users; they have already been implemented, evaluated, and proven successful.

Currently in the literature there are sets of examples of UIDPs, representing the user's perspective on using a system and the designer's perspective on building a system. Independently of the perspective, both kinds of UI design patterns benefit somehow the system's end user, providing well-established, familiar solutions to widely known problems. Among the major UIDP libraries, we find: Tidwell's patterns for effective interaction design [7], van Welie's pattern library for interaction design [6], and the Yahoo! Design Patterns Library [4]. Some other libraries target a more specific application domain, such as: user interface design patterns for games [8] and for mobile user interfaces [3].

For illustration purposes, let us consider the country selector pattern, described in [6]. Such a pattern belongs to the making choices category and it targets the problem of allowing the user of a website to navigate from the company's global website to a country-specific website, as illustrated by Figure 2.

Fig. 2. Example of the Country Selector User Interface Design Pattern

As sometimes international websites function as a proxy to country-specific websites, it is necessary that the website allow the user to navigate easily to her specific location. It is also important to state that this pattern is not the language selector pattern, which allows users to define the preferred language for the website. To provide a better solution in this case, both design patterns should be combined and used, because users interested in a certain location, like tourists, may not speak its local language.

Defining a user interface design pattern library is already a significant advance on the standardization and definition of best practices for building user interfaces. However, each library uses its own descriptive structure for its patterns, distinct if ever so slightly from one another. There is no standardization here, i.e., currently there is no consensus about a semi-structured standard description for user interface design patterns. For example, in [6] a pattern is described through the problem it targets, the description of the solution, the situation in which it may be used, how it is used and why a designer should use it to solve the targeted problem. In [4] a pattern is described by its title, the problem it solves, its context of use and the solution described by the pattern to solve its problem, without detailing how or why.

Besides the proposal of different structures by the libraries to describe patterns, another problem which arises is the lack of clear meaning of the relationships between patterns. Each pattern is often used together with other patterns; however, none of the design pattern libraries clearly describes how the patterns relate to each other, i.e., the explicit relationships (with clear semantics) between them.

3 User Interface Design Pattern Relationships

The main idea of the UIDPs is to apply the concept of separation of concerns, explored by the software engineering design patterns, creating components of user interface which aim to solve specific user interaction problems. Therefore, with such a concept it may be possible to compose a certain user interface solution through such components, instead of creating it from scratch.

Compositions between UIDPs are already informally documented by the user interface design pattern libraries, allowing identifying which UIDPs are commonly used together in certain situations. For example, it is specified in the van Welie library that the design pattern Accordion is often used within the Main Navigator design pattern.

The Pattern Language Markup Language (PLML) [9] is a proposal to standardize the documentation of user interface design patterns. This language describes important information as: the problems a UIDP solves, the solution the patterns propose and the relationship between the design patterns. This last property of the language is particularly important in the context of this paper, because it represents an explicit description of the relationship between patterns, which in our case is a basic concept to support the composition of UIDPs. An example of these kinds of relationships is described in the next excerpt of code of PLML. In such code, the pattern *Accordion* from van Welie's library, relates to the pattern *Closable Panels* from Tidwell's library.

Example of Relationship Description between the User interface Design Patterns Accordion and Closable Panels

```
<pattern-link type="is-a" patternID="Closable_Panels"
collection="tidwell" label="Closable Panels"/>
```

As illustrated by this example, through PLML it is possible to specify four properties in describing a relationship between patterns: *patternID*, which references the ID of the targeting UIDP from the relationship (Closable_Panels); *collection,* which describes to which library the pattern belongs (tidwell); *label,* which describes the label of the relationship between the patterns (Closable Panels) and *type* (is-a), which describes the semantic of the relationship between the two patterns. However, such feature is still not explored to be used in the current description of UIDPs.

Despite the capability of PLML to describe the types of relationships between UIDPs, the language only supports three types: *is-a*, *is-contained-by* and *contains* [9]. The first type of relationship, despite implying a generalization-specialization hierarchy, actually means that two design patterns are equivalent to each other, meaning that a designer can exchange the use of a UIDP with another in a certain context. The other two types are related to the composition of UIDPs; the *is-contained-by* type refers to a design pattern which is used within another UIDP, whereas the *contains* type refers to the design pattern which uses another UIDP.

On the language specification such a small set of types are may not be enough to describe more precisely the semantics of the relationships between the design patterns. Therefore, we intend to present an extension to these existing types to allow PLML to describe more precise the semantics of the relationships between the design patterns.

4 User Interface Design Patterns' Semantic Relationship

As object of our analysis we used the UIDP Library from Martjin van Welie [6]. Our study consists of identifying the relationships between the UIDPs and to identify other sorts of relationship between design patterns, based on the context they are proposed to be used and their relationship descriptions of PLML.

Based on such analysis we identified seven types of semantics that we propose to extend the current defined types by PLML. They are the following:

- **Used With:** describes that a UIDP x can be used together with another UIDP y and they are not hierarchically related, such as in a composition relationship. Such type of a relationship is illustrated in the library of van Welie by the following text excerpt: "*… Paging is also often used together with a List Builder …*". By this description we can identify that the Paging and List Builder patterns are independent but they are combined in this case through the *used with* relationship.
- **Equivalence:** describes that a UIDP x is like a UIDP y, making it possible to use both to solve the same problem in a certain context. Such type of a relationship is illustrated in the library of van Welie by the following text excerpt: "*… Accordions are alternative to Navigation Tree …*". By this description we can identify that in a certain situation both of the patterns can be used as alternatives to solve the same problem.

- **Similarity:** describes that a UIDP x has some characteristics similar to the UIDP y but, depending on the specific problem they target, one of the patterns should be slightly adapted. "… *the site has got some type of Main Navigation that allows* …". In this description, the similarity relationship is between the Breadcrumbs and the Main Navigation design patterns, which can be used as an alternative to each other in a certain specific context, if some slight modifications are performed.
- **Conditional Equivalence:** describes that a UIDP x may be considered to be equivalent to a UIDP y, if a set of conditions are satisfied. Such type of a relationship is illustrated in the library of van Welie by the following text excerpt: "… *If used for navigation it is (Accordion) conceptually equivalent to Tabs* …". In this description we can identify that the Accordion pattern can also be used as an alternative to the Tabs pattern only if it is used for navigation purpose.
- **Realization:** describes that a UIDP x implements the concepts described by a UIDP y. Such type of a relationship is illustrated in the library of van Welie by the following text excerpt: "… *Accordions can be a good way to implement a Frequently Asked Questions (FAQ)* …". In this case, the FAQ is a recommendation of a functionality which is desirable to have in an application, not an implementation of a concrete user interface component, as in the case of the Accordion pattern. Therefore, we established this kind of relationship between both patterns because the Accordion can implement the functionalities described by the FAQ.
- **Enhancement:** describes that a UIDP x builds upon an UIDP y, enhancing its functionalities. Such type of a relationship is illustrated in the library of van Welie by the following text excerpt: "… *This pattern (Advanced Search) builds on the Search Box pattern by adding some more search options* …". By this description we assumed that the Advanced Search pattern enhances the Search Box pattern with more options, which improves the search function of an application.
- **Conflict:** describes that a UIDP x and a UIDP y must not be used together. Such type of a relationship is illustrated in the library of van Welie by the following text excerpt: "… *Although Accordions are often used as part of a Wizard I strongly recommend against it since it is worse than regular implementations from a usability point of view* …". By such description we understand that although it is really common to use the Accordions pattern as part of Wizard pattern, the author discourage such a practice because the Accordion does not handle properly the navigation issues in Wizard-style applications.

5 Semantic Relationship Assignments

In this section, we describe how we have defined the set of proposed relationship types and the techniques we used to validate them.

Our methodology to extract and evaluate the proposed design pattern relationships consists of two steps: (i) the definition of the semantics of the relationship types, presented in the last section and (ii) the assignment of such semantics to the relationships of the design patterns.

In the first step we analyzed the UIDP descriptions and their references to other patterns, which helped us to select a set of representative labels for classifying the relationships with certain semantics. This list was already presented in the last section as an extension to the element *pattern-link* of the PLML language, to enhance the description of relationships between UIDPs.

In the second step our goal was to validate the proposed relationship semantics we defined. Our approach consisted of proposing to the users to assign the semantics we defined to existing UIDPs´ relationships, already described by the PLML language.

We selected 15 design patterns from the van Welie library to be evaluated. As the library is divided in 15 categories of patterns, we decided to select a UIDP of each of them. All of these patterns together contain 71 relationships, which should be classified by the users. The users should read the text description of a pattern, containing links to other patterns, and classify these links using the proposed relationship types.

We performed the experiment with five users, two men and three women all of whom are computer science professionals, and whose ages vary between 22 and 24 years old.

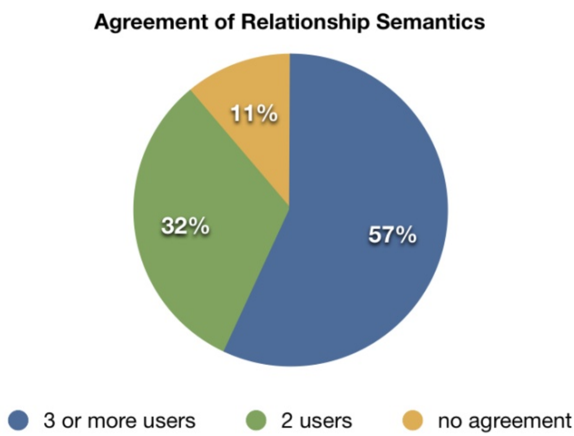


Fig. 3. Agreement of Users in Classifying a Relationship between Two User Interface Design Patterns

Figure 3 presents the agreement, in percentage, of the users in assigning semantics to relationships between design patterns. According to the chart, in 57% of the assignments, three or more users assigned the same type to a relationship; in 32% of the assignments, two users assigned the same type to describe a relationship; and in 11% of the relationships, none of the users' assignments matched.

We performed such experiment with a group of users to obtain the agreement of the majority in assigning certain semantics to the relationships of the design patterns. Therefore, if the majority of the users agree with the semantics to a relationship, it means that there is a higher probability that the chosen semantics represents a suitable term to describe the relationship.

As the assignments are important information to us, we analyzed carefully the three categories we registered in the experiment: the agreement of three or more users, the agreement of two users and no agreement. For the first case, we automatically added the assignments to our database, because the majority of the users who performed the experiment agreed in a common relationship type, which for us means that the assigned meaning sufficiently characterizes that relationship. In the second case, we included our own analysis as an additional assignment, i.e., we only considered the assignments of the two users if we agreed with them. For the relationships in the third case, we analyzed by ourselves the relationships to classify them.

6 Evaluation

In order to evaluate the effectiveness of providing typed relationships in searching and finding user interface design patterns, we prepared two usage scenarios.

In the first scenario, the participant should build a form for booking flights, and in the second scenario, the participant should develop a form for registering personal information in an e-commerce web site. For each of these scenarios, participants should select the user interface design patterns which they thought were the most suitable to build their user interfaces. For each scenario, participants had to search for the UIDPs using a different resource. We selected two different resources for the experiment: van Welie's user interface design pattern library "as is" (*Plain Library*); and a tool we implemented to help browse the library using the typed relationships between patterns (*Navi Browser*).

We selected a group of 8 participants, 6 men and 2 women, all of them professional computer scientists.

To reduce the bias from the order in which the tasks were performed and the resources used, we have distributed the participants in four groups, varying the order of tasks and resources, as presented in Table 1.

Table 1. Task and Resource Groupings Used to Reduce Bias in the Experiment

First Performed Task	Second Performed Task
Flight Booking / Plain Library	Personal Registration / Navi Browser
Flight Booking / Navi Browser	Personal Registration / Plain Library
Personal Registration / Plain Library	Flight Booking / Navi Browser
Personal Registration / Navi Browser	Flight Booking / Plain Library

After performing the proposed tasks for our evaluation, the participant should answer a few questions about their experience with each resource. Each question could be answered using a 5-point scale, ranging from 1 (lowest/worst) to 5 (highest/best). The questions proposed in the form are following:

- Q1: How easy was it to find the UIDPs using the proposed application?
- Q2: How adequate were the UIDPs to the user's solution?
- Q3: How easy was it to identify the relationships between the UIDPs?

- Q4: How accurate were the relationships between the UIDPs?
- Q5: How useful were the relationships between the UIDPs?
- Q6: How much time do you think the task took to be accomplished?
- Q7: How much did the application help to reduce the time to obtain the UIDPs to design the solution?
- Q8: What is your overall impression of using the proposed method to find the UIDPs?

From the experiments, we could identify that all of the users achieved the patterns for their tasks following a certain procedure: they searched first, in the list of all UIDPs, for patterns whose title contained keywords for the problems they needed to solve, selecting these for later detailed analysis; then, they analyzed all of the patterns linked to the ones within their list. Therefore, in the context of this experiment, the participants identified first the patterns *Forms* and *Booking* for the flight booking scenario, and *Forms* and *Registration* for the personal registration scenario. From there, participants identified additional patterns through the typed relationships, as presented in Table 2.

Table 2. User Interface Design Patterns Selected in Each Scenario

Flight booking		Personal registration	
Forms	Booking	Forms	Registration
Advanced Search Constraint Input Grid-Based-Layout Input Error Message Login Wizard	Advanced Search Homepage Processing Page Purchase Process Search Results Wizard	Advanced Search Booking Constraint Input Grid-Based-Layout Input Error Message Login	Form Login Product Recommendation Shopping Cart Wizard

We analyzed the collected data to understand the experience of the users in searching for user interface design patterns connected through the typed relationships. Figure 4 presents in a chart the comparison of the user satisfaction in using both applications for this experiment, meaning indirectly a comparison between the user satisfaction in making use of the semantics of the relationships and not.

The chart shows that using typed relationships was consistently rated better than the plain library. Questions 3 and 5 present valuable information to our approach, because they specifically evaluated the participants' satisfaction in using the typed relationships for searching design patterns. In the answers for these questions, the score for our approach is quite higher than the plain library, suggesting that the use of typed relationships makes the search process easier. During the interviews, the participants mentioned that they assigned a high score to our approach because they did not have to read the entire descriptions of the design patterns to understand the way they relate to each other, thus, they could save time in interpreting the meaning of the relationships and deciding whether other related design patterns should also be considered in the solution.

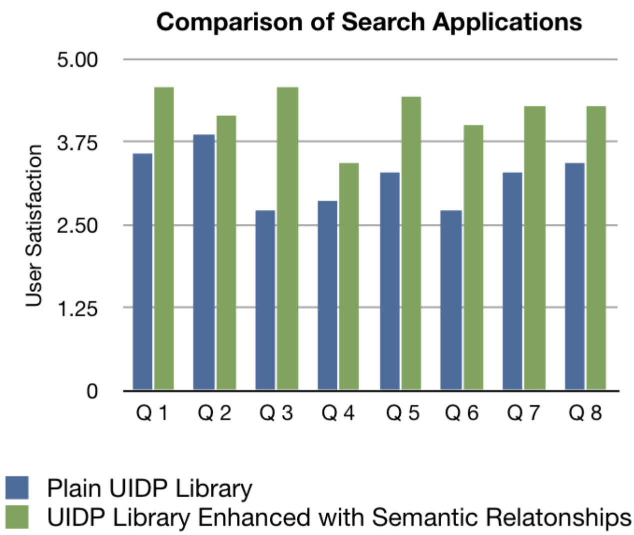


Fig. 4. Evaluation of Using Typed Relationships to Search for User Interface Design Pattern

7 Related Work

The Pattern Language Markup Language (PLML) [9] was developed to describe user interface design patterns in a standardized way, through a set of attributes. Van Welie has used PLML in his library. As we have seen, the types of relationships between patterns are very limited in PLML. Moreover, there is no established research work that proposes to use the information described by the semantic relationships to support the development of a search mechanism to retrieve design patterns more efficiently. In our approach, we have defined these new types of relationships between patterns as a first step towards such a search mechanism.

Previous efforts have investigated the use of user interfaces guidelines to support the process of developing applications [10]. However, the core problem of using guidelines is still the lack of tools and methods for describing and validating them, specially regarding to high-level guidelines. Our approach combines the contextualized information provided by the user interface design patterns with clear semantics of their relationships that can be used to develop a tool to support using patterns in user interface design.

Our underlying vision is somewhat akin to Semantic Web ideas introduced in [11]. In such work, the authors discuss the ideas of enhancing the Web through the attachment of additional information that describes the semantics of a resource in the Web. In the realm of web services, the World Wide Web Consortium (W3C) has proposed a standard for the Semantic Annotations for WSDL (SAWSDL) [12]. Such standard defines a set of attributes to annotated Web Services to better describe their semantics. Although such approach allows the attachment of any kind of semantic descriptions, the most common use and support focuses on the functionality descriptions of web services, and there is no support specifically defined to describe the user interfaces aspects associated to it. In our approach, we share the ideas introduced in the semantic description of web services to perform more refined searches, but our main goal is to model and support information related specifically to user interface issues, both problems and solutions.

Regarding the execution of semantic searches, there are many existing semantic service matchmakers which perform discovery processes based on the attached semantic, by non-logic, logic or hybrid matching techniques [13]. In future work, we intend to extend such matchmakers to perform matches of user interface design patterns to a specified user interface problem described by the user performing a search.

8 Conclusion

This paper proposes a new set of semantic types which enhance the description of user interface design pattern relationships. The proposed types were used by a group of participants to describe different kinds of relationships between UIDPs. We then conducted an experiment with another group of participants, in which we have identified the effectiveness of using the typed relationships to help search for design patterns. In our experiment, we compared two kinds of resources for finding user interface design patterns: without explicitly typed relationships, as it is currently available in the design pattern libraries, and with the explicitly typed relationships. The experiment showed that the typed relationships provide a better support to the search, because they make it easier and more efficient to determine the relationship between UIDPs, and thus to identify which sets of design patterns are suitable to solve a certain user interface problem, and how they can be combined.

We believe that our proposal of typed relationships can further be used to implement a new kind of UIDP search engine. Instead of browsing a design pattern library, interpreting and selecting compositions of UIDPs, the user should be able to perform such tasks semi-automatically; the search engine we envision will let the user specify a keyword-based query to describe her user interface problem, and the engine will be able to perform the search for such compositions. The goal is that the search engine customizes the recommendation of a set of design patterns for each situation, leveraging thus a new supported method for solving user interface problems.

Acknowledgments. We thank our colleagues at TU Dresden and at PUC-Rio for invaluable discussions related to this work. Simone Barbosa thanks the financial support of CNPq (557.128/2009-9) and FAPERJ (E-26/170028/2008). Jordan Janeiro, Thomas Springer and Alexander Schill thank the supporting of project CRUISe and the financial support of the Federal Ministry of Education and Research of Germany (BMBF).

References

[1] Alexander, C., Ishikawa, S., Silverstein, M.: A Pattern Language: Towns, Buildings, Construction. Oxford University Press, Oxford (1977)

[2] Gamma, E., Helm, R., Johnson, R.E.: Design Patterns. In: Elements of Reusable Object-Oriented Software. Addison-Wesley Longman, Amsterdam (1995)

[3] Mobile User Interface Design Patterns, http://patterns.littlespringsdesign.com/index.php/Main_Page
[4] Yahoo! Design Pattern Library, http://developer.yahoo.com/ypatterns/
[5] Borchers, J.O.: A Pattern Approach to Interaction Design. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 369–378. ACM, New York (2000)
[6] van Welie, M.: Welie.com - Patterns in Interaction Design, http://www.welie.com
[7] Tidwell, J.: Designing Interfaces: Patterns for Effective Interaction Design. O'Reilly Media, Inc., Sebastopol (2005)
[8] Bjork, S., Holopainen, J.: Patterns in Game Design, Charles River Media (2004)
[9] Fincher, S.: Perspectives on HCI Patterns: Concepts and Tools, Introducing PLML (2003)
[10] Cohen, A., Crow, D., Dilli, I., Gorny, P., Hoffman, H., Iannella, R., Ogawa, K., Reiterer, H., Ueno, K., Vanderdonckt, J.: Tools for Working with Guidelines. SIGCHI Bull. 27, 30–32 (1995)
[11] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American Magazine
[12] Kopecky, J., Vitvar, T., Bournez, C., Farrell, J.: SAWSDL: Semantic Annotations for WSDL and XML Schema. IEEE Internet Computing 11, 60–67 (2007)
[13] Klusch, M.: Semantic Web Service Coordination. In: CASCOM: Intelligent Service Coordination in the Semantic Web, pp. 59–104. Birkhäuser, Basel (2008)

Contributions of Psychology to the Design of Diagnostic Decision Support Systems

Gitte Lindgaard[1], Janette Folkens[1], Catherine Pyper[1],
Monique Frize[1], and Robin Walker[2]

[1] Carleton University, Ottawa, ON K1S 5B6, Canada
[2] IWK Health Centre, 5850/5980 University Avenue, Halifax, NS B3K 6R8, Canada

Abstract. This paper describes how psychological research can contribute to the requirements engineering, the design and usefulness of a Diagnostic Decision Support System (DDSS) intended to support pediatric residents' diagnostic decisions. Research on cognitive biases in Bayesian decision tasks is discussed. The design of the DDSS is briefly outlined, and a formative usefulness test is reported. Under the assumption that a particular cognitive bias could be overcome by showing it to participants, pediatric residents were given a set of Bayesian decision tasks. One half was given an opportunity to interact with NeoPeDDS and the other half was not. Results showed that NeoPedDDS usage improved the accuracy of the diagnostic decisions, but that formal training in Bayesian statistics appears to be necessary for residents to evaluate ambiguous information in a normatively correct manner.

Keywords: Bayes' Theorem, human decision making, probabilities, Diagnostic Decision Support Systems, diagnosticity.

1 Introduction

Numerous tools and techniques have evolved to support usability engineers in the user requirements gathering process [1,2,3]. One assumption in most of these is that the target users are already known along with their background, tasks and goals, the tools they currently use, and the context in which they will use the new application. It is also often assumed that these techniques suffice to derive the detailed understanding needed to design an effective application. Similarly, many knowledge elicitation techniques exist in Artificial Intelligence [4] to elucidate experts' mental models underlying their tacit problem solving and decision processes in complex, often highly dynamic, situations. Consequently, the design of many expert systems rests on the assumption that human performance is inevitably optimal and hence worth emulating [5,6]. This paper describes how understanding derived from fundamental psychological decision research was applied to the initial design of a prototype for a diagnostic decision support system in neonatology.

The next section introduces decision issues in medical diagnosis, highlighting the need for decision support systems to assist physicians processing ambiguous diagnostic information. It is followed by a brief discussion of the most popular medical decision support systems, and then by an outline of how probabilities relate to medical

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 15–25, 2010.

diagnosis in Bayesian decision tasks. The notion of diagnosticity is then discussed, leading into a short description of the creation of the decision support system called NeoPeDDS, and subsequently, the usefulness test of NeoPeDDS. Finally, the next steps in this research program are outlined, and concluding remarks are made.

2 Decision Issues in Medical Diagnosis

As the queues of patients needing urgent medical attention are growing rapidly worldwide, medical practitioners are under increasing pressure quickly to formulate a diagnosis and initiate treatment. Yet, diagnostic decisions are very complex. Patients rarely present with a clear, unambiguous clinical picture as described in medical textbooks. Many patients suffer from multiple diseases, and they may not know which symptoms are most important to report. In addition, many of the symptoms patients display may be ambiguous, equally indicative of different diseases. For neonatologists who deal with newborn babies, this complexity is further exacerbated because sick infants cannot say how they feel or where it hurts. Neonatology is therefore an attractive domain in which to provide effective computer-based Diagnostic Decision Support Systems (DDSSs).

Given this complexity, it is not surprising that misdiagnosis is a recognized problem in medicine. Figures released by the Institute of Medicine (IOM) on the annual number of deaths of hospitalized patients due to some kind of medical error are estimated to lie between 98,000 [7] and 115,000 [8]. It is believed that many of those deaths could be prevented [9,10,11]. However, figures vary widely. One recent review [12] found diagnosis-related errors to account for 10-30% of all errors recorded. Others estimate the proportion to be up to 76% [13]. Figures from autopsies have consistently yielded a misdiagnosis rate of 40% over the past 65years [6], but autopsies are no longer conducted routinely. The accuracy of recent figures is therefore debatable as are both the definition and the calculation of “preventable error” [12]. However, regardless of how these are counted, the number of misdiagnoses is high. The reasons are said to range from macro-level health system related problems [13] to micro-level cognitive errors [9]. Empirical evidence suggests that medical diagnosticians find it difficult to deal with ambiguous data [14]. This was the problem addressed in the design of the prototype introduced here, called NeoPeDDS. It was informed by research highlighting some of the cognitive biases to which human decision making, including medical diagnostic decisions, are prone.

3 Decision Support Systems in Pediatrics

Several DDSSs supporting pediatric decision making have emerged. The most popular of these are ePocrates and Isabel. ePocrates is an impressive DSS based on a comprehensive database of diseases, treatments, drugs, and much more [5]. It runs on virtually any mobile device including iPods and the BlackBerry. The online version provides instant access to a wide range of information and services. The database is updated regularly. It enables the diagnostician to compare the occurrence of symptoms in different diseases. However, it does not provide frequencies of occurrence of symptoms, making it impossible for the physician to use the information to weigh the probabilities of different diseases against one another. Isabel is a web-based natural-language DDSS that aims to reduce diagnostic errors. It applies word-matching searches through unformatted medical texts to arrive at a list of diagnoses. It provides a list of possible diagnoses in response to symptoms and other user-entered clinical findings [15,16]. Several studies, e.g. [17] and [18] have showed that Isabel can lead pediatricians to diagnoses they would otherwise not have considered. However, the output can be overwhelming, as Isabel provides up to 10 diagnostic categories, each of which may point to up to 35 different diseases. Each result links to a Knowledge Mobilizing System that allows perusal of medical texts about a particular disease. This can distract from the purpose of arriving at a final diagnosis quickly. Although it is popular, Isabel does not exactly yield quick diagnostic aid in a high-pressure clinical setting. Rather, it provides supplemental reading material for a more leisurely approach to diagnostic decision making where timing is not critical. One major drawback is that Isabel's suggested diagnoses appear implicitly to be equiprobable because it lacks quantified information about symptom diagnosticity. Its database is drawn from a cross-section of medical texts which do not quantify symptoms. Therefore, there is no way to calculate the frequency of occurrence of symptoms or the probability associated with different diseases in the light of the clinical picture a given sick infant presents. Such an approach continues to force clinicians to rely on personal experience and various decision heuristics when diagnosing ambiguous cases.

4 Probabilities and Bayesian Decision Tasks

The notion of probability is connected with the degree of belief warranted by evidence (epistemic probabilities) and with the tendency to produce stable relative frequencies (aleatory probabilities). Statistical probabilities concern the way evidence from various sources is combined into a numeric statement irrespective of the judge's belief. Epistemic probabilities incorporate an assessment of the judge's personal belief, generated from autobiographical experience and state of knowledge about the evidence. The human-generated epistemic probability reflects both arithmetic calculations and degree of belief. By contrast, a computer-generated statistical probability is an arithmetic computation of given numeric values. It is therefore unrealistic to expect the two to be identical. Subjective beliefs are more likely to attenuate than to increase judgmental accuracy because beliefs are derived from the judge's own experience.

The output of a Bayesian analysis is a distribution of probabilities over a set of hypotheses. The model is normative in the sense that it specifies certain internally consistent relationships among probabilistic opinions that prescribe how opinions should be revised with new incoming information. Existing knowledge is summarized in prior (aleatory) probabilities, the base rates, and incoming case-specific evidence provided through individuating information. The outcome of a Bayesian analysis, the posterior probability, is calculated by combining the base rates and the individuating information. Two hypotheses, H and $\hat{H}$, assessed against one another, are expressed in the base rates such that $P(H) + P(\hat{H}) = 1.0$. The model demands that the individuating information be considered in terms of its support for both hypotheses, the weighting of which leads to the posterior probability. This weighting results in a revision of the opinion contained in the original base rates. When the evidence supports both hypotheses H and $\hat{H}$ to an equal extent, no revision should occur. The resulting posterior probability is therefore identical to the base rate representing the hypothesis in terms of which the judgment is made.

Numerous early studies in Bayesian decision making led researchers to conclude that people are, by and large, good Bayesians [19], except that they tend to revise their judgments less than demanded by the model upon receiving additional case-specific information. They were “conservative” [20]. Numerous subsequent findings refuted that early belief, showing instead that people did not behave in a Bayesian manner at all [21,22]. People were found to ignore the base rates and instead rely exclusively on the individuating information, even when that information was completely nondiagnostic. That is, it equally supported both hypotheses or none of these. Objectively, such information should be ignored; judgments should rely exclusively on the base rates.

5 The Notion of Diagnosticity

Diagnosticity refers to “how much potential impact a datum should have in revising one’s opinion without regard to what the prior odds are” [22, p.778]. In order to determine the informativeness (diagnosticity) of the individuating information in cases where it consists of several items, a value must be assigned to each item. Early studies revealed a robust tendency of people to rely on the degree to which the individuating information was representative of the to-be-judged hypothesis. Fischhoff and Beyth-Marom’s [23] and Beyth-Marom and Fischhoff’s [24] explanation for this reliance on representativeness is that people judge the evidence solely by the degree to which the individuating information supports the hypothesis being entertained. They argue that people simply do not understand that diagnosticity is a measure of the relative support for both hypotheses, H and $\hat{H}$. The concept of diagnosticity is touched upon in some recent studies in the medical domain [7,24,25]. However, its importance for leading to a more accurate estimation of the posterior probabilities in the face of an ambiguous clinical picture is not strongly emphasized. Studies involving nondiagnostic individuating information have primarily focused on posterior probability estimates. By contrast, numerous studies in our lab focusing on nondiagnosticity in occupationally relevant Bayesian tasks have repeatedly confirmed the tendency for probability estimates to resemble judgments of representativeness [46-50]. Using very carefully constructed case-specific information in which the symptoms displayed by fictitious patients suffering from one of two competing diseases, our experiments have been conducted with both nurses and physicians and a range of different diseases. Medical texts provide lists of diseases and symptoms that are likely to be observed in each disease, but they do not quantify symptoms. The fact that many symptoms may occur in different diseases is not obvious. Therefore, it was necessary to pre-test symptoms on different samples of participants to ensure that highly diagnostic symptoms were perceived to occur very frequently in one, and very infrequently in the other of the two diseases to be exposed in Bayesian tasks.

Our results have consistently revealed a primacy effect. That is, the symptom shown first in a vignette was invariably found disproportionally to affect the probability estimate. Thus, a high-diagnostic symptom supporting the to-be-judged disease presented first resulted in a significantly higher probability estimate when the same symptom was presented later. Conversely, a high-diagnostic symptom supporting the alternative disease presented first suppressed the probability estimate, with a decreasing effect when presented later. The primacy effect suggests a tendency to focus on a particular diagnosis very early in the process. Graber [27] claims that "knowledge deficits are rarely the cause of cognitive errors in medicine; these errors more commonly involve defective synthesis of the available data" (pp. 1-2). This concurs with the suggestion that diagnosticians select a single, very salient symptom right away and use it as a pivot around which they collect additional information. Such a strategy could bias the integration of information in ambiguous cases, leading to "premature closure" whereby other possible diagnoses are not considered once a hypothesis is entertained [27]. For example, the judge may ignore available data that conflict with the current hypothesis; the fact that the selected pivot may point to different diseases may not be detected if only one hypothesis is entertained. A more detailed description of these experiments may be found in Lindgaard et al. [28].

6 Creating NeoPeDDS

In terms of supporting the task of diagnosing, we assumed that a display of the most likely diseases along with their respective probabilities would raise awareness of the possibility that more than one disease could account for the constellation of a patient's symptoms. Thus, upon entering at least one symptom into NeoPeDDS and telling the system to 'diagnose', it generates a list the five most probable diagnoses, complete with the relevant probabilities. Five was chosen as the maximum to keep the diagnostic task manageable. The diagnoses are based on the World Health Organization's International Classification of Diseases (ICD10) and on a Bayesian analysis of epidemiological data. The database comprises 97 complete records collected from some 1,200 infants admitted to the neonatal intensive care unit at the Children's Hospital of Eastern Ontario and diagnosed with respiratory distress. Respiratory distress was selected as the target condition because it occurs relatively frequently and because the signs and symptoms are ambiguous, pointing to different possible causes. The dataset enabled accurate quantification of the relative diagnosticity of each sign and symptom associated with every causal condition upon which $P(D|H)$ and $P(D|\hat{H})$ were calculated.

NeoPeDSS was developed by using Object Oriented Software Development (OOSD) and Usage Centered Design [1]. OOSD discusses requirements in terms of a use-case model, which consists of actors and use-cases. These were developed to define the possible sequences of system-actor interactions; use-case diagrams modeled the system requirements and boundaries using actors and use-cases to improve the breakdown of the system according to the user requirements. An abstract representation of the user interface was then designed. A navigation map tied the use-case narratives to a flow between interaction contexts. A content model as well as a navigation map was created from the essential use-cases and their relationships before designing the GUI prototype which was used to assess our assumption that a mere display of a set of diseases would suffice for pediatric residents to eliminate the primacy effect observed in our earlier experiments.

Usability was assessed of the early prototype before recruiting pediatric residents in the formative usefulness test. Thus, a heuristic evaluation was conducted by two usability experts, and two empirical usability tests were performed using HCI students who were naïve with respect to both the system and pediatrics. The task scenario exposed the core system tasks such as entering a patient case and retrieving a list of possible diagnoses as well as modifying the patient case. Problems revealed resulted in four relatively minor modifications to the prototype: (1) the layout of button locations was modified to separate the 'Diagnose' function from the 'Delete' functions to prevent accidental deletion of cases still being worked on; (2) tool tips were added; (3) a visual cue was added to indicate where to enter data, and (4) a history feature was added enabling users to retrieve an earlier case and compare it with the current results. The GUI is shown in Figure 1.

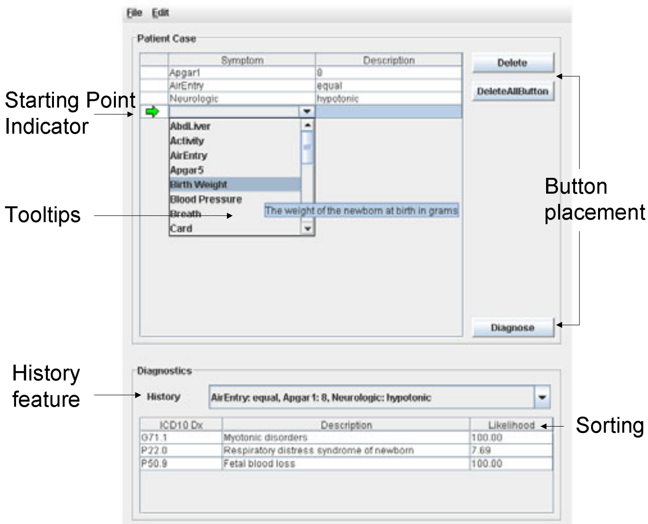


Fig. 1. NeoPeDDS GUI

Users are able to add more symptoms to refine the probabilities even after the 'Diagnose' button has been pressed. As more information is added, the relevant probabilities are adjusted accordingly. The 'Diagnose' button may be pressed as many times as the diagnostician likes.

7 Formative Usefulness Test

Three major issues were addressed in the formative test in which resident pediatricians took part. First, we needed to demonstrate that NeoPeDDS could improve diagnostic accuracy. To test that, five test cases were prepared for which participants proposed a preliminary diagnosis before, and a final diagnosis after, using NeoPeDDS. Accordingly, Hypothesis 1 predicted that more correct final diagnoses would be found after than before NeoPeDDS. Second, it tested the assumption that the display of the five most probable diagnoses would increase awareness of the possibility that more than one disease should be considered. Hypothesis thus 2 predicted that base rates would be used after, but not before, exposure to NeoPeDDS. Third, to recognize the relative worthlessness of the individuating information, participants must consider both diseases. Doing so should eliminate the primacy effect found in previous studies. Hypothesis 3 therefore predicted that a primacy effect would be found in probability estimates made before, but not in those made after, exposure to NeoPeDDS.

7.1 Method

Participants: Some 40 senior resident pediatricians were recruited from various university hospitals in Canada and the United States. NeoPeDDS was presented online, enabling the participants to complete the study in their own time and in several steps if they chose. Once a case had been evaluated, they were unable to go back over it. Upon completion of the test, they received a $100 gift certificate by email.

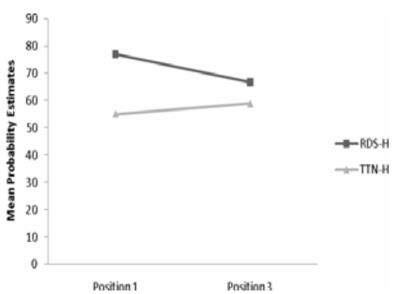
Materials: A fictitious cover story was created to provide a plausible explanation for the limited information available about each infant to be assessed. The 24 vignettes, each describing a sick infant and containing three symptoms, were constructed such that they were nondiagnostic or near-nondiagnostic. Each vignette contained three signs or symptoms: one, either high- or low-diagnostic, supporting Respiratory Distress Syndrome (RDS-H), another one supporting the alternative disease $\hat{H}$, Transient Tachypnoe, (TTN-H) and the third was nondiagnostic (e.g. runny nose). In nondiagnostic vignettes, both the diagnostic symptoms were either high or low in diagnosticity. In the near-nondiagnostic vignettes, one was high, and the other low in diagnosticity, yielding HL and LH vignettes. The three symptoms were combined factorially. Participants estimated the probability of RDS-H, disease H. Another five cases were described in more detail. These required the pediatricians to propose a preliminary diagnosis before using NeoPeDDS, and a final diagnosis afterwards. NeoPeDDS was not available for their preliminary diagnosis.

Design: One half of the participants were assigned at random to the Low Base Rate group (LBR) where the RDS-H base rate was low (28/100 fictitious cases) and that of TTN-H was high (72/100). This was reversed for the other half, assigned to the High Base Rate group (HBR). The test comprised three phases as well as some pre- and post-test questions. In Phase 1 participants estimated the probability of RDS-H in each of the 24 vignettes. In Phase 2 one half (half LBR and half HBR) were shown NeoPeDDS and how it worked as well as being given an opportunity to play with it before seeing the five cases to be diagnosed. The other half read an article on diagnostic error. In Phase 3, all participants again assessed the same 24 vignettes, presented in a different random order, with different, randomly assigned names. Finally, they indicated the perceived frequency of occurrence of each of the signs and symptoms exposed in the vignettes.

Procedure: Once a pediatrician had agreed to participate, a unique login code was emailed to them. Upon logging into the private and secure site, they completed an informed consent form before gaining access to the test. They then answered the pre-experimental questions and proceeded to Phase 1. They were told that they could log out at any time during the test. If they chose to complete the test in stages, they were told that the program would remember where exactly they had been before logging out and it would take them back to the next task. Upon completion of the entire test, they were presented with a debriefing form, which they could print out, thanked for their time, and advised that a gift certificate would be sent to the email address to which their login code had been sent. All results were sent automatically to a server at Carleton University.

8 Results and Discussion

A comparison of the number of correct preliminary and final diagnoses proposed in Phase 2 was significant ($t(19) = 2.93$, $p < .008$), thereby supporting Hypothesis 1. To test Hypothesis 2 predicting that base rates would be used in judgments of the 24 vignettes after (Phase 3), but not before exposure to NeoPeDDS (Phase 1), a 2 x 2 x (2) mixed-design ANOVA was conducted for base rate groups (HBR, LBR), exposure to NeoPeDDS (exposure, no exposure), and experimental Phase (1, 3). This analysis should ideally have yielded a two-way interaction of Phase and exposure as well as a main effect for exposure and Phase respectively. None of these predicted effects occurred ($F < 2$), thereby refuting Hypothesis 2. However, contrary to the expectation that the base rates would not affect probability estimates before exposure to NeoPeDDS, there were no interactions involving the base rate groups, and the main effect for base rate group was highly significant ($F(1,36) = 15.85$, $p < .000$). This suggests that base rates were used to some extent even before exposure to NeoPeDDS, but that this usage did not increase after exposure. The result is puzzling, as our previous research has consistently shown base rate neglect in many similar experiments [26]. Since roughly 75% of the participants were not familiar with formal decision models, this finding cannot be attributed to prior knowledge of Bayes' Theorem. Other researchers have [26] shown that people relied exclusively on the base rates when the nondiagnosticity of the individuating information was palpably worthless. As the pediatricians adjusted their probability estimates less than Bayes' Theorem demands, the evidence was apparently not deemed palpably worthless.



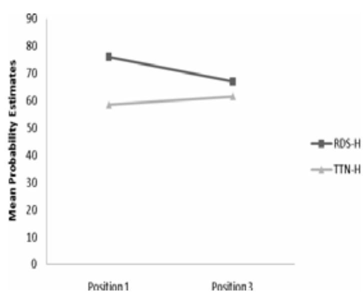


Fig. 2. The serial position effect in Phase 1 (left panel) and in Phase 3 (right panel)

To test Hypothesis 3, predicting that a primacy effect would be found in the probability estimates made before, but not after, exposure to NeoPeDDS, a repeated-measures ANOVA was conducted separately for phases 1 and 3. This was justified by a non-significant t-test comparing estimates of participants who had been exposed to NeoPeDDS with those who had not ($t < 1$). The main effect for serial position was significant both for Phase 1 ($F(1, 39) = 5.45$, $p < .05$) and Phase 3 ($F(1, 39) = 10.86$, $p < .01$), as were the interactions of symptom and serial position ($F(1, 39) = 10.86$, $p < .01$ Phase 1; $F(1, 39) = 28.89$, $p < .001$ Phase 3). Figure 2 shows that estimates were higher for RDS-H in serial position 1 than in serial position 3, and that the reverse was true for TTN-H estimates, resulting in a clear convergence of estimates across serial position. Both of these findings indicate the presence of a primacy effect. As such an effect had been predicted for Phase 1 but not for Phase 3, the results partially supported Hypothesis 3.

One explanation of the persistent primacy effect is that individuating information is considered only in terms of the nominator, here RDS-H, as Beyth-Marom and Fischhoff [29] claimed, because they do not understand the concept of diagnosticity. Another possibility is that participants simply weighed information confirming RDS-H more heavily than information disconfirming it, perhaps because they did not know how to deal with conflicting data in the Bayesian framework. The data are insufficient unequivocally to discern which of these possibilities may account for the results. In order to determine if they were able to assess the frequency of occurrence of the different symptoms, they were given two lists at the end of the test. For one list they were asked to indicate in how many infants out of 100, all diagnosed with RDS-H, they would expect to find each symptom. For the other list they were asked the same question, but this time the infants were said to have been diagnosed with TTN-H. The lists were identical, and they both displayed all the symptoms that featured in the short cases, albeit shown in different random orders in the two lists. The findings showed that participants were fully aware of the relevant frequencies of occurrence. Thus, for example, the H-diagnostic symptom supporting RDS-H was seen to occur very frequently in RDS-H and very infrequently in disease TTN-H, and vice versa for the H-diagnostic symptom supporting TTN-H. Apparently, they were sensitive to symptom diagnosticity, but they did not know how to combine symptoms pointing to the two diseases into a normatively correct judgment. It would therefore appear that at least some training is necessary for a DDSS based on a Bayesian algorithm such as NeoPeDDS to provide optimal assistance to physicians. One most encouraging feedback was that all but a single participant said that they would use a DDSS such as NeoPeDDS if it were made available to them.

9 Conclusions and Next Steps

The above findings suggest that NeoPeDDS did facilitate the task of diagnosing to some extent. However, awareness of the possibility that several diseases may account for a highly ambiguous clinical picture, did not suffice for participants to utilize the base rates optimally in their probability estimates. People may generally have a poor understanding of the concept of diagnosticity because they do not understand the relevance of the denominator term, $P(D|\hat{H})$ for the posterior probability, $P(H|D)$. Eddy [29] has shown that physicians have difficulties distinguishing between the terms $P(H|D)$ and $P(D|H)$. It is conceivable that this difficulty extends to the necessity of estimating $P(D|\hat{H})$ even when clinicians are capable of estimating the frequency of occurrence of individual symptoms as was the case here. The above data are insufficient to determine participants' understanding of diagnosticity, as they could have relied either on the absolute frequency of occurrence of the symptoms under RDS-H (H), or on the relative difference in frequency of occurrence under both competing hypotheses, RDS-H and TTN-H (H and $\hat{H}$). Either approach would affect the estimates in a similar manner because the H-and L- diagnostic symptoms differed along both dimensions. A H-diagnostic symptom was high in absolute frequency of occurrence under the hypothesis it supported as well as in the difference in frequency of occurrence under both hypotheses. Similarly, a L-diagnostic symptom was low in both absolute and relative frequency of occurrence. One way to test people's understanding of the concept of diagnosticity could be to present problems where

$$P(D|H) = 0.85 \text{ and } P(D|H_1) = 0.83$$
$$P(D|H) = 0.04 \text{ and } P(D|H_1) = 0.02$$
$$P(D|H) = 0.85 \text{ and } P(D|H_1) = 0.20$$

If the concept is not understood correctly and people rely only on the absolute frequency of D under Hypothesis H, the resulting P(H|D) should be approximately equal for (a) and (c) but lower for (b). If people rely on the difference in frequency of occurrence of D under both hypotheses H and $\hat{H}$, the resulting P(H|D) should be approximately equal for (a) and (b) but higher for (c). If P(H|D) is calculated in a normatively correct manner, taking both the absolute and the relative frequency of occurrence into account, then (c) should be highest, followed by (b) and (a). This will be tested in a future experiment. Finally, we will add a short training module showing how Bayes' Theorem works, and add more practice examples. This will be tested independently.

Acknowledgements

We thank Dr. Satid Thammasitboon for reviewing and adjusting the cases and for giving us access to participants. and all the student and pediatric resident participants who so willingly took part in the usefulness test.

References

1. Constantine, L.L., Lockwood, L.A.D.: Software for use: A practical guide to the models and methods of usage-centred design. Addison-Wesley, Reading (1999)
2. Mayhew, D.: The usability engineering lifecycle: a practitioner's handbook for user interface design. Morgan Kauffman, San Francisco (2003)
3. Preece, J., Rogers, Y., Sharp, H.: Interaction design: Beyond human-computer interaction, 2nd edn. John Wiley & Sons Ltd., Hoboken (2007)
4. Ford, D.N., Sterman, J.D.: Expert knowledge to improve formal and mental models. System Dynamics Review 14, 309–340 (1998)
5. Borra, R.C., Andrade, P.D., Corrêa, L., Novelli, M.D.: Development of an open case-based decision-support systemd for diagnosis in oral pathology. Europran Journal of Dental Education 11, 87–92 (2007)
6. Goggin, L.S., Eikelboom, R.H., Atlas, M.D.: Clinical decision support systems and computer-aided diagnosis in otology. Otolaryngology 136, 521–526 (2007)
7. Hughes, C.M., Phillips, J., Woodcock, J.: How Many Deaths Are Due to Medical Errors? JAMA 284, 2187–2189 (2000)

8. Miller, R.M., Elixhauser, A., Zhan, C.: Patient safety events during pediatric hospitalizations. Pediatrics 111, 1358–1366 (2003)
9. Berner, E.S., Graber, M.L.: Overconfidence as a cause of diagnostic error in medicine. American Journal of Medicine 121(SA), 2–23 (2008)
10. Kohn, L.T., Corrigan, J., Donaldson, M.S.: To err is human: building a better health system, pp. 1–2. National Academic Press, Washington (2000)
11. Leape, L.L.: Institute of Medicine Medical Error Figures Are Not Exaggerated. JAMA 284, 95–97 (2000)
12. Hayward, R.A., Hofer, T.P.: Estimating hospital deaths due to medical errors: Preventability is in the eye of the reviewer. JAMA 286, 415–420 (2001)
13. Schiff, G.D., Kim, S., Abrams, R., Cosby, K., Lambert, B., Elstein, A.S., Hasler, S., Krosnjar, N., Odwazny, R., Wisniewski, M.A., McNutt, R.A.: Diagnosing diagnosis errors: Lessons from a multi-institutional collaborative project. Advances in Patient Safety 2, 255–278 (2005)
14. Croskerry, P.: Critical thinking and decision making: Avoiding the perils of thin-slicing. Annals of Emergency Medicine 48(6), 720–722 (2006)
15. Bavdekar, S.B., Pawar, M.: Evaluation of an internet-delivered pediatric diagnosis support system (ISABEL®) in a tertiary care center in India. Indian Pediatrics 42, 1086–1091 (2005)
16. Larissa, A., Lyman, J., Borowitz, S.: Impact of a web-based diagnosis reminder system on errors of diagnosis. In: Poster Session Presented at American Medical Informatics Association Annual Conference (2006)
17. Maffei, F., Nazarian, E., Ramnarayan, P., Thomas, N., Rubenstein, J.: Use of a web-based tool to enhance medical student learning in the pediatric intensive care unit and inpatient wards. In: Proc. 15th. Ann. Ped. Criti. Care Coll., Interactions in Pediatric Critical Care, September 30-October 2, vol. 6(1) (2004)
18. Peterson, C.R., Beach, L.R.: Man as an intuitive statistician. Psychological Bulletin 68(1), 29–46 (1967)
19. Edwards, W.: Conservatism in human information processing. In: Kahneman, D., Slovic, P., Tversky, A. (eds.) Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge (2002)
20. Kahneman, D., Tversky, A.: The simulation heuristic. In: Kahneman, D., Slovic, P., Tversky, A. (eds.) Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Boston (2002)
21. Koehler, J.J.: The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. Behavioral and Brain Sciences 19(1), 1–53 (1996)
22. Fischhoff, B., Beyth-Marom, R.: Hypothesis evaluation from a Bayesian analysis. Psychological Review 90(3), 239–260 (1983)
23. Beyth-Marom, R., Fischhoff, B.: Diagnosticity and pseudodiagnosticity. Journal of Personality and Social Psychology 45, 1185–1195 (1983)
24. Sonnenberg, A.: We only see what we already know – a modified Bayes' formula to explain inherent limitations of diagnostic tests. Medical Hypotheses 63, 759–763 (2004)
25. Alvarez, S.M., Poelstra, B.S., Burd, R.S.: Evaluation of a Bayesian decision network for diagnosing pyloric stenosis. Journal of Pediatric Surgery 41, 155–161 (2006)
26. Graber, M.L.: Diagnostic errors in medicine: What do doctors and umpires have in common? Morbidity & Mortality 2, 1–6 (2007)
27. Lindgaard, G., Pyper, C., Frize, M., Walker, R.: Does Bayes have it? Decision support systems in diagnostic medicine. International Journal of Industrial Ergonomics 39(3), 524–532 (2008)

A Composite Task Meta-model as a Reference Model

Steve Goschnick[1], Liz Sonenberg[1], and Sandrine Balbo[2]

[1] University of Melbourne,VIC 3010, Australia
[2] Deakin University,VIC 3217, Australia
{stevenbg,l.sonenberg}@unimelb.edu.au, sandrine@acm.org

Abstract. In this paper we develop a comprehensive composite meta-model from Task Analysis models called the *Reference Task Meta-model* (*ReTaMeta model*) for the purpose of comparing numerous Agent-Oriented meta-models. The reference model needed to be derived from a field independent of the Agent-oriented paradigm, yet based on Psychology. To arrive at the ReTaMeta model we first extracted the meta-models from several well-known cognitive task models including GOMS, GOMSL, TKS, GTA and also the CTT and Diane+H Task Modeling notations for fine grain task detail, and then combined their respective concepts in a complementary and comprehensive way.

Keywords: Task Model, Meta-model, Task Analysis, Interaction model, reference meta-model.

1 Introduction

Milton and Kazmierczak [1] created a method called *The Method of Conceptual Evaluation and Comparison*, for comparing meta-models in one discipline, against a reference meta-model drawn from another discipline. Using it they successfully compared a number of data modeling languages against Chisholm's Ontology from Philosophy [1]. In this paper, we do not detail their method, but highlight its foundational requirements for *a reference model* that comes from an independent field of study while also having related concepts to the area under investigation.

Our larger goal was to evaluate competing meta-models in the Agent-Oriented (AO) paradigm, so, in order to use Milton and Kazmierczaks method, we needed such a *reference model*. This paper outlines the path we took in creating the *Reference Task Meta-model* (*ReTaMeta Model*), a comprehensive composite model, drawn from many existing meta-models from the Task Analysis (TA) paradigm.

The TA paradigm is an ideal source for a reference model independent of AO, since both paradigms cover many similar concepts, despite their different histories. Both cover mentalistic notions assumed to be represented in some way in the cognitive functioning of the human mind, each drawing upon Psychology.

We studied a number of cognitive task models and noted a progression of sophistication over time, as we moved from HTA [2] to GOMS [3], TKS [4], GTA/TWO [5, 6] and GOMSL [7], towards a cognitive model about 'how the mind' goes from goals to finished tasks. We also needed concepts that covered *temporal and relational sequencing of tasks*, to achieve the level of detail needed in our reference model, so we

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 26–38, 2010.

abstracted a meta-model from the Diane+H notation [16,18]. We then merged concepts from all meta-models examined, into a comprehensive composite meta-model from the TA paradigm. **Note 1**: The models chosen were not done so for a comparison across TA models, nor to generate some TA model with new features. They were selected as a representative subset of models, to cover the progression of TA over time, to arrive at a single comprehensive reference meta-model, for the purpose outlined above. **Note 2**: UML is designed to allow the modeler to represent models at various stages of expressiveness, as needed. It has been used as such with meta-models here: sometimes presentation of attributes and methods detract from the discussion, and explanation of them can unnecessarily expand a paper for little value. **Note 3**: As with application modeling, when different Analysts abstract a model from written description, their models are invariably different. Other researchers have abstracted some of the models we investigated – we believe our work complements theirs by adding to a larger body of meta-model research.

2 Extraction of Task Meta-models

This section describes the meta-models of the various TA models considered: GOMS, GOMSL, TKS, GTA/TWO and then the Diane+H notation (with reference to CTT).

GOMS and GOMSL

The GOMS (goals, operators, methods and selection rules) model by Card et al [3] was an attempt to represent human behaviour with regard to interfaces, with a high-level model that software engineers could apply without empirical testing. Beyond earlier approaches, GOMS models user *intention* and is mainly used to *predict* the time taken for an expert to complete given tasks, as a metric.

The entities and concepts in the left-side of figure 1 represent GOMS as follows:

Goal represents a state that needs to be accomplished.

Methods are required to achieve specified *Goals*, they consist of a series of steps, each are either *Operators* to perform, or s*ub-Goals* to achieve. Methods are like plans that the person has already learned.

Operator: *Operators* are basic or simple *actions* that a person can perform on the system or device in question. Card et al define them as elementary acts that are either: perceptual, motor or cognitive in nature.

Selection Rule, Condition: When alternative *Methods* exist to achieve a *Goal*, a *Selection Rule* uses *Conditions* to decide which to use. We included *Enactment* to allow a given *Method* to be used by multiple *Goals* – but not in parallel (a limitation of GOMS).

While GOMS was invented to evaluate interfaces by predicting task performance, it has been used to understand user activity, by choosing *Operators* at the right level of granularity. [7] portrays such a GOMS model as a description of the *procedural knowledge* required to do tasks on a device or system – a representation of '*how to do it*'. *Operators* are usually defined by the hardware (e.g. mouse, keyboard, button) with the analysis concentrated on defining the Goals and Methods.

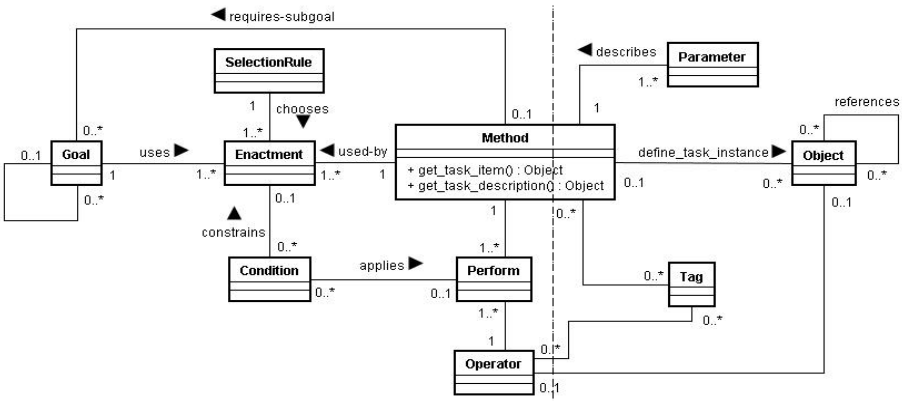


Fig. 1. On the left is the GOMS meta-model, while the superset meta-model is for GOMSL

GOMS is limited to the representation of pre-learned, error-free tasks - unsuitable for analysing *problem solving* tasks. An extension to GOMS named NGOMSL [9,10] - short for *Natural GOMS Language* – is a structured language used to capture *production-rules* that model the aspects of GOMS in analysis. The authors argue that the rendition of GOMS as production-rules, approximates the way people structure their task knowledge. When new UI features are added to an existing technology, the NGOMSL analyst is able to make estimations of the amount of learning required, based on the new production rules needed. A more recent executable version of NGOMSL called GOMSL – for *GOMS Language* [7,9] - is available in GLEAN, a programming environment. GOMSL retains all of the underlying cognitive aspects and direction of the original GOMS, but adds some additional functionality.

The language is partly declarative, with a few procedural branching constructs (*If Then* and *GoTo*), and a *Decide* construct - which is operationally equivalent to the *Switch* statement in C/Java. *Method* declarations have been extended to take a list of parameters, allowing the analyst to generalise methods – providing the ability to abstract procedural knowledge. The GOMSL meta-model is represented in figure 1. The entities *Goal, Method, Operator, SelectionRule, Condition, Perform* and *Enactment* – have the same meaning as they do in the earlier GOMS meta-model, while *Object, Tag* and *Parameter* are innovations introduced in GOMSL:

Object, Tag is basically a symbolic *name* with a list of *property/value* pairs. GOMSL can store them as lists of like-objects – called the Object Store (represented by Object in the meta-model). In cognitive terms, the objects in the store represent '*Long Term Memory*' (LTM). A value set in the properties of an object can be another object, hence the *self*-relationship in the model. Both Object and Tag are used to hold the *task data*. A *Tag* holds a value stored under a symbolic name like a *variable* in a programming language. All *Tags* in GOMSL are said to remain in '*Working Memory*' in cognitive terms. *Objects* can be retrieved from the LTM object store via a Tag, making their *property/value* pairs available to working memory.

Parameter: Represents each parameter used in a Method. The two functions shown in *Method* in the meta-model (which are not in GOMS), have the following purposes:

get_task_description(): retrieves the list of parameters of a Method;
get_task_item(): is then used to retrieved an instance of the parameter property/value pairs for all parameters, from the Object Store. This retrieved set of values are then be used in the Method instantiation.

Note: Using an Operator called *thinking_up* GOMSL can retrieve a *task_item* from the object store based on *search* using a single parameter value (i.e. like retrieving a record from an SQL database table of parameter values, based on the value of one attribute only). It simulates a person '*thinking up'* tasks they may need to achieve a Goal, recalling them from memory. *thinking_up* is an example of a *mental operator*, hypothetical or unobservable in the user, usually inferred by the analyst.

GOMSL retains the GOMS heritage through making Goals, Operators and Methods declarative. He has used an analogy to LTM (long term memory) as an Object Store, and Working Memory as an analogy for the *local memory* in a given simulation. Bringing objects '*into focus'* is his version of dynamic instantiation. Existing objects automatically going *out of focus* when a like-object is brought into focus, is a rudimentary garbage collection mechanism.

Although GOMSL is still focused upon the evaluation of user interfaces in a GOMS analysis, it has the ability to put timing measures (e.g. 50 ms) against various *methods* and *operators*, to simulate an experienced user. Executing code written in GOMSL leads to a GOMS model. The high level Methods with their set of parameters, are extracted from the Object Store and translated into the lower-level Methods needed to achieve the corresponding Goals, repeatedly, until they are all reduced to primitive Operators, each of which takes an assumed time.

The Task Knowledge Structure (TKS)

The cognitive ability humans bring to the interface predates the machines we apply them to, so the originators of some task models tried to make theirs more general, with application beyond the computer-human interface. The Task Knowledge Structure (TKS) model [4, 12] is and example. It takes GOMS as a starting point, then adds knowledge structures held in the mind, specifically, knowledge related to a *task*. They assert that people gather knowledge structures about tasks they have learned and performed, for application in future similar tasks. The TKS model represents this *task knowledge* held in memory then activated during task execution.

They thought that the four dimensions of GOMS with some extension, '*might be considered the basic structural components of any task knowledge that a person might recruit*' [4]. Not just in doing tasks, but structures outlining how particular knowledge is brought to the task. TKS represents *a summary* of the sorts of knowledge a person stores about doing a task. Figure 2 (left-side) is our meta-model of TKS, abstracted from their key publications, as follows:

Goal: Goals are represented in a goal substructure, which '*includes enabling and conditional states, if a goal is to prevail*', i.e. *Pre-conditions* and *post-conditions*. Goals and sub-goals represent states, and are not executable.

Plan: A method to achieve a goal in terms of state change, using tasks and sub-goals.

Task: A task is an activity to achieve a change of state. Task *procedures* (represented via the task *self*-relationship, and associated sub-goals) are one of the knowledge types, and are executable.

Object: They assume that knowledge is stored about both *physical* and *informational* objects. Johnson later added *conceptual* [12]. Objects are configured into taxonomies.

Agent: While they did not initially use the term 'agent', instead using 'person' and 'individual', P. Johnson later defined *agent* as either: person, animal or machine [12].

Role, Responsibility: They saw roles as '*heavily implicated'* in the way that people bring particular sources of knowledge to a task. A TKS role is defined by a particular *set of tasks* for which an individual is *responsible* – a many-to-many relationship here.

Similarity is a measure of the similarity between Tasks performed for different Roles. E.g. whether a person is organising a meeting at work or at home, will involve a different set of tasks, however, the skills in one, help in doing the other.

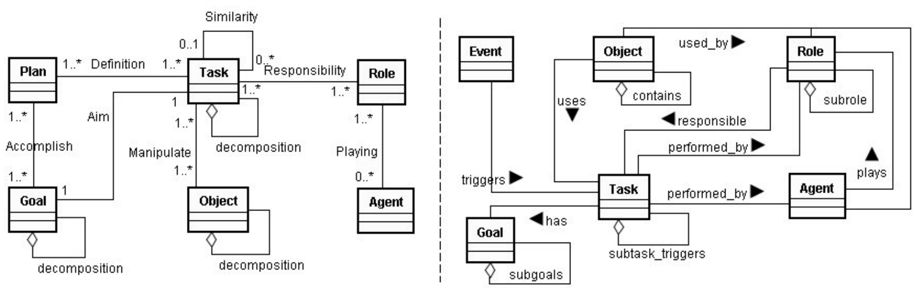


Fig. 2. Meta-model of TKS (left) abstracted from [4,11]; Meta-model of TWO (right)

The originators of TKS do have several other concepts that we have not singled out here from the task hierarchy in the meta-model in figure 2 above, namely:

Action, Procedure and Strategy: *Actions* are simply the terminal or leaf nodes of the *task hierarchy*, so they can be thought of as constituent parts of the task hierarchy. A TKS *Plan* varies within the paper [4]: at one point plans consist of '*goal substructures'*, at other times they consist of '*task substructures'*. In addition to that they have *procedures* described as a '*macro-action',* which a Task may call upon to achieve a goal. Elsewhere they refer to a procedure as '*an element of a subtask'*, elsewhere, as a '*task procedure'*. Their concept of a *Strategy* is simply one that separates procedures for achieving the same sub-goal. I.e. A logical OR in the task hierarchy caters for such alternatives. P. Johnson later removed Strategy from the TKS model himself [12]. In abstracting the TKS meta-model here, we have taken the more recent view of *a task hierarchy*, as defined in various *task notations*. These task model notations have *temporal* and *logical operators*, and can represent *selection* and *concurrency*. (E.g. Diane+H and CTT) As such, the Task *hierarchy* in the TKS meta-model in figure 2, is more than expressive enough to encompass *procedures*.

In several ways, the TKS model seems quite an advanced TM for its time, however, the variations in terminology used across the three defining publications, particularly between the terms: *task, procedure, macro action, task procedure*, and *plan* – impedes the abstraction of a meta-model from those descriptions. In comparison, the definitive descriptions of HTA, GOMS and GTA/TWO are very concise in the meaning of their inclusive concepts and terms.

We found TKS most interesting for its introduction of *Object, Role, Responsibility* and *Agent*, beyond the other entities it has in common with HTA and GOMS. The TKS authors cite research that backs up their use of Objects within knowledge structures held in the mind. P. Johnson later played down the *Role* and *Responsibility* concepts in TKS in his book [12].

Groupware Task Analysis (GTA), Task World Ontology (TWO)

Groupware Task Analysis is a generic task analysis model for which the creators present an ontology in later literature, called the *Task World Ontology* (TWO) [5, 6]. TWO is effectively the meta-model of GTA and it includes as its primary entities: *Event, Task, Goal, Agent, Role* and *Object*. GTA was an early task method that allowed for the analysis of people using *collaborative* technology, supported with tools, in the field of computer supported cooperative work (CSCW). In a comparison of TA models which included GTA/TWO [13], those authors characterise the changes in task models in order to deal with the cooperation, as: '*A role is defined by the task the role is responsible for. Roles are assigned according to organisational rules*'. Its support for *Agents, Roles* and *Responsibilities*, is not more than what the TKS model added to task modeling, unlike TKS, it is very clearly defined. The researchers around the TWO model extended support beyond conceptual modeling into tools for researchers and practitioners. As such, GTA/TWO gained a reputation as a TA and modeling framework used to support CSCW application development.

The TWO meta-model is rendered here in UML class notation in figure 2 (right-side). It includes an extra entity beyond the Task Models considered so far: it has ***Event*** as an entity within the model itself, which *triggers* Tasks to swing into action.

The concept of a *trigger* here leads to mandatory actions being fired via a Task hierarchy, without any deliberation. The TWO meta-model does have a Goal entity separate from the Event interaction, such that Goals can be related to Tasks and subtasks. However, TWO Goals are sub-ordinate to Tasks, in turn triggered by Events, rather than Goals being unrelated to external events, so the concept of goal-driven behaviour (as seen in software agent meta-models) is not represented within the conceptual framework built around TWO - but it goes conceptually very close.

GTA/TWO is coming from a more practical modelling and tools perspective. While it does have an entity for Object, TKS included the Object entity from a psychological basis, drawing from both theory and empirical studies. TWO includes the *Role* and *Agent* entities along with the *responsible* relationship, coming from the practical need to support the multiplicity of users involved in a CSCW application.

The Meta-model Behind Diane+H Notation, with a comparison to CTT

Diane+ is a methodology with a task model *notation* [8,17]. Diane+H is a sub-set of Diane+ covering the formal notation, as implemented in software tools such as *Tamot* [16,18]. Diane+H has a significant amount of expressive power in portraying task compositions from high-level user goals to low-level actions. It allows for temporal and logical relationships between tasks. It differentiates tasks by *type* of either: interaction by the *user*; a task performed solely by the *application/system*; and *manual tasks* by the user not involving the system at all. Additionally, it can represent tasks that run in parallel, and tasks that need to follow on from one another in sequence.

Preconditions may *trigger* a task and *link-conditions* may be placed between sequential tasks. These and other capabilities make the Diane+H meta-model a complex one. Figure 3 portrays a meta-model of it we abstracted from [16,18].

Diane+H is included in our selected examination of meta-models as a representative of the graphical notations that have been developed in the TA discipline, which include CTT [19] and MAD. In a cross-comparison of six notations [14], Balbo et al rated Diane+ most highly overall, across 10 dimensions, including aspects of the SDLC. That said, we still examined CTT and the related tool CTTe as it appears more flexible than Diane+H and we comment on some differences, below.

The range of applications it has been applied to demonstrates the flexibility of Diane+H. In addition to the two early stages of the SDLC (Requirements/Analysis; Design) that Diane+ was originally created, it has been used [14] to: automatically generate online help and user documentation; identify the place of new technologies in the work environment; in the later phases in the SDLC (Development and deployment); and more recently in displaying Task Models for web navigation paths and options, automatically generated from web sites [15].

A detailed description of the significant *entities*, *relationships* and most of the *attributes* needed to represent a Diane+H modeling tool such as Tamot, follow:

Task may be performed either *manually* by the user, by the system (*automatically*), or via the user *interacting* with the system. They can be *optional* or *mandatory*, *elementary* or *composites*. Composites may appear visually expanded in the tool, or not – hence the boolean attribute *expanded*. Tasks may be performed repeatedly with a minimum and maximum number of *iterations* declared in the notation. A Task may be *terminating*, identified here with the boolean attribute named *terminal*.

LogicalOperator: Sub-tasks of a Task may be linked via a logical operator - OR, AND or XOR. Two *or more* sub-tasks may be grouped together with an AND. Interestingly, a Task ***hierarchy*** may start with a *LogicalOperator* at the top of the tree, so the cardinality on relationship *leads_to* is *zero-or-one* at both ends.

RelatedTask: Diane+H notation is most often used to represent ***hierarchies*** *of tasks*, but in some application models, a Task may be linked to more than one 'parent' Task, representing a ***network*** *graph* rather than a tree graph. The *RelatedTask* entity here, ***allows for both sorts of graphs***. In addition, it caters for the ability of Diane+H to represent Tasks that happen in *parallel* (for which the *sibling* attribute is set to *true*, or *false* – for parent-child relationships that happen *sequentially*). The *elder* relationship links it to a *parent Task* in the case of parent-child relationships, and it points to the *next oldest sibling* in the case of *parallel Tasks.* Note: In the notation, parallel Tasks are visually stacked in a vertical column without lines between them [18] – the business-rule used here to present such a stack, is to place the eldest sibling Task topmost in the stack, and so on down to the youngest sibling.

LinkCondition: Tasks that are linked together sequentially may encounter a condition that needs to be met before task execution continues. This *LinkCondition* is expressed in a string attribute called *expression* – which is set to *null* when defining a sequence with *no* conditions. In [19], Paterno portrays several task notations which have specific icons in CTTe which don't have equivalents in Diane+H. However, we find that '*LinkCondition*' could be broadened in meaning to enact several of them including: *concurrent communicating tasks*; *task independence*; and *suspend-resume*.

Precondition: In addition to LinkConditions, *Preconditions* may trigger a Task in the first place. The Precondition may *involve*: the progress of *other Tasks* elsewhere in the model; some *Event*; or else a change in a *Data* value.

Feedback: Diane+H has two attributes in a *Feedback* entity, one that describes the *purpose* of the Task, while the other represents a message to a user when the Task is complete. The description held in the *task_purpose* attribute is useful to the designer of an application when prototyping it via a tool such as Tamot, as it gives some context-sensitive information.

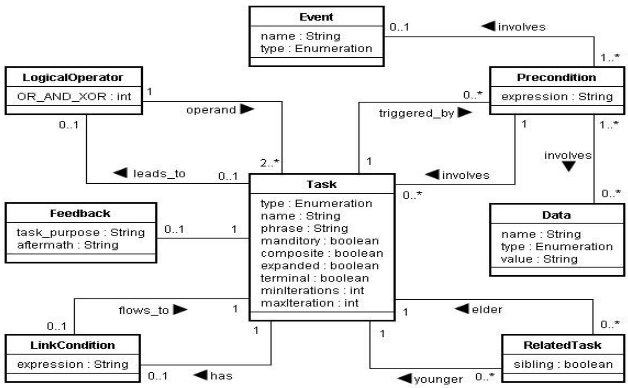


Fig. 3. Meta-model of the Diane+H Task Model Notation, abstracted from [16,18]

CTT has a good treatment of *Objects* which is absent in Diane+H (except via *Data*). CTT has a relationship between User and Objects called *Rights*, to cater for multiple users with differ access to Objects. CTT's *Task* also has an attribute called *Platform*, used to allow or disallow objects available to a Task, based on Platform.

Note: Several researchers have embedded Task Models in tools which facilitate much of the SDLC, notably: Tamot using Diane+H as their notation [16, 18]; and CTTe by Paterno [19]. Paris et al were able to automate the extraction of task models from standard text using WordNet, and from UML class and sequence diagrams. Inversely, they are also able to generate UML diagrams from Task models, visually represented in Diane+H and declaratively in an XML format. These extracted UML models are limited by the '*user-oriented nature of the task models*' they built [16].

3 A Composite Task Meta-model

In the introduction we touched upon an evaluation technique used to compare meta-models called *The Method of Conceptual Evaluation and Comparison*. It requires a *reference meta-model* from a field independent of the one being studied (the Agent-oriented paradigm, in our case). In this section we put forward our composite meta-model drawn from the TA paradigm as just such a reference model, one that we think is a sensible composition of concepts from the TA meta-models presented above.

There are several ways that such a reference model could be constructed from separate TA meta-models, all of them with some subjective reasoning, which is worth a brief discussion: Firstly, even if the models had clearly overlapping concepts, we could have taken the *intersection* of concepts between the models; however that would leave us with a minimalist model, not very helpful with the comparison of agent meta-models which often have an extensive range of concepts. E.g. some agent-meta-models have upwards of 30 entities and as many relationships. It was instructive to us, that in the comparison of *data modeling languages* done by Milton and Karzmierczak [1], they deemed a *subset* of the Chisholm ontology as a relevant reference model for their purposes (note: the Chisholm ontology from Philosophy is considered a '*commonsense realism*' model of the world around us, that includes *states* and *events*), since data modeling languages are not intrinsically interested in *states* and *events* - other *process-oriented* languages are. What we needed was a *superset*, a comprehensive model representing a *union* of the constituent concepts in the representative TA meta-models we examined, to reach a comparatively complex meta-model to those in AO. When there were *similar* concepts in two different meta-models that did not perfectly line up, we took the more flexible one.

There was also some chronology to the TA models. The cognitive task models we examined were largely built upon the assumed knowledge structures in peoples' minds and together they represent an accumulating set of concepts, over the time period in which they were invented. From them we have *Tasks*, *Goals*, *Plans* (*Methods + Operators/Actions + Rules/Conditions*), *Objects (with ontological structure), Roles, responsibilities* and *Events* – with adequate connection to psychological theory and/or the backing of empirical studies, to place them somewhere within the cognitive architecture inside peoples' heads. To those, we further considered the concepts in the Diane+H and CTT notations, to adequately allow for the decomposition of tasks with sophisticated expressive power in order to cater for the temporal and logical relationships between tasks, and other features.

Figure 4 represents our *composite task meta-model*, which carefully mixes and matches concepts from the Task meta-models examined above, into a model that includes all key concepts in one meta-model. We claim that it is a *broad reference model*, but we certainly did not examine all task meta-models in our process, and so we do *not* claim that it is a *unified task model for all TA purposes*.

The definition of the concepts and terms as they are represented in figure 4 follow, with an explanation of the choices we made when alternatives were available:

Goal, Task: In a summary of his ConCurTaskTrees notation (CTT), Paterno defines a *goal* in a task model as follows: '*A goal is either a desired modification of a state or an enquiry to obtain information on the current state. Each task can be associated with a goal, which is the state change caused by its performance.*' [19]. Goals are represented in a substructure, which '*includes enabling and conditional states, if a goal is to prevail*'. That is a good general definition of a Goal, across the TA models examined. A TKS *Task* is an activity that results in a change of state in a given domain – and that aligns with a GOMS *method*. As in TKS and in Paterno's definition, there is a one-to-one mapping between Goal and Task, a *Task* sets about *satisfying* a *Goal* as portrayed in figure 4 - top-left. They could both be represented as predicates/method-signatures (as is the case in both GOMSL and GTA/TWO), but the Goals in the goal hierarchy will just list the parameter names, whereas the specific

Task/method-signature, will have many if not all of the parameters set to *values*. I.e. The *Task* that comes via the *responsibility* relationship from *Role*, may calls upon *Goal*, for the list of sub-Goals that need to be answered/achieved.

RelatedTask: From the Diane+H meta-model, this is the most flexible approach to task decomposition of all the meta-models examined. As discussed above, it allows for serial and parallel execution of tasks, and for *networked* graphs (multiple parent tasks) in addition to *hierarchies*. As with the GOMS *Method*, the Task in this model consists of related sub-tasks and *Operators*. Operators are represented via the Perform entity (as the same Operator, may be used in many Tasks, e.g. a left *mouse-click*). The Goal associated with any given sub-task can be located via the *satisfying* relationship.

Plan: A Plan is a way to achieve a goal in terms of state changes. *Plan* here takes the GOMS meta-model entity *Enactment* as the better more detailed approach, with *SelectionRules* and *Conditions* that may be placed upon the elements of the Plan, e.g. upon the sub-tasks which the sub-goals. *SelectionRules* allows for the OR, AND and XOR logical operators that are in Diane+H, in addition to other possible rule constructs which may include *Conditions*. The Condition entity can also be used to represent the *LinkCondition* entity as seen in the Diane+H model, which represents a *PreCondition* for any sub-tasks a given Task model.

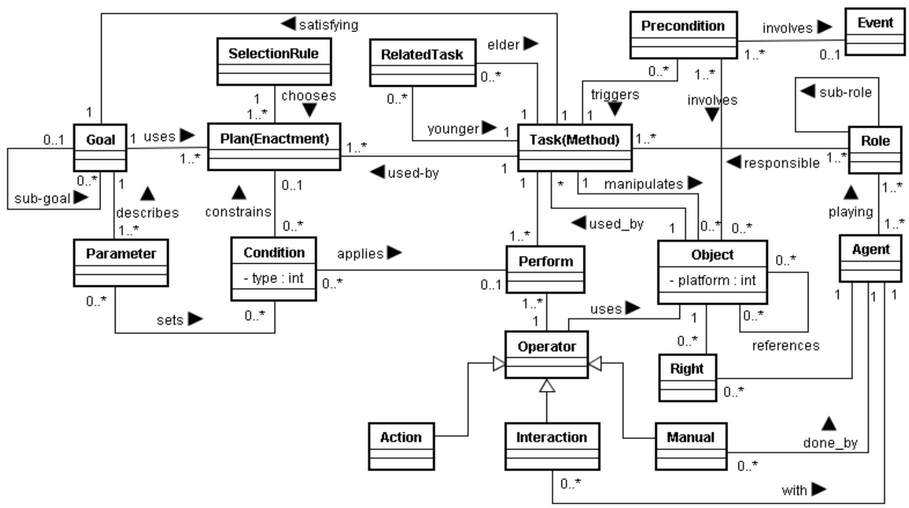


Fig. 4. The Composite Reference Task Meta-model (named the *ReTaMeta Model*)

Event, Precondition, Object, Right: The *Precondition* for the topmost Task to be set in motion, can be either an *Event* (as in GTA) or some change in state represented in the *Object* ontology, or both. What was the *Data* entity in Diane+H is represented here as *Object*, but not just involved in the preconditions of Task, since the other meta-models examined (i.e. TKS, GOMSL, GTA and CTT) use the concept Object to represent a structure of related domain objects, and so it is the case here too in the composite task meta-model. We also include *Right* here from CTT's *Rights*.

Operator, Action, Interaction, Manual, Agent: Diane+H and CTT can be used to model tasks that are performed by either: the user (*Agent*), the system, or both (via interaction between the two). While this variation in Task was allowed for via a *type* attribute in the Diane+H meta-model (and the *category* attribute in CTT), in our Composite meta-model it is refined by sub-classing the *Operator* entity three ways: *Action* (e.g. an internal basic system action, requiring no further fine-grained representation in the task Model); *Interaction,* from which a resulting action is returned via an interaction process not worth modeling in a specific task model – e.g. a standard GUI component such as a *FileChooser* dialog widget, used to locate a filename in a standard way; *Manual* (some basic manual task the user must do themselves, outside of the system being modeled – e.g. *stand-up*).

Agent, Role, responsible: The human user is represented in the *Agent* entity, which can represent either: person; system/machine; or animal (as in TKS). Both TKS and GTA/TWO have *Role* as an entity that an Agent can play, such that an Agent can play one-or-more Roles and a Role can be played by one-or-more Agents – the Role is *responsible* for a set of tasks (one-or-more). While TKS has Role as a simple entity, GTA/TWO has it as a Role *hierarchy*, so in our composite meta-model, we include Role as a possible hierarchy of roles, with each *responsible* for a set of *Tasks*.

Parameter is included here from the GOMSL meta-model. The use of *Parameters* in GOMSL is dynamic, to retrieve all sorts of *method signatures*, in a '*thinking'* like manner without a specific plan. In GOMSL, *Parameter* is attached to Method, but here we attach it to Goal. The reason is that a Goal is declarative, just as a *Method Signature* is declarative. So it makes good sense to attach *Parameters* to *Goal.* Any *Task* that needs such methods/*sub-goals*, can get to the definition via the *satisfying* relationship between *Task* and *Goal*. Also, once a Goal has been reached, the associated *Parameter/s* may be used to *Set* a *Condition* upon other tasks.

4 Usage and Potential of the ReTaMeta Model

Our composite meta-model was drawn from the TA paradigm to be completely independent of Agent meta-models. We have used it successfully to do a *Conceptual Evaluation and Comparison* of many AO meta-models, to be reported elsewhere. While that was the initial reason it was devised, the *ReTaMeta Model* has other possibilities, including in the service of designing interactive systems. A computational system based on the ReTaMeta Model would inherit several interesting properties from the meta-model:

- The Interaction sub-class of Operator allows for any number of pre-existing UI components/widgets to handle standard user-system interaction.
- The three forms of *Operator* - *Action* (the System), *Interaction* (User and the System), and *Manual* (just the human User) - means that it would be highly suitable for *mixed-initiative* systems, including *human-in-the-loop* agent systems.
- The *Action* entity could be mapped to Internet services (e.g. where an *operator* is outsourced to an Internet/web service).
- The *Goal* hierarchy lends itself to a *declarative* language approach (which in turn, allows for proactive behavior by the system), while the *SelectRule, Parameter* and *Condition* entities allow for *procedural* language constructs, in the one system: a more flexible yet concise approach to solving goals, than a purely declarative one.

- *Precondition* and *Event* facilitate reactive behavior - suitable for 24/7 systems.
- The *RelatedTask* entity allows for both *network* and *tree* graphs of task structures, and the *Object* hierarchy allows for ontologies of objects to be involved.

It would have some features comparable to agent systems (e.g. proactive goal-oriented behavior), but still clearly focused on the human-user from several directions, including: the origin of the meta-model itself (from the TA paradigm); the flexibility of the *Operator* entity, and via the *Role hierarchy* with related *responsibilities* in the form of Tasks to be undertaken and achieved. In short, it could be pressed into service in designing new interactive systems in the direction of People-Oriented Programming [21] – and that is where we intend to concentrate some future research effort.

References

1. Milton, S., Kazmierczak, E.: A Study Using a Common-sense Realistic Ontology. Journal of Database Management 15(2), 19–38 (2004)
2. Annet, J.: Hierarchical Task Analysis. In: [20], pp. 67–82 (2004)
3. Card, S., Morgan, T., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale (1983)
4. Johnson, P., Johnson, H., Waddington, R., Shouls, A.: Task-Related Knowledge Structures: Analysis, Modelling and Application. In: Fourth Conference of the British Computer Society on People and Computers IV, Manchester, UK, pp. 35–62 (1988)
5. van der Veer, G., van Welie, M.: Groupware Task Analysis. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Analysis Design, p. 808. Lawrence Erlbaum Inc., Maywah (2003)
6. van Welie, M., van der Veer, G.: An Ontology for Task World Models. In: DVS-IS 1998. Springer, Adington (1998)
7. Kieras, D.: GOMS Models for Task Analysis. In: [20], pp. 83–116 (2004)
8. Tarby, J.-C., Barthet, M.-C.: The Diane+ Method. In: The Second International Workshop on Computer-Aided Design of User Interfaces. University of Namur, Belgium (1996)
9. Kieras, D.: Towards a practical GOMS model methodology for user interface design. In: Helander, M. (ed.) Handbook of Human-Computer Interaction, pp. 135–158. North-Holland, Amsterdam (1988)
10. Kieras, D., Polson, P.G.: An approach to the formal analysis of user complexity. International Journal of Man-Machine Studies 22, 365–394 (1985)
11. Johnson, P., Johnson, H.: Knowledge Analysis of Task: Task analysis and specification for human-computer systems. In: Downton, A. (ed.) Engineering the Human-Computer Interface, pp. 119–144. McGraw-Hill, London (1989)
12. Johnson, P.: Human Computer Interaction: Psychology, Task Analysis and Software Engineering. McGraw-Hill International Ltd., London (1992)
13. Limbourg, Q., Vanderdonckt, J.: Comparing Task Models for User Interface Design. In: [20], pp. 135–154 (2004)
14. Balbo, S., Ozkan, N., Paris, C.: Choosing the Right Task-Modelling Notation: A Taxonomy. In: [20], pp. 445–465 (2004)
15. Balbo, S., Goschnick, S., Tong, D., Paris, C.: Leading Web Usability Evaluations to WAUTER. In: Proceedings of the 11th AusWeb, Gold Coast, Australia (2005)
16. Lu, S., Paris, C., Vander Linden, K.: Tamot: Towards a flexible Task Modeling Tool. In: Proceedings of Human Factors, Melbourne (November 2002)

17. Paris, C., Tarby, J.-C., Vander Linden, K.: A Flexible Environment for Building Task Models. In: People and Computer XV - Interaction without Frontiers, ICM-HCI-2001, pp. 313–330. Springer, London (2001)
18. CSIRO: Diane+ Formalisms, `http://www.ict.csiro.au/staff/cecile.paris/from-cmis/projects/isolde/tamot/onlinehelp/Diane+H.htm`, (last accessed: February 15, 2010)
19. Paterno, F.: ConcurTaskTrees: An Engineered Notation for Task Models. In: [20], pp. 483–501 (2004)
20. Diaper, D., Stanton, N. (eds.): The Handbook of Task Analysis for Human-Computer Interaction. Lawrence Erlbaum Associates, Mahwah (2004)
21. Goschnick, S.: People-Oriented Programming: from Agent-Oriented Analysis to the Design of Interactive Systems. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part I, HCII 2009. LNCS, vol. 5610, pp. 836–845. Springer, Heidelberg (2009)

Future Dining Table: Dish Recommendation Based on Dining Activity Recognition

Tomoo Inoue

University of Tsukuba, Graduate School of Library, Information and Media Studies,
Kasuga 1-2, Tsukuba 305-8550 Japan
inoue@slis.tsukuba.ac.jp

Abstract. Dining table is one of the environments that remain to be explored by information technology and that could be important for happy and healthy life of everybody. We have been studying the Future Dining Table system, dining environment of the future. The system features the real time recognition of dining activity, which is represented by the history of the hand movement to the dishes on the table. The history of dining actions is used to know the current status of dining. Currently the system recommends additional dishes according to the dining status. It is expected to be useful as a substitute of a food server in the age of labor shortage.

Keywords: dining computing, tabletop interaction, physical computing, activity recognition, food service.

1 Introduction

Information and communication technologies (ICT) have been positively applied in almost all fields. Because it is known that the wide use of ICT in a nation is closely related to the competency of the nation, ICT use in all fields has been promoted [1]. ICT is now in our daily life and is changing it. The major goal of ICT use has been the pursuit of efficiency and competency in business so far. We think it could also contribute to the pursuit of happiness and quality of life. Because eating is clearly indispensable for our life and closely connected to the feeling of happiness and the quality of life, application of ICT to eating activity is significant to our life and is likely to improve eating environment.

In this paper, we propose Future Dining Table as an example of the realization of the idea of eating environment innovation by ICT. The Future Dining Table recommends dishes to a user. The situation of the system use is assumed in a restaurant or typically in a Japanese pub where many dishes are ordered during dining. It recognizes dining action in real time. Dining action is recognized as a hand movement with chopsticks to the dishes on the table. The system stores the history of dining actions and uses it to know current status of dining. It is used to know whether the user is in the beginning of dining or is in the middle or is in the end, for example. According to the dining status, the system recommends the user the dishes that would fit to the current meal in the right timing for additional order.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 39–49, 2010.

Considering the recent and future shortage of labor supply, it will become more difficult to keep good employees who are capable of recommending another dish which matches the dining status of the customers. The system is expected to be useful as a substitute of a food server in such situation.

In this paper, the chapters are composed in the following way. Related research is described in Chapter 2. The system is proposed in Chapter 3. Brief evaluation of the recognition element of the system is described in Chapter 4. Conclusion is given in Chapter 5.

2 Related Research

2.1 Tabletop Interface, Use of Physical Objects and Its History

People live and act in the physical world. They use various physical objects everyday. Tangible User Interface (TUI) uses physical objects as the user interface, and is thus known as more natural and intuitive way of interacting with computers [2]. Methods of information display on a table have been researched [3][4] because a table is often used when working on something. They are called tabletop interfaces. People put physical objects on a table and use them for many tasks. Thus table interface is closely related to TUI.

The proposed system falls on to the tabletop interface in the ubiquitous computing environment as well as a type of TUI. Dining activity includes the use of dishes, physical objects, on a table.

Because one of the features of a computer is that it can store information, command history of a user or various logs have been often used. For example, they are used for the adaptive support of a user or for the marketing of an Internet shop. As such, use of history information has been popular in computer systems. Although TUI and physical computing has been common, use of history in physical computing has not been exploited very much. One of the examples is a collaborative learning support system that multiple users operate multiple cubes on a touch-sensing table. In this system, history of who operates which cube is shown to the users [10]. However, this system only shows who has not touched which cube yet. Also, its usefulness is not very clear.

Dining activity can be viewed as operating history of dishes. Use of the history enables a system to recognize dining status of a user, not just a beginning and ending of dining.

There are common ways of recognizing physical objects by a computer. Use of visual markers such as barcodes is one of the easy and robust ways. For example, the Interactive Textbook which is one of the applications of the EnhancedDesk uses a textbook. Pages of the textbook are recognized by the barcodes that are printed on the corners of the pages [5]. SmartSkin is an example of other ways of recognition. It is a table where networked electrodes are embedded. It recognizes location of a hand or a physical object with a piece of metal on the table by detecting the capacitance change [6]. Although simple visual markers are used in our current system, dining table is one of the good environments to utilize these preceding studies.

2.2 Dining Recognition

"Sixth dish" is a system to help interpersonal communication during dining. It projects pictures on vacant dishes from above so that pictures give clues to start conversation [7]. This research supports meal times in terms of communication but it does not use dining status or a user's behavior.

In welfare and medical care, meal has been recorded for management of health. This record is typically a written memorandum showing the given menu or sometimes the remains after meal. Because making a written meal record is time consuming and is not easy when there are many clients, automatic recording of meal has been researched in welfare information technology.

Recognition of dining behavior in a video has been proposed. It uses the video of a dining room to detect a hand movement and recognizes eating by applying hidden Markov model [8]. It does not recognize what the diners eat and the order of eating things.

An eating monitoring system has been proposed. This system uses dishes with RFID tags and a tray with a RFID reader and a load meter and pressure sensors. It can measure every step of weight reduction of food. What and how much it was eaten can be recognized. From the analysis of eating pattern with the system, it has been suggested that there is a stable eating pattern in the same person and eating patterns are different between different persons [9]. Though the system recognizes detailed reduction of food, it does not concern about the behavior of a user.

We plan to improve dining environment in not only recommending another dish, but other services such as communication support in the future. Because of this, behavior recognition of a user is necessary. In this assumption, we still like to recognize what food and how much a user eat on the way of the dining for providing fine adaptive service.

2.3 Future Role of ICT Use in the Society

Until recently, major part of the ICT research has pursued economical effectiveness and efficiency. There has not been amount of research on ICT use for happiness, good feelings or for the mental quality of life. However, there is a trend to pursue mental value as well as economical value recently. For example, Japan has turned its helm from being "a nation of economic power" to being "a nation with intellectual presence" [11]. Norman proposed importance of esthetic and emotional value of things in contrast with usability of things in his book published in 2004 [12]. The Workshop on Designing Cute Interactive Media that was held in 2008 focused its attention on the effects of cuteness, i.e. mental value of human, of interface agents or other things on user experience [13].

Our research is heading toward increasing mental value with the application of ICT in our daily life, although the proposed system is only a first step toward this goal and thus is not directly related to the quality of life yet.

3 Future Dining Table

Despite the prevalence of ICT in many fields, dining environment is one of the few that has not changed much. One of the few well known examples is RFID tagged dishes for calorie calculation or for checkout in cafeteria. So there is room for innovation.

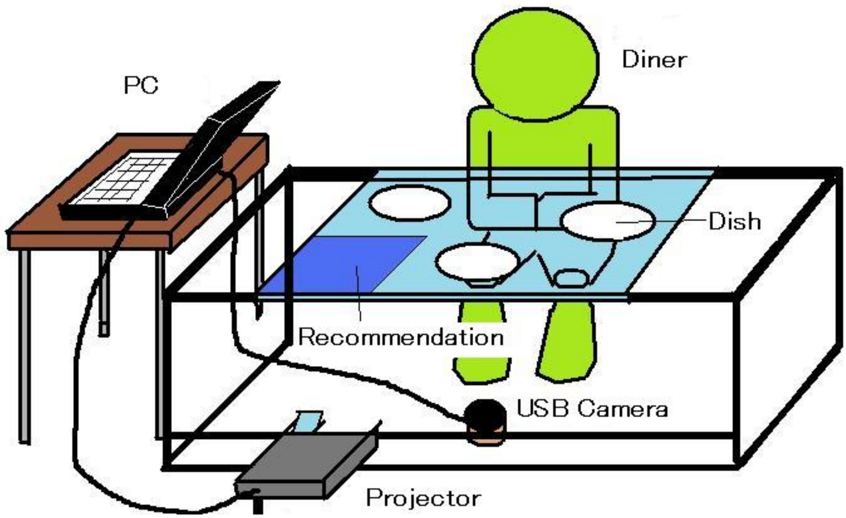


Fig. 1. Future Dining Table

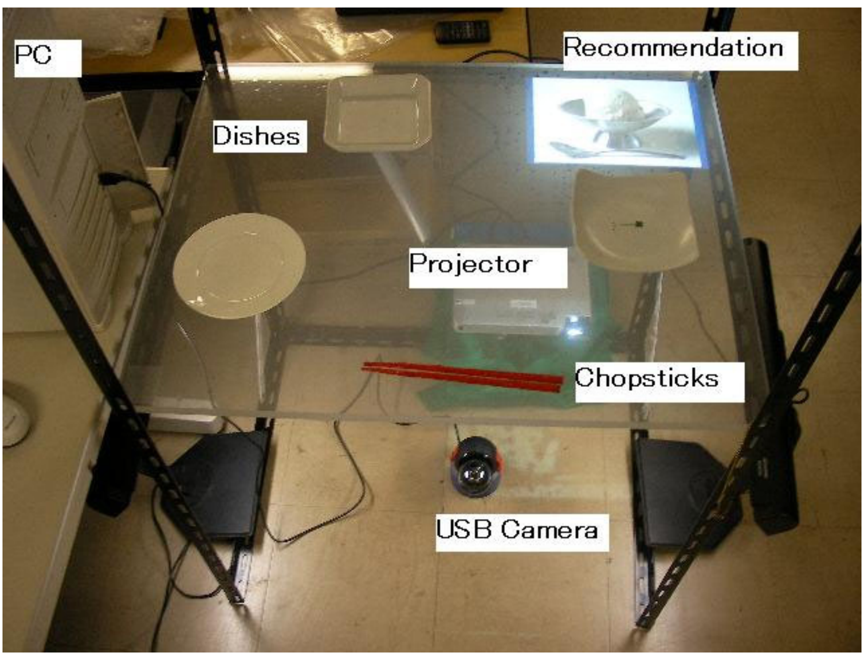


Fig. 2. Future Ding Table overview

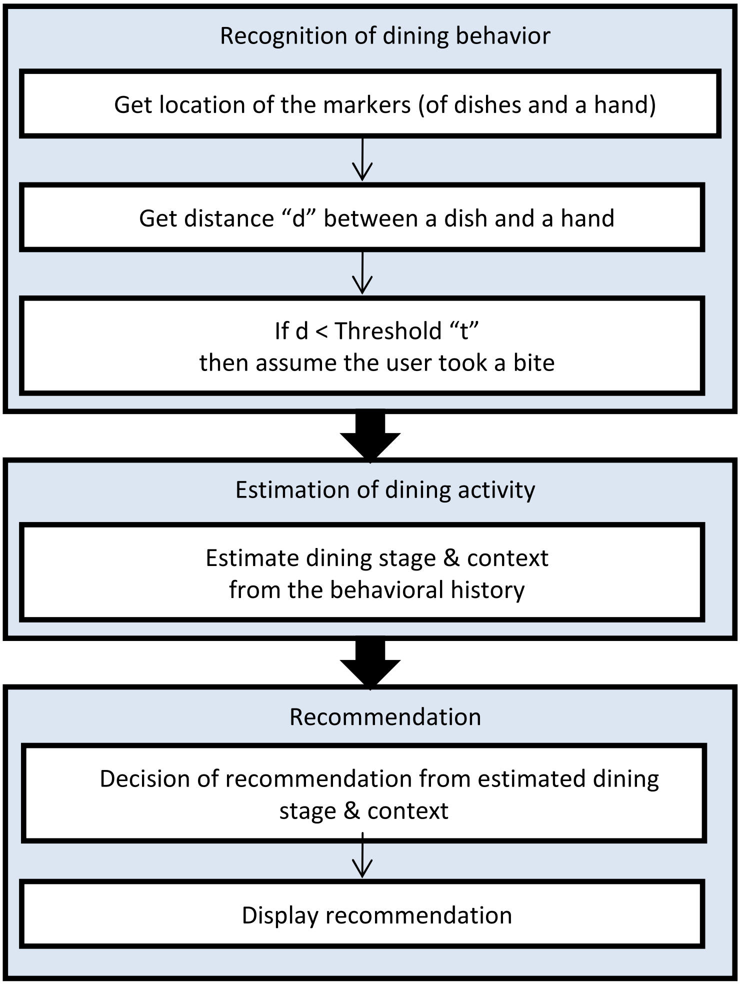
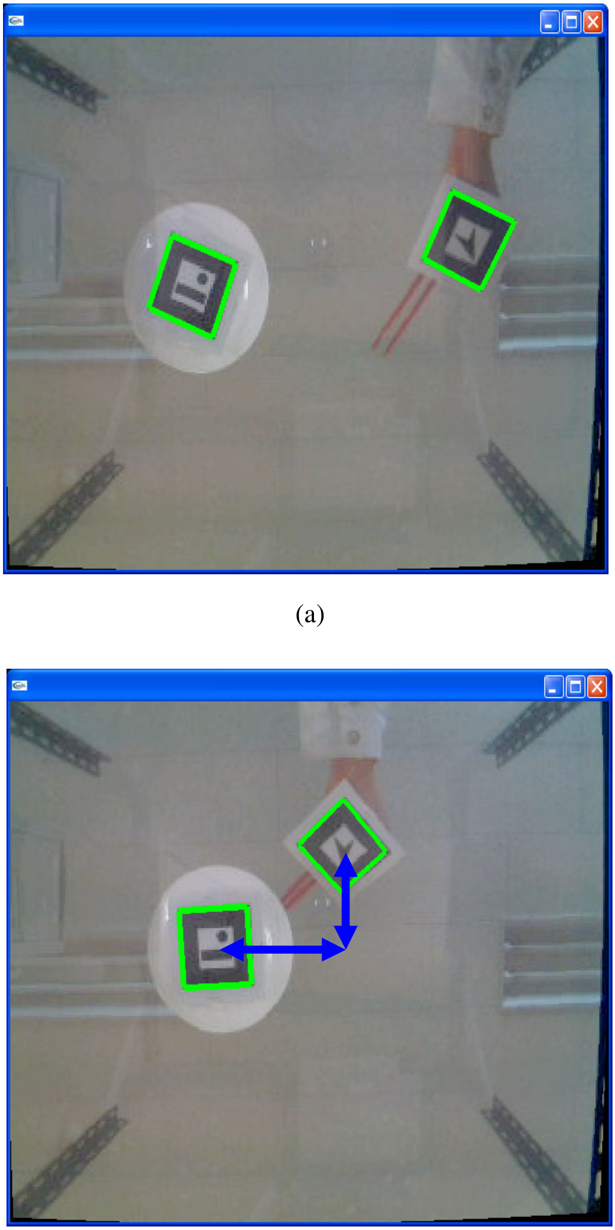


Fig. 3. Future Dining Table algorithm

Dining activity includes handling of dishes and cutleries, i.e. physical objects. Therefore use of software alone is not enough for supporting dining environment. This is different from some other business activities that can be completed by computers and the networks. To recognize the user’s dining status, we need to detect dishes and cutleries, and also need to use their operated history.

We propose the Future Dining Table (FDT) that recognizes dining activity of a user in real time, and recommends another dish for additional order according to the dining status. The system is assumed to be used in a restaurant where additional dishes are usually ordered. It is assumed to replace or to help waiters when human resource becomes more precious.



(a)

(b)

Fig. 4. Dining activity recognition
(a) Recognition of the visual markers
(b) Closeness of a dish and a hand means a bite

The picture of FDT is shown in Figure 1. The FDT system overview from a user is shown in Figure 2. From the dining recognition to the dish recommendation is processed mainly in 3 steps. They are the recognition of dining behavior, the estimation of dining activity, and recommendation (Figure 3). Each process is explained in the following subsections.

3.1 Recognition of Dining Behavior

As in shown in Figure 2, the table is a transparent 15mm thick acrylic board with 60cm depth and 75cm width. The board is sealed by the transparent screen sheet for video projection (Kimoto Co,Ltd. Dilad Screen T40SI-40) so that the recommendation can be displayed on the table. Each physical dish has a visual marker, for which ARToolkit [14] is used, on its bottom.

The user also attaches the visual marker on his/her hand with chopsticks. The markers are recognized from the image taken by a USB camera (Logicool QVR-13R). The video image from the camera is 320x240 (pixel) and 15fps, which covers the table. 1pixel of the image is set to 2mm in the real world in this condition.

Dining behavior is recognized as the recognition of each dish and the distance between a dish and a hand. The user always uses chopsticks when eating, which is normal in Japanese dining. The hand holding the chopsticks moves to a dish closely when the user takes an eating action. The system recognizes this movement as a bite, a unit of eating. Figure 4 shows the recognition process. The visual markers are recognized as in Figure 4 (a). When the marker of a hand (with chopsticks) and the marker of a dish come close enough as in Figure 4 (b), it is recognized as a bite. Currently, it is recognized as a bite when the distance is less than 80 pixels (160mm) in the width direction and is less than 100 pixels (200mm) in the depth direction. These values are determined by experience. If the threshold distance is set closer, judgment becomes more severe. If it is set more distant, judgment becomes more relaxed.

Table 1. Example menus and the number of bites for finishing

Menu	Number of bites for finishing
Fried chicken (Oily)	10
Salad	20
Kimchi one-pot dish (Spicy soup)	30

3.2 Estimation of Dining Activity

What the user eats is known from the dining behavior explained in 3.1. To know how much the user eats, a possible method is to measure the weight of each dish as in the related research [9]. Without such a mechanism, another method may be the use of history of bites. If the number of bites to finish a dish is known, how much remains on the dish can be estimated from the bite counts to the dish. Investigation is needed to know the former factor. It may also depend on the user and other contexts. Here we prepared 3 example menus and assumed the number of bites to finish the dish for the prototype, which is shown in Table 1. By using the history of bites to each dish, more detailed dining activity can be estimated than using the dining behavior only.

3.3 Recommendation

Another dish is recommended using the result of the estimation of dining activity. There are a number of methods for recommendation. Because recommendation algorithm itself is not the focus of this research, we have applied a simple rule-based recommendation for the current FDT. The system displays recommendation according to the operation of physical objects, which has not been exploited yet. Example conditions and the dishes of recommendation are shown in Table 2. With the detailed recognition of dining status and the investigation of actual relation between the condition and the ordered dish, the rule could be refined.

Presentation of recommended dishes is another issue of research. The current system displays the picture of the recommended dish on the table by using a video projector (CASIO XJ-S46).

Table 2. Example rules for recommendation

Condition	Dish of recommendation
All dishes are less than 20% and more than 3 consecutive bites of fried chicken in the recent 15 bites	Sherbet
More than 10 bites of kimchi one-pot dish in the recent 15 bites	Milky ice cream
Less than 25% of Kimchi one-pot dish	Rice for Soupy Rice

4 Initial Evaluation of Dining Activity Recognition

We have developed the FDT prototype as described. We have confirmed the system works by using the example settings. We have also conducted initial evaluation of dining activity recognition of the system.

The subject was 3 right-handed male university students aged 22 to 24. All were the first time users of the system. They were asked to eat normally as in everyday. 3 dishes were set on the table as shown in Figure 5. A snack of small pieces was chosen as the food on the dishes because we could control the number of bites by this. 8 pieces of the snack were set on each dish. This meant a subject bit 24 times. The dining behavior was videotaped. The record by the FDT was compared with the videotaped behavior, which provided the correct answer.

The result is shown in Table 3. Precision, recall and F-measure were used as the measures. These are originally from information retrieval, and have become generally used as the measure of such evaluation. Precision is defined as true positive / true positive + false positive. When different dish from the actual dish is recorded by the system, precision decreases. Recall is defined as true positive / true positive + false negative. When there is no record while there was an actual bite, recall decreases. F-measure is defined as the harmonic mean of precision and recall; F = 2 / (1/Precision + 1/Recall). It is used to represent the performance of both precision and recall in a single measure.

Fig. 5. Future Dining Table in use

Table 3. Result of the dining activity (bite) recognition

Subject	Precision	Recall	F-measure
A	.95	.79	.86
B	.82	.75	.78
C	.68	.54	.60
Average	.82	.69	.75

There was a little dispersion by the subjects. Low recall rate of subject C was because of his unique form of using chopsticks. Usually people pinch food with the chopsticks, but he laid down them and scooped the snacks. The recall was .69 and the precision .82 on the average, resulting .75 of F-measure. It could be acceptable taking into account that the service provided by the system is not mission-critical, although the author does not know of any similar evaluation.

5 Conclusion

Future Dining Table was proposed, designed, and the prototype has been implemented. The FDT recognizes dining activity by a camera and the visual markers, which are fairly simple and inexpensive equipments. It can be improved by the application of image recognition technique because the current system requires the user to put a visual marker. By making use of the history of physical objects, FDT can recommend additional dish to order in the right timing of eating. Considering recent and future shortage of labor power, it will become more difficult to keep good employees who are capable of recommending another dish with caring the dining status of the customers. The system is expected to be useful as a substitute of a food server in the kind of restaurants where additional dishes are often ordered. For the future research, we will improve the prototype and will extend the system to better support dining activity including interpersonal communication.

Acknowledgments. The author would like to thank Yuuki Seto for his contribution to the development of FDT. This research was partially supported by the JSPS Grant-in-Aid for scientific research 22500104, 2010.

References

1. Ministry of Internal Affairs and Communications of Japan, The Panel on ICT Growth, Final x ICT Vision Report (2008) (in Japanese), `http://www.soumu.go.jp/joho_tsusin/eng/Releases/Telecommunications/news080703_3.html`, `http://www.soumu.go.jp/s-news/2008/pdf/080703_6_bt2.pdf`
2. Ishii, H., Ullmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In: Proc. CHI 1997, pp. 234–241 (1997)
3. Wellner, P.: Interacting with paper on the digital desk. Communications of the ACM 36(7), 86–96 (1993)
4. Matsushita, M., Hijikata, Y.: Sugihara, Survey of tabletop systems. Journal of Human Interface Society 9(1), 35–58 (2007)
5. Koike, H., Kobayashi, Y., Sato, Y.: Integration of Paper and Digital Information and Real Time Interaction by hand on Augmented Desk System. Transactions of Information Processing Society of Japan 42(3), 577–585 (2001)
6. Rekimoto, J.: SmartSkin: An Infrastructure for Freehand Manipulations on Interactive Surfaces. In: Proc. CHI 2002, pp. 113–120 (2002)
7. Amano, K., Nishimoto, K.: pHotOluck: Communication Activating System at A Dining Table by Projecting Pictures on Dishes. IPSJ SIG Notes 2004(31), 103–108 (2004)
8. Gao, J., Haupymann, G.A., Bharucha, A., Wactlar, D.H.: Dining Activity Analysis Using a Hidden Markov Model. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR), vol. 2, pp. 915–918 (2004)

9. Kawashima, T., Tanisugi, Y., Mitsudo, Y.: Dining Monitoring System Using Sensing Tray and ID-ware, IEICE technical report. Welfare Information Technology 106(285), 61–66 (2006)
10. Watanabe, K., Takeuchi, T., Inoue, T., Okada, K.: Face-to-face collaboration system that identifies individual user. IEICE Transactions on Information and Systems J91-D(12), 2755–2764 (2008)
11. Ministry of Education, Culture, Sports, Science and Technology of Japan, Council for Science and Technology, Toward a nation with intellectual presence (1999) (in Japanese), http://www.mext.go.jp/b_menu/shingi/12/gakujutu/toushin/990601.htm
12. Norman, D.A.: Emotional Design. Basic Books, New York (2004)
13. Cute Interactive Media – Workshop on Designing Cute Interactive Media (2008), http://kawaii.wikidot.com/
14. ARToolKit Home Page, http://www.hitl.washington.edu/artoolkit/

Developing a User-Centered Mobile Service Interface Based on a Cognitive Model of Attention Allocation

Julia Niemann[1], Volker Presse[2], Jessika Reissland[3], and Anja Naumann[1]

[1] Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
Julia.Niemann@telekom.de, Anja.Naumann@telekom.de
[2] Berlin Institute of Technology, Chair for Technology and Innovation Management, Sekr. H 71, Straße des 17. Juni 135, 10623 Berlin, Germany
Volker.Presse@tim.tu-berlin.de
[3] Department of Human-Machine Systems, TU Berlin, Franklinstr. 28/29, 10587 Berlin, Germany
jre@mms.tu-berlin.de

Abstract. Mobile devices have conquered our daily life. They are carried and used at almost all times and in any situation sometimes even against legal restriction - e.g. in the car. Services running on these devices (e.g. email, text messages, etc.) include graphical (GUI) and voice interface (VUI) causing visual distraction for the driver although they could solely be operated by speech in- and output. As a result services should adapt their interfaces due to the specific usage scenario. The aim of this work is therefore to develop design recommendations based on a cognitive model for the voice user interface taking the particularities of the specific scenario (e.g. reduction of off-road eye gazes) into account. We assume that distraction is mainly due to an increased effort and a decreased expectancy of the VUI compared to the GUI. Design recommendations derived from these considerations will be described by means of a concrete example.

Keywords: Mobile services, in-car infotainment, car services, voice-user-interface.

1 Introduction

Nowadays there is an enormous demand for all kinds of services that can be used on portable as well as mobile devices satisfying the users' needs for communication, entertainment, and mobility. These devices and their use are an integral part of today's life in industrialized societies. As a matter of fact it seems to be obvious that these services (e.g. email, music, news feeds, text messages, etc.) also find their way into the vehicle and are more and more used even while driving – careless of any legal restrictions. This leads to the point that the development and design of mobile and portable services which could also be used in the car should take this specific usage scenario into account.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 50–57, 2010.

Within this scenario the driver has to deal with several tasks and is confronted with various interfaces. First of all the interface provided by the car in order to fulfill the primary task [1] of driving the car, secondly integrated devices such as radio, navigation (GPS) devices or high-end entertainment systems and finally portable and mobile devices such as mp3 player, portable GPS devices and mobile phones. All these systems demand attention from the driver in order to be used whereas the primary driving task has to be fulfilled by the driver at all times. This poses on the one hand a large challenge to designers and developers of in-car services, since distraction from the primary driving task has to be kept as low as possible. On the other hand, the specific safety-related driving context also implies that the usage of the infotainment system might generally be interrupted in more demanding or even critical driving situations.

2 Challenge

Usually, services are developed due to design guidelines and ease-of-use principles to improve the use of the service in general and these prototypes are then tested within a driving situation and adjusted afterwards.

Within this paper the authors are proposing an approach based on a cognitive model which takes human capabilities into account in an early phase of the design process. An interdisciplinary team work (consisting of Psychologists, Designers and Engineers) is required to enable user centered engineering in order to create services which minimize the distraction caused by the visual sources [2, 3] and avoid costly adaptation in the late phases of service development.

If attention is needed to be divided between two tasks, Wickens [4] states that those tasks will interfere less if they demand different cognitive resources. Since driving as the primary task is visually and manually demanding, it is most suitable to use speech as the main interaction modality for secondary tasks like interacting with in-car infotainment systems. In line with that, Vollrath & Totzke [5] showed that processing acoustic information while driving has less negative effects on the driving performance than processing of visual information. While attending the auditory channel, drivers can keep their eye gaze on the road. But do voice user interfaces (VUI) actually minimize glancing at in-car displays? Kun et al. [6] examined the influence on visual distraction of the presence of a graphical user interface (GUI) of speech-based portable navigation devices (PND). In case of an available display, more visual distractions compared to a speech based PND without a display were identified although the secondary navigation task was accomplishable with speech interaction only. This leads to the question why users still focus some attention on the GUI in spite of a VUI with which their task is solely achievable. Even more important is the subsequent question how to avoid these visual distractions.

One possibility to avoid eye glances towards the GUI is to cover the display while driving. But this is contradictory to the fact that drivers prefer to have a GUI to get visual information [6]. Hence, taking the GUI away is not a solution. Furthermore, it is not clear what the reason for the glance to the GUI is. What kind of benefits does a graphical user interface provide compared to a voice interface? Based on a task analysis of the primary and secondary tasks while driving, mainly two situations seem to play an important role for requesting the visual feedback of the GUI: reorientation of the user to figure out at which process stage of the actual task he is and which input possibilities he has. The need for reorientation is a main characteristic of an in-car secondary task as the interaction might be interrupted by the primary task.

3 Attention Allocation between GUI and VUI in Dynamic Environments

According to Wickens [7] attention allocation in dynamic environments (like driving a car, or flying an airplane) is influenced by four aspects: salience, effort, expectancy, and value (so called SEEV model). These aspects affect the probability of drawing attention to specific regions (PA) in the following way:

$$P(A) = sS - efEF + (exEX + vV) \tag{1}$$

Salience represents the strength of the stimulus: the more salient a stimulus, the higher the probability of shifting attention to it. The effort of the user (e.g. long scanning paths) that needs to be invested to allocate attention to this stimulus inhibits P(A). Salience and effort are so called bottom-up processes, while expectancy and value are top down processes: if the expected bandwidth of a stimulus is high, the probability of attention shifting increases. Not only a high expectancy but also the importance of the stimulus (value) enhances P(A) in a top-down manor. Up to now, the SEEV model is evaluated for visual attention allocation [8] and can predict percent dwell times for different areas of interest, e.g. while car driving, with correlations from 0.65 to 0.98.

We assume that the effects of salience, effort, expectancy, and value on attention allocation do also play a role in the auditory domain. Therefore, theoretical considerations based on the SEEV model were taken into account to answer the question why the probability of attending the visual display (P(A)GUI) is obviously higher than the probability of shifting the attention to the VUI (P(A)VUI).

As “salience is a maximum for auditory events“ [9], there is no benefit for GUIs over VUIs in this aspect.

Considering the importance of the inherent information of GUI, respectively VUI, they seem to have the same quality and therefore it is assumed that they have the same value in the first place. But detracting the visual attention from the road has high costs which in turn implicates a lower value for attending the GUI.

Hence, neither salience nor value give reasonable causes for an increased benefit of allocating attentional resources to the GUI despite safety-critical issues instead of attending the VUI.

Since visual information is constantly present, so that the driver can initiate an attention shift at any time the expectancy to get information that is needed via GUI is very high. In contrast, a speech output is non-permanent and requires attention in the very moment it is presented, which reduces the expectancy to a value near zero. While in the special case of driving the interaction with the infotainment system is frequently interrupted by the primary driving task, the drivers expediently tend to use the expectancy high-valued GUI instead of the low-valued VUI for information acquisition.

Another difference between GUI and VUI is evident considering the aspect of effort, which is mainly expressed by time and cognitive effort required for attention allocation: a considerably high benefit for displays is the faster detection of information compared to voice outputs, because those require serial listening from the beginning to the end, while the relevant information cannot as easily be picked out as from a GUI. Furthermore, listening to a text generally is more time-consuming than reading a text [10]. Another aspect is that GUIs often provide more information compared to VUIs: different colours, fonts or grouping of information on the display transfers implicit information to the user (e.g. buttons in a specific colour or font mark input options). To present equivalent additional information via voice output, such kind of meta-information has to be provided verbally (e.g. “you have the following options”). This will increase the time needed for information acquisition and will also affect the cognitive workload for the user because he has to keep the verbal information in the phonological loop of his working memory [11]. This again might lead to annoyance and cognitive overload of the user. To avoid annoyance and increasing cognitive workload, it frequently results in design solutions of speech output prompts, which represent only parts of the whole information depicted on the GUI. In turn, this influences the value for the variable expectancy to receive information, because not all the information which is actually relevant for the user might be presented – the bandwidth of information is low. Usually these reduced ranges of prompts are limited to the actual interaction (like system information to give a brief orientation in the menu or a possible voice command that can be used to follow the main interaction path). Obviously there is a trade-off for designers of speech-based in-car infotainment systems: Increasing expectancy by presenting all the information of the screen via speech (increasing the bandwidth of information), which will lead to an increased cognitive and time effort, or decreasing the effort by presenting only a part of information which will decrease the bandwidth of information (decreased expectancy).

Summing up, voice outputs imply higher effort and lower expectancy, which might lead in a higher probability of drivers focusing their attention on the GUI of an infotainment system rather than on a VUI. Based on these considerations, implications for designing infotainment systems are deduced in the following part.

4 Design Recommendations

In a project of Deutsche Telekom Laboratories different mobile services were implemented on a G1 mobile phone. The application includes a graphical user interface as well as a voice user interface and can be solely operated by speech input and output. If the driver intends to give a speech input he has to press a button located at the steering wheel to activate the automated speech recognition (push-to-activate button). In the development process the VUI was especially designed to gain a lower probability of attention shifts towards the display and as a result off the road. To achieve the objective it was intended to increase the expectancy and to decrease the effort. The “email” application was selected in order to derive design considerations and will be described in the following (Figure 1).

Fig. 1. Screen shots of the email application

4.1 Increasing the Expectancy

As mentioned before, not all information that is represented on the screen is usually available by using speech input. This fact and also the single representation of voice output reduce the user's expectancy to get all information via the VUI. This in turn leads to a lower probability of focusing attention on the screen. To increase the level of expectancy, all information on the screen needs to be available for the driver so it can be requested anytime. To avoid a higher manual effort, an additional functionality was introduced to the push-to-activate button: while the speech recognition is activated with a short-term push of the button, a long-term push of this button will provide more elaborated verbal information. The aim is to represent all information which is present on the GUI – not only information that is related to the actual interaction. Thus, there are two different output levels:

- direct speech prompts for the main interaction path (system initiated speech output)
- long-term push prompts to get all information presented on the screen (driver initiated speech output)

This concept changes the way voice output is usually provided to the driver. Typically speech output is system-driven while in the system described the driver initiates the request of speech output. With the help of the long-term push on the push-to-activate button, the driver can get the desired information whenever he intends to do, which increases the expectancy of this information up to 100%. The user is now able to orientate him- or herself again after interruptions through the primary driving task. Furthermore, missing a system initiated direct speech prompt is no longer fatal, because relevant information can be repeated on the long-term push prompt. This also reduces the cognitive demand of attending and memorizing the system output.

4.2 Decreasing the Effort

To present all the information via voice output increases the expectancy, but this in turn involves an increased time effort for speech output. In order to decrease that amount of time and to provide all the information presented on the screen for the driver without provoking an increased mental workload, there are three different kinds of design solutions that we developed for the VUI:

- non-speech sounds
- up-tempo speech
- acoustics instead of visualization (semantically enriched information representation)

Non-speech sounds. According to Brewster [12] “icons can present information in a small amount of space compared to text; non-speech sounds can present information in a small amount of time as compared to speech”. There are three main types of nonverbal sounds: auditory icons, earcons and spearcons.

Auditory icons are metaphorical representations of a word or a concept [13]. Important for the use of auditory icons is a strong intuitive link with the word that should be presented through the auditory icon. The stronger the association with the word the less learning is required.

In the project presented, we tried to replace speech output with auditory icons whenever there was a strong and clear associated sound available. For example, instead of the speech output: ”Your Mail has been sent” there is a wind sound played - comparable to an object flying through the air.

Earcons are short musical motives that represent menu items. They differ in the combination of rhythm, pitch, timbre, register and dynamics.

To achieve the same rapid information acquisition about the current status in the menu as in case of glancing at the screen earcons were used. For example: every menu item on the start screen (email, sms, telephone and news) has its own sound motive. The four sound motives exhibit a high differentiation. However, to represent the single steps of the main interaction path (e.g. writing an email) within the specific menu items, the particular sound was varied by changing the pitch. The pitch rises with every new step that brings the user to the end of the operation. Since absolute data for pitches is hard to perceive for humans, we present every increased pitch with the sounds of the preceded interaction steps. This in turn not only reduces the cognitive workload caused by detecting a difference between the sounds, it also enables the user to associate the sound with one of the interaction steps by counting the number of sounds. However, it is important to take into account that earcons “are abstract so their meaning must always be learned” [12]. Therefore the earcons that will be used for orientation in the menu in the email application initially need to be presented with a speech prompt.

Spearcons are speech audio cues that are sped up to the point speech is no longer comprehensible. As the sound of the accelerated word is unique – like a fingerprint – it is very easy to produce sounds that are different enough from each other. Spearcons as well as earcons could also be used for orientation in the menu and also need to be learned before presenting them on their own.

One important aspect of non-speech sounds is the consideration of annoyance and user preference. Especially since in-car infotainment systems are comfort systems which add various requirements to the design in terms of acceptance and joy of use. To decide which kind of nonverbal sound will finally be applied to the email application as well as performance indicators, user acceptance has to be tested.

Uptempo Speech. Speeding up the paste of the voice output is another design solution for the VUI to decrease the time effort of speech output. This in fact should solely be used with content that the user has knowledge of. In the described project we sped up content that had already been heard via the direct system initiated speech output. For example: after opening a new email the user gets a direct feedback from the system. The voice output reads the email header via text-to-speech. If the user holds the push-to-activate button he gets all information on the screen and so the email header can be requested again, but this time it plays faster.

Acoustics instead of visualization. As mentioned before, by using different colours, fonts or grouping items together, a GUI can represent additional information to the user. In the described project, different voices and pauses were used to represent such indirect information. Buttons or interaction elements on the GUI were magenta coloured. By drawing visual attention to the screen, the user knows which next interaction options are available. Additionally, he or she knows which speech commands he can use devoid of being told “you have the following options…” (meta-information). As mentioned before, we do not only want to let the user know via direct speech prompts which are the next possible interaction steps to follow the main path of interaction (as it is frequently used in speech based infotainment systems). We also want to enable the user to hear all the speech commands he or she can use. By using a different voice for interaction elements than for other information, the information is implicitly clustered and does not require further explicitly verbal announcing. This has a positive effect on the time effort, because prompts generally shorten while the content of information stays the same.

Variations in the duration of pauses were used to represent the graphical grouping of elements in different locations on the screen. To give an example: in the email in-box the system pauses are shorter between reading out the possible interaction elements “replay” and “forward” compared to the length of the pause between these two options and the navigation interaction elements “next” and “previous”.

Another way of presenting the implicit information given on the screen can be found in spatial representations of nonverbal sounds or speech. For example, instead of using different voices for interactional elements as well as other information, this content could be represented via different spatial alignments of these speech prompts. A critical aspect by using spatial representation of sounds for in-car infotainment systems could be seen in the stronger orientating responses towards the source location of acoustic cues rather than visual cues [14]. Since it is not intended to affect these orientation responses while driving, it was initially renounced to semantically enrich sounds and speech prompts by spatial representation for the email application.

5 Conclusions and Implications

The present project focuses on the question of how to prevent drivers from allocating their attention to displays of in-car infotainment systems if a VUI is available. The design recommendations that were made by taking the SEEV-Model [7] into account allow the integration of human factors in an early phase of the engineering process. Furthermore in a next step the design recommendations are going to be tested in a driving simulator study to iterative verify if the design goals (to develop an intuitive and safe interface for in-car use) are achieved. Aim of further studies will be to answer the question, which of the above mentioned methods for decreasing the time effort can substitute graphical information best.

Acknowledgments. Our thanks to Frank Oberle, Thomas Scheerbarth, Stefan Seide, Jianshen Zhou and Felix Burkhardt.

References

1. Bubb, H.: Fahrerassistenz - primär ein Beitrag zum Komfort oder für die Sicherheit? Der Fahrer im 21. In: Jahrhundert, vol. 1768, pp. 25–44. VDI Verlag, Düsseldorf (2003)
2. Wierwille, W., Tijerina, L.: Eine Analyse von Unfallberichten als ein Mittel zur Bestimmung von Problemen, die durch die Verteilung der visuellen Aufmerksamkeit und der visuellen Belastung innerhalb des Fahrzeugs verursacht wird, pp. 164–168. Zeitschrift für Verkehrssicherheit (1995)
3. Wierwille, W., Tijerina, L.: Darstellung des Zusammenhangs zwischen der visuellen Beanspruchung des Fahrers im Fahrzeug und dem Eintreten eines Unfalls, pp. 67–74. Zeitschrift für Verkehrssicherheit (1997)
4. Wickens, C.: Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3, 159–177 (2002)
5. Vollrath, M., Totzke, I.: In-vehicle communication and driving: An attempt to overcome their interference. Driver Distraction Internet Forum Sponsored by the United States Department of Transportation (2000)
6. Kun, A.L., Paek, T., Medenica, Z., Memarovi, N., Palinko, O.: Glancing at personal navigation devices can affect driving: experimental results and design implications. In: Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications. ACM, Essen (2009)
7. Wickens, C.D., Goh, J., Helleberg, J., Horrey, W.J., Talleur, D.A.: Attentional Models of Multitask Pilot Performance Using Advanced Display Technology. Human Factors: The Journal of the Human Factors and Ergonomics Society 45, 360–380 (2003)
8. Wickens, C., McCarley, J.: Applied attention theory. CRC Press, Boca Raton (2007)
9. Wickens, C., McCarley, J., Alexander, A., Thomas, L., Ambinder, M., Zheng, S.: Attention-Situation Awareness (A-SA) Model of Pilot Error. In: Foyle, D., Hooey, B. (eds.) Human Performance Modeling in Aviation, pp. 213–239. CRC Press, Boca Raton (2007)
10. Kozma, R.B.: Learning with Media. Review of Educational Research 61, 179–211 (1991)
11. Baddeley, A., Hitch, G.: Working memory. In: Baddeley, A., Hitch, G., Bower, G. (eds.) Recent Advances in Learning and Motivation, vol. 8, pp. 47–90. Academic Presse, New York (1974)
12. Brewster, S.: Nonspeech auditory output. In: Jacko, J., Sears, A. (eds.) The Human Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 220–239. Lawrence Erlbaum Associates, Mahwah (2002)
13. Gaver, W.W.: Auditory icons: using sound in computer interfaces. Human-Computer Interaction 2, 167–177 (1986)
14. Proctor, R., Tan, H., Vu, K., Gray, R., Spence, C.: Implications of compatibility and cuing effects for multimodal interfaces. Lawrence Erlbaum Associates, Mahwah (2005)

Structured Digital Storytelling for Eliciting Software Requirements in the ICT4D Domain

Daniel Sinnig[1], Kristina Pitula[1], Richard Becker[2], T. Radhakrishnan[1], and Peter Forbrig[2]

[1] Concordia University, Montreal, Canada
{d_sinnig,pitul_87,krishnan}@encs.concordia.ca
[2] University of Rostock, Germany
{richard.becker,peter.forbrig}@uni-rostock.de

Abstract. Due to the specific challenges that ICT4D projects present, conventional requirements elicitation techniques are often inapplicable or insufficient. We address this shortcoming by introducing *structured digital storytelling* (SDS) as an alternative elicitation technique especially suited for the ICD4D domain. SDS is supported by our mobile elicitation tool designed around the needs and capabilities of the targeted user population. We embed SDS in a requirements elicitation and specification process which commences with an initial domain analysis of the needs, goals, and tasks of the target population with respect to the high-level social development goals. The outcome of the preliminary domain analysis is used to formulate a set of questions for guiding the story narration. Once the stories are elicited, they are processed and the results are fed into a traditional requirements specification process. In this paper we propose a task-analytic approach for determining the topics of the guiding questions.

Keywords: Requirements Engineering, Requirements Elicitation, Structured Digital Storytelling, ICT4D.

1 Introduction

Requirements engineering is an important stage in any software development. It is more so in the case of software development for social development projects in rural areas of developing countries. Currently, a significant effort is being put into providing people in rural areas with access to digital content and services by using Information and Communication Technologies (ICT). Such projects, which seek to apply ICT in redressing socioeconomic inequalities, are referred to as "ICT for Development" or ICT4D. This article is concerned with requirements engineering in the ICT4D domain.

Unfortunately, often ICT4D projects fail because their development was driven by the technology available and not by the very needs and social problems of the people living in rural communities [1]. Existing technologies are often applied in a non-inclusive manner with respect to the local population, without sufficient adaptation or re-invention, and often without regard for user's needs and their social contexts.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 58–69, 2010.

Many of the failures are attributed to requirements miss-management [2] and could have been avoided if suitable requirements elicitation methods were applied.

It has been shown that traditional elicitation techniques assume that end users are able to understand and articulate their problems and needs in a reflective manner. Expressing oneself, with the right problem descriptions and the right needs, however, are all learnt skills, which because of their socioeconomic situation, people in rural communities may not have had the opportunity to develop [3]. An oral based mode of communication is more intuitive among such communities [4].

In order to address this shortcoming, we propose using *storytelling* as an alternative and complementary requirements elicitation technique. Storytelling is successfully applied in multiple domains [5;6], and is particularly suitable for people with limited literacy to express themselves and their needs. For the sake of eliciting requirements, we adapt the traditional storytelling concept with a task-oriented specialization and provide tool support, which leads to our notion of structured digital storytelling (SDS). With SDS, the narration is guided by a set of thematic questions. Before SDS can be applied, it is important to narrow the topic of the story to be told. For this purpose, a preliminary domain analysis is carried out determining needs, goals and tasks of the target population related to the project's high-level social development goals. Based on the task hierarchy a set of appropriate open-ended questions is selected and presented to the rural population. The elicitation phase is supported by a software tool (such as our E-Tool), which serves as a facilitator. The collected stories are analyzed and the resulting information is fed into a traditional top-down development process.

The remainder of this paper is structured as follows. The next section reviews relevant background information about requirements engineering and digital storytelling. Section 3 introduces the E-Tool for eliciting stories in a rural context. Then, in Section 4, we demonstrate how SRS can be embedded into a requirements elicitation and specification process. Section 5 discusses the case studies we have conducted. In Section 6, we discuss relevant related work. Finally, we conclude and provide an outlook to future work.

2 Background

In this section we remind the reader of the importance of proper requirements management and discuss key challenges when applied within an ICT4D context. We then describe key concepts of digital storytelling and discuss its advantages and disadvantages compared to traditional requirements elicitation techniques.

2.1 Requirements Engineering in the ICT4D Domain

Requirements engineering (RE) is a crucial step in the development of any software system. It is the process whereby the intended purpose of a system is discovered and documented so that it can be analyzed, communicated and eventually culminate in a software implementation that meets that purpose. How well that purpose is met is the primary measure of a system's success. Thus, RE is essential in determining what a system will do and how this will be measured. The process is inherently iterative, and consists of three major activities: elicitation of needs, requirements specification and requirements validation. The process starts with some ill-defined 'ideas' of what the system should do. These are elicited, analyzed, and systematically transformed into a technical requirements specification that defines the software system to be built completely and unequivocally.

The RE discipline offers a wide range of established methods and techniques for accomplishing the various activities, appropriate for different problem domains and development styles. The early involvement of end-users is a well-established principle of software engineering, with standard methods (such as interviews, workshops, focus groups and ethnographic studies) to facilitate the elicitation and communication of software requirements between stakeholders and analysts.

Eliciting software requirements in an ICT4D context poses a number of challenges due to the nature of the projects. Such projects target marginalized communities with the goal of assisting them in improving their socio-economic situation. The projects are driven by high-level social and economic development goals that almost invariably are initiated from outside the targeted community. Many involve multiple stakeholders such as social workers, agronomists, government and business representatives, etc. coming from the public, private, and non-profit sectors, and working in partnership to be the most effective. The intended beneficiaries typically have limited schooling, low literacy levels, and low disposable incomes while the developing regions and countries where the projects take place are characterized by inadequate infrastructures, intermittent power and connectivity, underdeveloped economic markets and distribution and support networks, and a lack of trained personnel.

2.2 Structured Digital Storytelling

Storytelling has been proven to be useful for people with limited literacy to express themselves and their needs. A recent study by Kerr [7] demonstrated the potential of digital storytelling for expressing community information, issues and frustrations; the very sources that will shed light on the real needs of users in rural areas. At its simplest, storytelling consists of someone telling their personal story on some topic and the narration being recorded. A facilitator may be present to ensure that items of interest are clarified and expanded. Recent digital technologies support the authoring of sophisticated multimedia stories that can be made accessible to a broad audience. Interactive Voice Response (IVR) systems that use structured dialogues to ask a sequence of questions, are an alternative approach for collecting user input..Our approach combines the two. We propose adding a multimedia, structured dialogue interface onto digital storytelling technology to assist people in expressing their information needs through stories which can then be shared in the community. Instead of asking direct questions about their information needs, the target population's needs can be elicited through a series of questions, short stories, 'what if scenarios', or by hearing their neighbors' views regarding the problems they face and their potential solutions. The structured dialogue ensures that relevant themes are covered, while hearing stories told by their neighbors will inspire people to tell their own stories.

In an ICT4D context, storytelling has a number of advantages over commonly used requirements elicitation techniques such as interviews, focus groups and ethnographic studies [8]. A major difference between SDS and the other elicitation techniques is that the narrator is largely left on their own to tell their story in their local language.

Among the advantages we foresee with SDS is that storytelling capitalizes on the villagers' primary mode of communication. Although the questions provide some general guidance, an interviewer is not present to influence the narration. Thus, it is possible to identify problems and needs not initially envisaged and contextual factors that might otherwise be overlooked. Moreover, collecting stories involves fewer resources in terms of facilitators, preparation, and elapsed time, and it does not require facilitators who speak the local language.

A disadvantage with respect to other elicitation techniques is that there is no one present to provide clarifications, guidance or immediate follow-up on items of interest. Additionally, participants may focus their story on a single aspect leaving other equally relevant aspects unmentioned.

3 The Mobile E-Tool

In order to facilitate structured digital storytelling, we have developed a mobile tool, the E-Tool. It is a Java application and runs on cell phones with support for audio/voice recording. The motivation behind choosing a mobile elicitation device is the high penetration and ubiquitous availability of cell phones both in urban and rural contexts, especially in developing countries. For example, mobile phones have a 65% market penetration rate in India[1]. The technical requirements for the application are fulfilled by almost every low-budget cell phone available today..

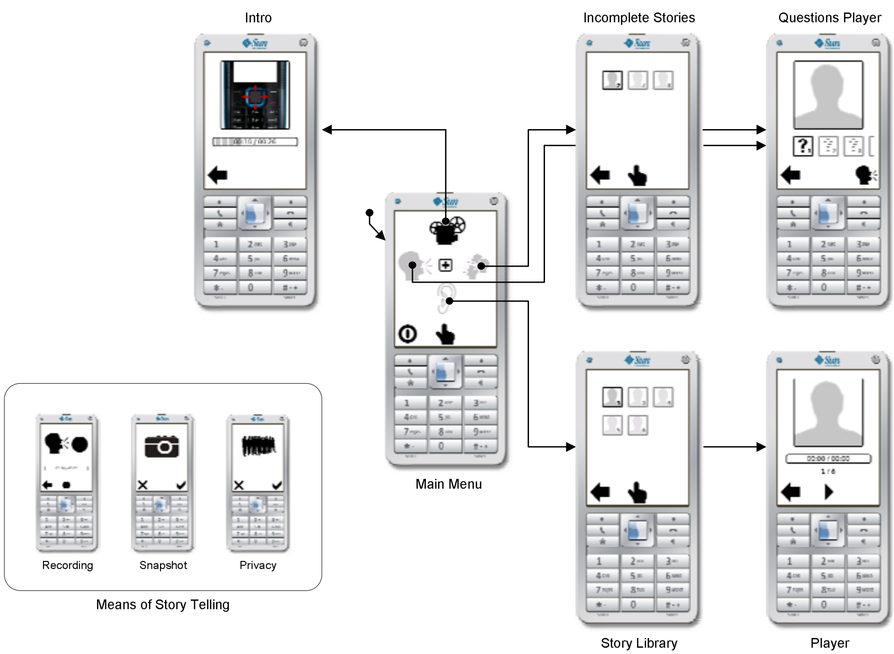


Fig. 1. Features and Navigation Structure of the E-Tool

[1] Telecom Regulatory Authority of India.

The E-Tool is designed to be easy to use by a non-literate population. It provides (1) an introductory video explaining the reason for collecting the stories as well as an overview of how to operate the application, (2) a story library where the stories are stored and villagers can listen to them, and (3) an interview feature which guides users through a series of questions, letting them tell their own story. To keep the UI simple, it is based on a text-free design relying on iconic symbols, numbers from 0 to 9, and audio prompting instead. Navigation is made quite simple by using symbols that are native to the specific community of people. Some of the features of the E-Tool as well as its navigation structure are portrayed in Figure 1.

While the absence of human facilitator bears the risk that the narrator will drift from the actual topic, we believe that without supervision the narrator has more liberty in determining the story's direction and thus put importance on things of his choice. Among the challenges to address, storytelling is by nature a social exchange that is enriched by the presence of an audience. As with any technique, consideration must also be given to confidentiality and self-censure, particularly when dealing with sensitive subjects. This can be achieved by providing each participating local community with only one cell phone at a time. The E-Tool offers the option to share stories in a public library or to keep them in a private space.

4 SDS-Driven Requirements Elicitation and Specification Process

In this section we detail a requirements management process based on the ideas and principles defined in the previous sections. That is, the process promotes the application of *structured digital storytelling* in an expanded context to elicit requirements from users. Our process is illustrated through a set of real-word models and artifacts stemming from a case study that conducted with farmers in rural India.

4.1 Preliminary Analysis of Needs, Goals and Tasks

RE is by nature difficult because it starts in a largely unconstrained problem space and much of the effort revolves around defining system boundaries, identifying relevant environmental conditions and prioritizing and selecting which requirements to pursue among the many possible options [9]. In the ICT4D domain requirements elicitation generally starts with determining what categories of information are vital to a rural society in some prioritized order. These categories identify the potential areas of intervention for a project, and invariably relate to a community's economic activity, and the health and welfare of its families.

In applying SDS, it is important to determine what topics to ask participants to speak about so that we can obtain a better understanding of their needs in the context of their activities, and thus bring out their specific concerns, issues and problems in situ. For this purpose, based on the high-level social development goals and with the assistance of domain experts, a preliminary domain analysis is carried out determining high-level user needs, goals and tasks. Needs are further decomposed into a set of goals. According to [10], a goal denotes a desired 'state' and is achieved by satisfying all of its sub-goals. Goals may support each other or be in conflict, and constraints will apply to certain goals.

Under the premise of understanding the needs and goals of village farmers in rural India, we conducted an on-site domain analysis. The primary objective was to understand the root causes behind apparent problems and to relate them to a set of needs and goals. This comprised a detailed investigation of each contributing problem and a quantification of its individual impact. The analysis was carried out through a series of consultations and brainstorming sessions with domain experts as well as through observations and discussions with local farmers - the intended beneficiaries of the project.

Figure 2 portrays a subset of a needs-goal model. Farmers' needs are satisfied if the corresponding (a) farming, (b) financial, and (c) household goals are achieved. The farming goal is comprised of the goals associated with growing crops such as access to water and land, seed, fertilizers, pesticides and resources as well as the financing necessary to cover the significant upfront investment required to buy supplies, most often through loans. The financial goal is comprised of the goals arising from financing child education, parents' health care and farming activities. The household goal is comprised of the goals to educate children, care for aged parents and so forth.

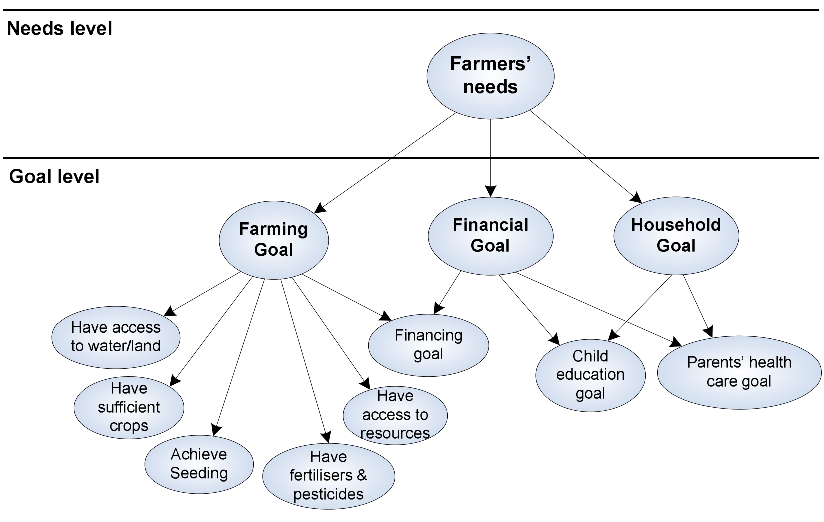


Fig. 2. Needs-goal model for Village Farmers in India

The needs-goal model is useful for representing the desires and concepts in the user's world; however it does not provide much insight into what activities they do. Therefore, as a next step, we associate each high-level goal with a task model that defines what tasks/activities are involved in attaining that goal. Most task modeling formalisms like HTA [11], GOMS [12], GTA [13], or CTT [14] make it possible to decompose tasks hierarchically into sub-tasks and to specify temporal constraints.

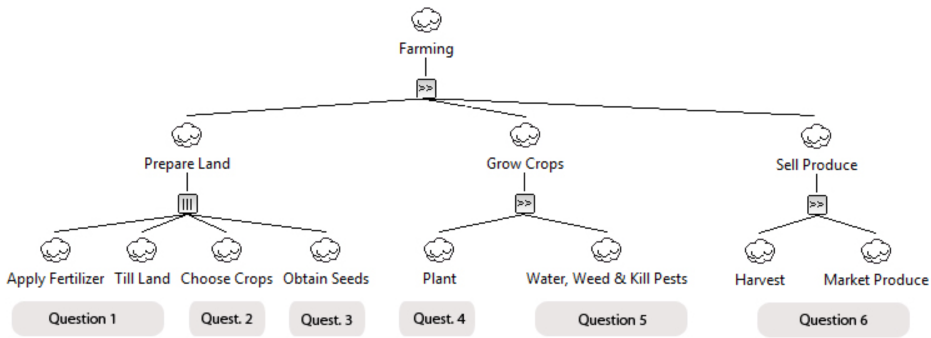


Fig. 3. Task Breakdown for Attaining the Farming Goal

The task-model depicted in Figure 3 provides a coarse-grained decomposition of the various activities necessary to attain the high-level goal of "Farming". In the example, the various subtasks are related by two temporal operators *enabling* (») and *concurrency* (|||) denoting sequential composition and parallel composition, respectively. To farm, one must first prepare the land before one can grow crops and subsequently sell the produce. Preparing the land involves applying fertilizer, ploughing, choosing crops, and obtaining seeds. Growing crops includes the actual planting and the subsequent watering, weeding and elimination of pests; and finally, the sale of produce encompasses its harvesting and marketing.

4.2 Question Selection and Requirements Elicitation

Once a clear understanding of the high-level needs, goals and tasks of the intended beneficiaries has been obtained, a set of corresponding questions is formulated. The questions are chosen based on the tasks in the task model (as indicated in Figure 3) while the question context is obtained from the corresponding goals and sub-goals (Figure 2). By relating the questions to a particular context in which the task occurs, the questions will be closer to the farmers' actuality. Table 1 depicts a subset of the questions presented to village farmers in one of our case studies. Each question consists of a set of sub-questions and is suffixed by the request "Please tell us about such things" - inviting the participant to tell a comprehensive story.

Once the set of questions has been chosen, the collection of stories is relatively straightforward. An application (such as the E-Tool presented in the previous section) that plays the questions and records responses can be made available on a suitable mobile device in a location such as a community center, and villagers invited to try it. The various questions will be presented in the temporal order of the corresponding tasks so that the farmers are prompted to tell their story within a framework that is situated with respect to what they do. Participation can be encouraged by enlisting the support of respected members of the community and having them record their own stories to serve as examples in the story library. The time involved corresponds to the time it takes to tell the story, allowing a reasonable number of stories to be recorded in a few days. Because resources such as interviewers and facilitators are not involved, the application can readily be deployed in a number of villages, increasing the number of the stories collected and coverage of issues. Different topics can be addressed by changing the application prompts while different linguistic regions are easily supported by simply translating and rerecording the application prompts, making this elicitation method highly adaptable.

Table 1. List of Questions Presented to Village Farmers in Rural India

Q1	"How many acres is your land? Do you own it or lease it? Do you have enough water? How does water affect your yield? How do you prepare the land for planting? Do you use fertiliser? Do you have difficulty getting it? Do you use manual labor or machines? Is it costly or difficult? What if anything do you wish to be better? Please speak about such things."
Q2	"What are the different crops you cultivate? How many months for it to mature? How much is the yield when it is the best? Do you cultivate always one type of crop? If someone helps you to cultivate a different crop that can be more profitable, will you try? What kinds of help will you need, like new seeds etc? Please speak about such things."
Q3	"How do you obtain the seeds for each crop? Do you buy or make your own? How do you get good quality seeds? Please speak about such things."
Q4	"To plant a crop, what expenses do you have? Do you have enough money or do you have to borrow? If borrowing at what interest rate? Please speak about such things."
Q5	"Do you always have enough water for your crops? How does water affect your yield? Tell us about the weeds and insects destroying or affecting your crop. How do you tackle them? When there are too many insets, whom do you consult? How do the pests affect your yield? Please speak about such things."
Q6	"To harvest, do you use manual labor or machines? How much yield (bags or kilos) do you get in a bounty crop? What do you do with the crop after harvest? Where do you sell, in the 'mahndy' or to the local middle man? Where is the price better? Do you have trouble to transport the produce? Please speak about such things."

4.3 Requirements Abstraction and Specification

Following the elicitation phase, the stories collected are processed, analyzed and categorized. Since stories are told in people's local language, the first step after data collection consists of translating the stories into English. This is projected to be a manual activity, performed by local individuals (ideally from the same rural community) who possess an adequate degree of literacy in both the local language and English. Although relatively labor intensive, this process is straightforward and does not require highly skilled labor.

Following this, the transcriptions are analyzed to determine what issues the participants mention, their importance, and the various factors that play a role therein. It is here that a skilled qualitative analysis is required, to identify the various issues and related factors or conditions. For example, farmers complain of unfair pricing at the markets where they sell their produce. At the same time, farmers who have taken out loans are compelled to sell to the lender, thus bypassing the open market system. Any attempt to address the issue of unfair pricing must also consider how many farmers are constrained by loans and how they might benefit from a possible solution. Such issues are prioritized based on how critical they are and the frequency with which they are mentioned, and these along with the other factors are related back to the high-level goals. Additionally needs, domain concepts and desired features are abstracted as well. Due to the often large amount of information, we plan to employ manual as well as automated or semi-automated natural language processing to extract the information stated above. The applicability of corpus linguistics to document processing in requirements engineering has been shown in several problem domains and at different levels. E.g., Rolland and Proix [15] provide a general background for the applicability of natural language processing to requirements engineering.

Throughout the process of requirements production and refinement, statements from the different sources are integrated into a single set of requirements in the final *software requirements specification* (SRS). Finally, the SRS is validated to ensure that it reflects stakeholders' needs, and is sound and complete. The former is achieved by establishing traceability links to the user stories gathered during elicitation. It is ensured that each user need is addressed by at least one requirements statement and any unnecessary system features unrelated to any user needs are excluded. It is important that the requirements engineers remain in close contact to domain experts who will be able to reconcile differing goals and needs.

5 Case Studies in Rural India

In order to test the viability of SDS for requirements elicitation in an ICT4D context, we conducted three field studies in two rural areas of India using a prototype laptop version of the elicitation tool described in Section 3. In India there are wide disparities in socioeconomic conditions. Whereas the new economy centered on urban areas is giving rise to a comfortable middle class (and aspirations to those below), the traditional economy of the rural areas, largely based on farming, has left much of the rural population in extreme poverty with few options out [16;17].

We elicited stories on two different topics in two different rural regions with distinct regional languages. A total of three studies were conducted – two near Chennai in Tamil Nadu state and one near Bangalore in Karnataka state. All three took place in farming villages typical of the rural Indian context. In Tami Nadu, stories on farming and higher education were elicited while in Karnataka the topic was farming only. All together 30 stories were collected, 17 on farming and 13 on higher education. These were told by both male and female participants representing a broad age range, from children to the elderly, and a cross section of financial situations, from the very poor to those considered well off by local standards.

The acceptability of the SDS approach exceeded our expectations. All the participants were able to tell their story, and were enthusiastic about doing so. Villagers participated readily and quickly picked up the operation of the application. Once they began talking, they became engaged in telling their story and were not distracted by the mechanics of recording. While in almost all cases they participated in groups, their stories were highly personal and did not show any signs of "groupthink". At the same time, the group provided an audience for the teller, making the narration a natural communicative exchange.

Our analysis of the stories indicated that they were highly useful in identifying and prioritizing the participants' concerns, and revealed an abundance of contextual information regarding their activities. While a discussion of this experiment is out of scope here, the results of our study are presented in [18]. Additionally, by conducting this experiment on different topics in different regions and languages, we demonstrated that this approach is both repeatable and readily adaptable to new contexts in terms of population group, language, and topic.

6 Related Work

In the case of ICT4D projects, the characteristics of the targeted user populations make it difficult to apply standard elicitation techniques. To date, although there is a growing body of research related to software development and deployment across national boundaries, little work has been done in the area of cross-cultural requirements gathering with stakeholders from disadvantaged socioeconomic backgrounds.

Over the past decade, although numerous ICT4D projects have been attempted, few have achieved long-term sustained success. A too great emphasis on technical success with inadequate consideration of end-users' needs and of the social development aspect of the projects are among the factors that have contributed to this lack of success [1]. The need to involve intended beneficiaries in determining project goals and constraints has long been recognized in social and economic development circles, and a number of participatory approaches such as *Participatory Rural Assessment* (PRA) based on the work of Chambers [19] exist for accomplishing this. However, such approaches are primarily geared towards identifying and introducing social and economic interventions and would need to be adapted to assist in determining software requirements. Furthermore, participatory approaches rely on techniques such as interviews, focus groups and community meetings, and thus encounter the same disadvantages as previously described when used for eliciting requirements in an ICT4D context.

In sociology, storytelling has proven itself as a means for developing a collaborative analysis from the bottom-up. In a recent study by Kerr [7] it was applied to identify the problems and constraints of homelessness as perceived by the homeless themselves. The resulting analysis revealed a number of significant issues that do not emerge from conventional top-down analyses where input is solicited from people such as social service providers, public officials and academic experts. There was a comparable divergence in the nature of potential solutions and associated issues as viewed from the top-down versus bottom-up. Furthermore, the research process of telling and listening to stories served as a catalyst to the homeless to become active in changing their situation.

In the ICT4D context, digital storytelling has been proposed as a means for sharing information among semi-literate people in rural villages. A study by Frohlich et al. [2] has shown the viability of storytelling as a means of communication in rural India. The study showed that villagers were enthusiastic about creating and listening to stories. Interestingly, the study also revealed a certain tension between those interested in creating and disseminating serious "development" content, and others more interested in creating personal and cultural content for entertainment purposes.

More recently, storytelling has been proposed as a means for eliciting requirements in domains where access to end-users and the actual context-of-use is restricted. In

[6], the author describes a field study in the healthcare domain in which the requirements elicited using focus groups and interviews are compared to those elicited using focus groups and stories. The focus of this study was to determine any differences in the number, breadth and depth of themes addressed and the amount of time required by participants. The study concluded that there was no significant difference with respect to the number and breadth of themes addressed. However, storytelling and focus groups were more effective in eliciting more diverse context-of-use and social information, and storytelling required less time than interviews.

7 Conclusion

Due to the specific challenges that ICT4D projects present, the sole use of conventional requirements elicitation techniques is often insufficient. In this paper, we addressed this shortcoming and proposed structured digital storytelling (SDS) as a complementary technique for eliciting software requirements. Storytelling capitalizes on the primary mode of communication of the people living in rural areas. As such it is particularly suited for participants with low literacy. Compared to traditional elicitation techniques, the narrator is largely left on their own to tell their story in their local language. The use of an SDS approach does not exclude the use of other elicitation techniques. Indeed we see it as complementary, assisting stakeholders in validating that they are focusing on the right problem, and that the problem is thoroughly understood in the context in which it occurs.

SDS is facilitated by our mobile E-Tool, which allows for recording and playback of elicited stories. The tool is designed to be easy to use by a non-literate population and thus is equipped with a minimalistic, text-free user interface, which instead of text, uses graphical icons, buttons with distinct colors and audio prompts to identify and guide users through the various options. We have embedded SDS into a requirements management process, which starts off with an initial domain analysis phase during which high-level needs and goals of the rural population are elicited. Goals are further refined by a set of tasks performed to achieve the goals. Based on the various tasks a set of questions is selected which are used to configure the E-Tool. Once the stories are elicited, they will be processed and the information gained is fed into a traditional top- down requirements specification process.

We demonstrated utility of SDS in an ICT4D context by conducting three case studies in rural India using a prototype laptop version of the E-Tool. This will be followed by a field study whose aims are to validate the proposed requirements management process and to determine the usability of the mobile (cell-phone based) E-Tool. One challenge that remains to be addressed is the remote configuration and remote collection of the recorded voice clips.

References

[1] Heeks, R.: CT4D 2.0: The Next Phase of Applying ICT for International Development. IEEE Computer 41(6), 26–33 (2008)

[2] Frohlich, D.M., Rachovides, D., Riga, K., Bhat, R., Frank, M., Edirisinghe, E., Wickramanayaka, D., Jones, M., Harwood, W.: StoryBank: mobile digital storytelling in a development context. In: Proc. of Human Factors in Computing Systems, Boston, MA, USA, pp. 1761–1770. ACM, New York (2009)

[3] Pitula, K., Radhakrishnan, T.: A Multimedia Tool to elicit Information Needs in Rural Communities. In: Proc. of HCI for Community and International Development Workshop at CHI 2008, Florence, Italy (2008)
[4] Ong, W.J.: Orality and literacy: The technologizing of the word. Routledge, New York (2002)
[5] Snowden, D.: Story telling: an old skill in a new context. Business Information Review 16(1), 30–37 (1999)
[6] Gausepohl, K.A.: Investigation of Storytelling as a Requirements Elicitation Method for Medical Devices, Masters Thesis in Industrial and Systems Engineering, Virginia Polytechnic Institute (2008)
[7] Kerr, D.: We Know What the Problem Is: Using Oral History to Develop a Collaborative Analysis of Homelessness from the Bottom Up. The Oral History Review 30(1), 27–45 (2003)
[8] Leffingwell, D., Widrig, D.: Managing software requirements: a use case approach. Addison-Wesley, Boston (2003)
[9] Cheng, B.H.C., Atlee, J.M.: Research Directions in Requirements Engineering. In: Proc. of Workshop on the Future of Software Engineering (ICSE 2007), pp. 285–303 (2007)
[10] Van Lamsweerde, A.: Goal-Oriented Requirements Engineering: A Guided Tour. In: Proc. of Fifth IEEE International Symposium on Requirements Engineering, pp. 149–163 (2001)
[11] Annett, J., Duncan, K.D.: Task Analysis and Training Design. Occupational Psychology 41, 211–221 (1967)
[12] Card, S., Moran, T.P., Newell, A.: The Psychology of Human Computer Interaction (1983)
[13] Veer, G., Lenting, B., Bergevoet, B.: GTA: Groupware Task Analysis - Modeling Complexity. Acta Psychologica 91, 297–332 (1996)
[14] Paternò, F.: Model-Based Design and Evaluation of Interactive Applications. Springer, Heidelberg (2000)
[15] Rolland, C., Proix, C.: A Natural Language Approach for Requirements Engineering. In: Loucopoulos, P. (ed.) CAiSE 1992. LNCS, vol. 593. Springer, Heidelberg (1992)
[16] Rezwan (2009). India: A wave of suicides among farmers. Global Voices Online (2009)
[17] Robinson-Sunna, S.: Seeds of Despair. Time.com (2007)
[18] Pitula, K., Radhakrishnan, T.: On Eliciting Requirements from End-Users in the ICT4D Domain. Submitted for Review to Journal of Requirements Engineering (2010)
[19] Chambers, R.: Rural Development: Putting the Last First. Prentice-Hall, Englewood Cliffs (1983)

Experiencing User-Centered Design (UCD) Practice (Case Study: Interactive Route Navigation Map of Bangkok Underground and Sky Train)

Waralak Vongdoiwang Siricharoen

Computer Science Department, School of Science,
126/1, Vipavadee-rangsit Road,
University of the Thai Chamber of Commerce
Bangkok, Thailand
waralak_von@utcc.ac.th

Abstract. The research is aim to create MRT: Metropolitan Rapid Transit Authority of Thailand (Underground Train) and BTS: Bangkok Mass Transit System Public Company Limited (Sky Train) interactive route navigation map. This map shows the whole routes of both MRT and BTS train stations which are located in the interactive manner. The design is not only a proposal for showing the whole information about direction and price information of transportation by MRT and BTS in Bangkok. But also this research took these concepts one step further, by introducing interactive map which the viewers/passengers can access the information from everywhere via the internet. The goal of this interactive navigation map is to satisfy informative. Problem space includes 2 different user groups. Regardless of age or background, they do all have the same purpose; travel with the train. The User-Centered Design (UCD) methodology has been applied in developing processes of this research.

Keywords: User-Centered Design (UCD), Human Computer Interaction (HCI), Interactive, Map.

1 Problem Statement and Motivation

The BTS Sky train is the comfortable and convenient way to get around Bangkok. In service since 1999, it has transformed the face of public transportation in the Thai capital, for the first time offering both residents and visitors a comfortable ride through central Bangkok. The Bangkok's second major mass transit system (MRT - Mass Rapid Transit: underground train) is on track for a test run on 2004 that reaches from the Northern train station of Bangsue to Hua Lumphong main railway station in a loop, connecting with the BTS on 3 different stations, namely: Silom, Sukhumvit, and Chatuchak Park[1]. MRT have a similar ticket system as BTS but both systems are for the time being not compatible. All stations of MRT can be reached via ramps and elevators, if needed, or alternatively via escalators.

[1] www.bts.com

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 70–79, 2010.

The finalized system of this research is web based version of MRT and BTS interactive route navigation map. The three main problems for the tourists/residents, who travel around Bangkok, are:

- The first problem is that MRT and BTS systems and services are still unknown to the tourists/residents; because the complete information of underground and sky train maps are not provided on the both MRT and BTS official web site. But there is some information that provided in web site of private organization for example, http://bangkok.sawadee.com/skytrain.htm, www.2bangkok.com, http://bangkok. sawadee.com/mrta.htm. As we can see information of MRT and BTS routes separately in Fig. 1 and 2.
- The second problem is that tourists/residents do not know the necessary details in order to using these transportation; such as, if where the train stations are, and how they connected along the train route or, how both BTS and MRT connected, How much does it cost for traveling from one station to another?. If we look at many interactive systems we can see that they have been designed by computer programmers and software engineers on the false assumption that everyone else in the world understands (or should understand) computers as well as they do. How can we make sure that interactive systems are better designed? [1]. Most of the problems we address in this course fall under the topic "how can we design it better?"
- The third problem is concerning the developing processes, the designers of the information system are actually very intelligent. Most designers of the system are computer experts and programmers: people familiar with computer systems, how they work and not really intimidated by them. In the mistaken that they are typical members of the public [1].

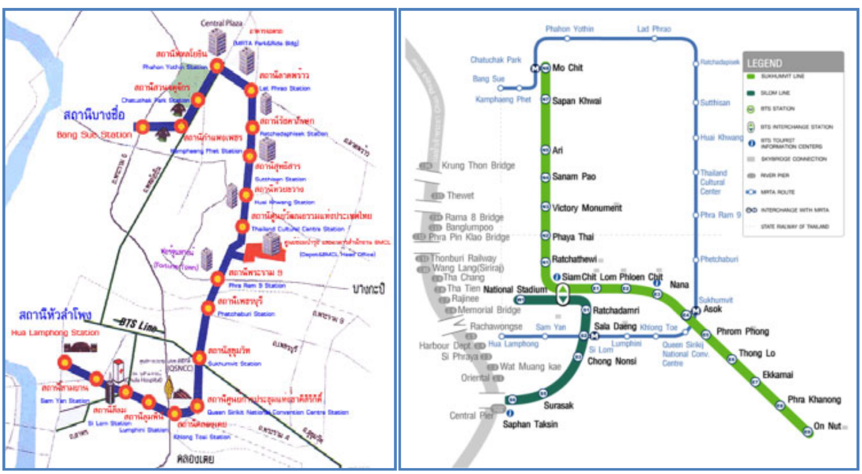


Fig. 1. MRT underground route map[2] (left) and BTS sky train route map[2](right)

---

[2] http://bangkok.sawadee.com

2 Research Objective Statements

"Getting It Right the First Time"...is the most desirable thing that every developer wants. This research intends to apply the user-centered design methodology to create the interactive route navigation map. So the primary problem specification is to refer to any technological system which requires interaction with user, such as how can we design an interactive MRT and BTS map by integrating the user design center method, which is both convenient and informative?. Research Objectives is to help identify the primary needs the users will have off the train map, we have chosen to split the goals into usability goals, user experience goals, and, publication goals; as suggested [2].

- Usability Objectives: Because we assume a wide variety of users, it is important that the system is easy to use.
- User experience Objectives: First and foremost we wish the user to have a nice and easy experience using the public online interactive map. We try to imagine; *a user on the station who looks at the interactive route pop up from inactive background and sees the clear train route*. Maybe he can see the reality location of all the train stations. The users know at any given time where their closest station is located, our goal of making a service that is informative has been met.
- Publication Objectives: As the main industry in Thailand is tourism [3]; this research can be benefited to the tourists as well as residents. This work can provide the main transportation gateway of BTS and MRT in Bangkok. The first time visitor can use this web-based interactive map to see and learn how to travel around Bangkok using MRT and BTS.

3 Interactive Manners and User-Centered Design (UCD)

Interaction design aims to reduce the learning curve and to increase accuracy and efficiency of a task without diminishing usefulness. The objective is to reduce frustration and increase user productivity and satisfaction. Interaction design tries to understand certain users' requirements and then designing to gather and go above them by observing who wants to use it, and how they would like to use it. Relating users, designers gain the ability to better understand user goals and experiences. There are also positive side effects which include enhanced system capability awareness and user ownership or sometimes called participatory design. To enhance overall approval of the final product by adding a sense of ownership, because users who have been dynamic contributors in a system's development are more expected to feel that they are the owner of the product. The powerful, interactive systems will be relevant to its intended users, will support the tasks for which it is intended and make good use of technology. According to Ben Shneiderman, "The old computing is about what computers can do, the new computing is what users can do." [1]. Then any concern of the present and future of computing was restricted. Only especially skilled people could work with them. Ben Shneiderman also said in foreword (Sears and Jacko, 2003) that there are three essential topics to address in Human Computer Interaction (HCI). They are usable, universal, and useful. It is vitally important for system designers and HCI practitioners to distinguish between usefulness, usability and accessibility (universal) if they are to be as effective as possible in their work [1]: *Useful* means that the system supports user objectives, *Usable* means that it is supports these objectives in easy-to-use ways, *Accessible* means that it can be used by the full range of intended users.

In order to accomplish these concepts, the UCD is the answer of how we come up with the system that can perform these essential concepts. Combining with a new methodology which is prototyping, the final result of prototyping studies would be directly convertible into efficient application code. On the usability evaluation side, the scenarios and prototypes would be readily converted into a simulation that would support human studies of effectiveness of the resultant designs [4]. The phrase interactive design refers to both the process and the results of design. User-centered design is itself an interactive process. If you believe in creating interaction between the system and the user, then it makes sense to employ the same concept of interaction between when the user is, in this case, the designer [4]. UCD was developed at the University of California at San Diego; the key features include [1]: 1. A central focus on the people who will use the systems, on their preferences and requirements, 2. Building simple models of the users, the tasks and the technological systems, 3. An iterative process and 4. Prototyping and the evaluation of alternatives by users."

The designer cannot passively go through a number of stages in linear manner, doing each stage once before signing it if and moving on like waterfall model. In UCD, by contrast, the designer and the design process must both work interactively. The emphases are design and evaluation as an interactive cycle. Designers cannot become stuck with their initial ideas and assumption, but must respond to data about what works and what does not. UCD forces designers to be explicit about decisions made, reviewing them through prototype and storyboards with the team and potential users. Designers need ways of sharing design thinking by producing physical representations of design ideas. These could be working computer-bases prototypes or as simple as a paper sketch [1]. It means that the design team and potential users can share in the process, evaluating the design ideas and suggesting improvements.

There is certainly the current high level of interest in multimedia interaction parts from up to date technical advance. While guidelines and research findings exist, so do large and significant gaps in knowledge. This is due to the inherent problem of predicting how system features will succeed or fail in realistic task setting. Moreover, iterative design and testing method are necessary [5]. One of the whole suggestive processes of UCD starts from Plan → Research → Design → Adapt → Measure [6]. UCD is an approach to design that grounds the process in information about the people who will use the product. UCD processes focus on users through the planning, design and development of a product [7]. There is an international standard that is the basis for many UCD methodologies. This standard (ISO 13407: Human-centered design process) [8] defines a general process for including human-centered activities throughout a development life-cycle. ISO 13407 provides guidance on achieving quality in use by incorporating user- centered design activities throughout the life cycle of interactive computer-based systems. The sequence in which these are performed and the level of effort and detail that is appropriate varies depending on the design environment and the stage of the design process [8].

4 Case Study Walk through: The Development Processes

In less well-defined projects, many designers have found real scenarios helpful to characterize that happen when users perform classic tasks. During the early design stages, data about current performance should be collected to provide a baseline. UCD is a flexible process for software development projects that enables teams to more effectively meet the needs of users and customers. In this version, the UCD activities are broken down into simple four phases: ***Analysis, Design, Implementation and Deployment***, with suggested activities for each phase. They are as following:

4.1 Analysis and Design

4.1.1 Finding the Objectives of the Research

It should focus on usable, universal, and useful.

4.1.2 Developers Brainstorming

- Developers who have experience in coding for 5- 7 years, got together to discuss the experience on using the similar accessible system. One of the developers did not have any experience in using or creating the interactive map at all. Consequently they discussed how the system should work in their own perspective.
- The first problem occurred in this phase is that the requirements are not clear to developers. And as we set the methodology as UCD, then they must turn to users in order to get what users really need. They all have their own area of expertise such as some are very good in design, or coding, or graphic design.
- Learning from the other available systems, developers have the good design from these three accessible train interactive systems which are the Paris interactive map[3], the Seoul Metropolitan Rapid Transit interactive map[4], and the Tokyo interactive map[5] as see in the Fig. 2.

Regardless of the method we choose to identify the appropriate tasks and requirements for your Web site, it is important that we conduct these activities before implementing any new design or major redesign of a site. This critical first step in the Web site design process will ensure that final design is user-friendly, flexible and adaptable, and will eliminate expensive and unnecessarily frequent redesigns.

Considering available interactive Thai interactive train map, they cannot show the whole route pop up from the screen and there is no information about the price from beginning to destination stations which are selected by the users. Also there is no information about connection of BTS and MRT route. Nevertheless the advantage of these existing interactive map Web sites is that they can show the surrounding area/tourism spots of each train stations.

4.1.3 Finding What the Users Need

Introduction an integral part of an effective user-centered design (UCD) by IBM; process is the gathering of requirements and the identification of common user tasks.

[3] http://www.ratp.info/orienter/cv/cv_en/carteparis.php
[4] http://www.smrt.co.kr/Train/Subwaymap/Eng/Subwaymap.jsp
[5] http://www.japaneselifestyle.com.au/tokyo/tokyomap.htm

This part of the UCD process is just as important in Web site design and it is in software design. Interview user and focus group techniques had been applied in this stage. User questionnaires are familiar, inexpensive and generally acceptable companion for usability tests and expert reviews. The keys to successful questionnaires are clear goals in advance and development of focused items that help to attain those goals. Experienced questionnaire designer knows that care is also needed during design, administration, and data analysis [9].

Fig. 2. Korea[6], Paris[7], and Tokyo[8] interactive map

As the users cannot imagine how the system work. Developers have to draw the few samples of system screen interface in order to show to the users. And then they discuss and choose the most preferable design or function that they want to be included in the screen. The reason that why we have to interview users before creating the prototype because we do not want to narrow the users' idea. This process might take very long time, because the users might not have any clue of what the system will look like, but we can discover what the important functions of the system is, and what kind of information the user needs. 10 open questions had been selected by developers and were distributed the 2 groups of the 40 users (age between 19 – 44 years old). The first group is the users which are unfamiliar or little familiar to the BTS and MRT transportation; they do not often travel or never travel by BTS or MRT. The other group is the users which travel very frequently by BTS and MRT.

The results from the questionnaire conclusion (Fig. 3) show that the high similarity requirements are 1. Route distinction (30% and 29 %), 2. Price (18% and 17%), 3. Travel time (11% and 17%), 4. Station name (16% and 10%), 5. Connection point of BTS and MRT (11% and 6%). Other suggestion topics are; color of each route, printable, zoom (in-out), bold font, bi-language (Thai-English).

4.2 Implementation and Deployment

After getting the system requirements; the developers pull out the similar users' needs from both 2 groups of users and make it dominant on prototype screens. The problem found in this stage is the coding problem. The first coding style, we applied PHP script to generate the database (it works as server-site script), the result system

[6] http://www.smrt.co.kr/Train/Subwaymap/Eng/Subwaymap.jsp
[7] http://www.ratp.info/orienter/cv/cv_en/carteparis.php
[8] http://www.ratp.info/orienter/cv/cv_en/carteparis.php

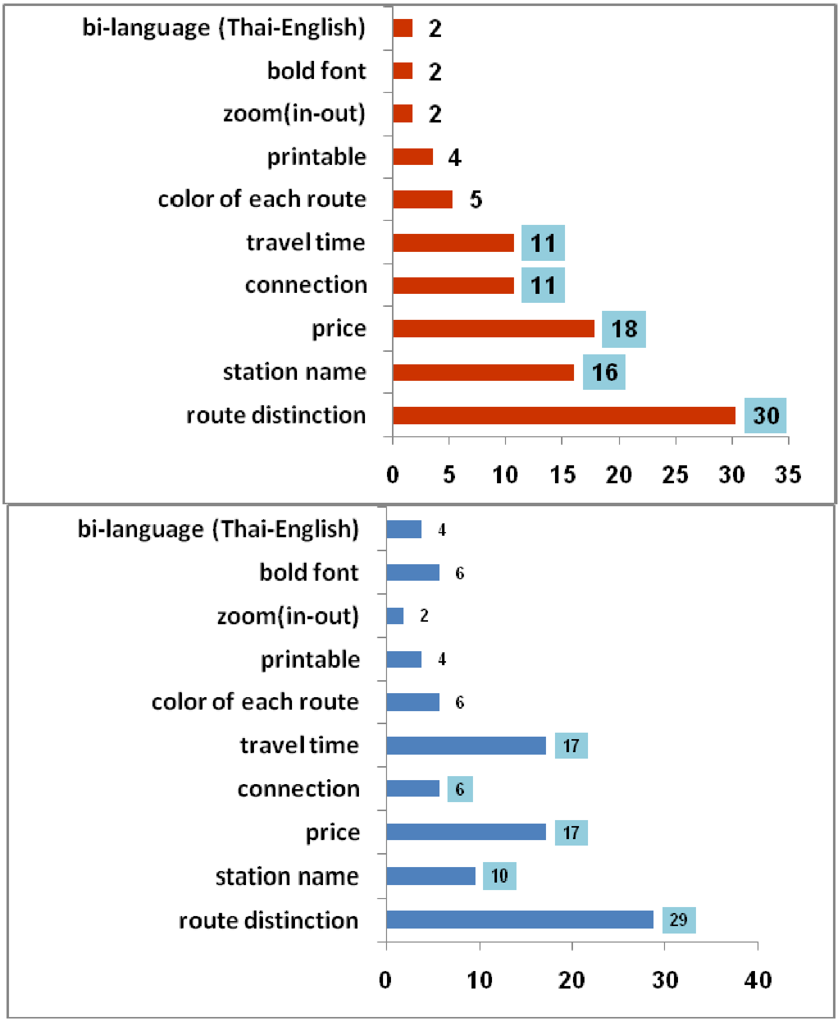


Fig. 3. Percentage comparison graph of suggestion topics between 2 groups of users: *group1*: unfamiliar user (above) *group2*: familiar user (below)

did not impress users, because it took a long time to show the train's route which user selected, even though the database is not so large. The developers decided to change-coding style from PHP to array instead. There are pros and cons here: Array style of coding took less time to process and made the system more efficient than PHP because it works on client site, regardless of the operating system, platform, speed of internet connection. However it might make difficulty in the future for developers to edit the code of the program.

Fig. 4. The first interactive map which is not been selected (too much information on one page) (left) the first prototype screen with white background (right)

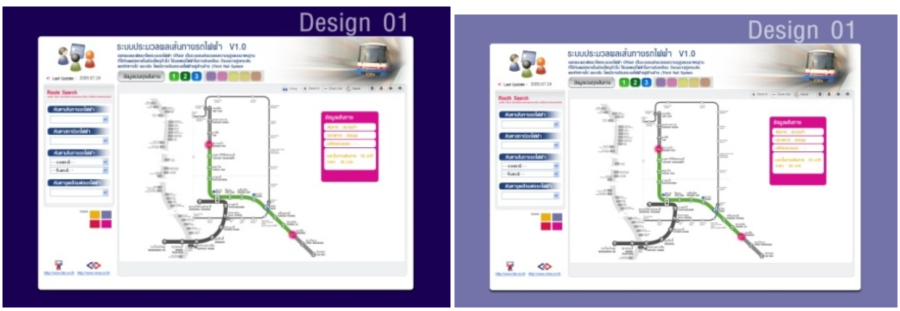


Fig. 5. The in-process screen design

The first draft design is based on the existing Bangkok interactive map, which contain too much information (Fig. 4 left). Accordingly, the developers considered the information. From the questionnaire and interview, the users do not need so many details, because they need simply 9-10 topics of content on the screen. They determined drop this design (Fig. 4 left). The next design is white background with the main train route distinction (Fig. 4 right), the users do not like the simple color, and they found them unattractive (Fig. 5 left). The developers show them to user then the user want the more colorful screen (primary color are yellow, red, and blue[9]) and then we select the blue color theme.[10] Hence the next version draw on the dark blue background with a same blue shade of menu's color with the white background with the same train route distinction as same as the first design (Fig. 5 right). The users prefer this design but they want to make change to cooler tone of background color. And developers totally agree on this suggestion; the reason is that the interactive intend to provide the content in text and the user usually take time to look at the information

[9] http://www.creativepro.com/article/learning-to-use-color-on-your-web-site
[10] http://www.ziva-vatra.com/mirrors/nemesis.lonestar.org/reference/internet/web/color/charts/blue_01.html

concern the route of the trains, as a result; in this case the user need comfort tone of color to their eyes for their eyes.

The last suggestion made by users is to construct the information universal use, which is to make web site available in both Thai and English language as see the last adjustment in Fig. 6 below.

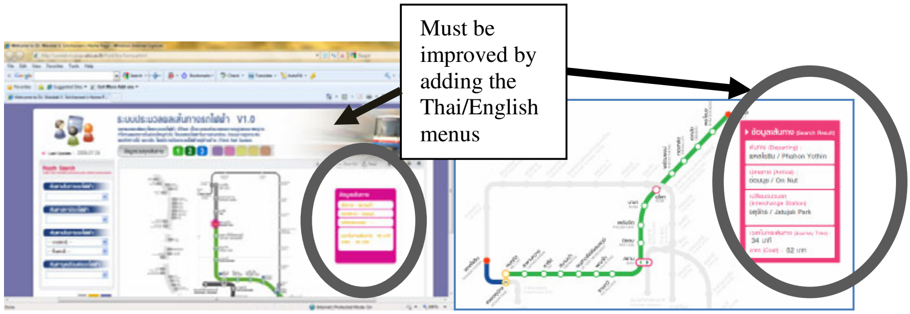


Fig. 6. The final web based interactive map

5 Conclusions Recommendations

User-centered design is not just about the user interface and the screen layout, colors, and buttons. It is much more, to come to a good user interface design at the end, it is essential to lay the foundation in the beginning. This foundation is an understanding of what end users really need and want “the user requirement”. As already stated in the name of the process, "User-Centered" Design, the end-user should be involved in every step of software development processes and increase the ownership of the product. This investment of involving end users and customers early in the development process ensures that software products are produced that are demanded and accepted by users at the end. The recommendation further works are the links to tourist spots nearby station or link to the official web of BTS and MRT. There is the intention of the researcher to give this work to MRT and BTS and to make it available via their official web site. Now the final version is available at http://utcc2.utcc.ac.th/public_online/(Fig. 7.). Clearly after the last evaluation from the users, they say that the research work is very save time and money for the residents and tourists in order to prepare the trips before go to the train stations. The bottom line is UCD is the methodology which let the user get involve with the developing processes since the beginning of the project; there are advantage and disadvantage here. From the case study; the advantage are: increasing the acceptance of the product because of participatory and ownership in products, getting the real requirement from the users, and usable products. The disadvantages are: time consuming and costly because of iterative design, change of requirements, and argument between developers and users.

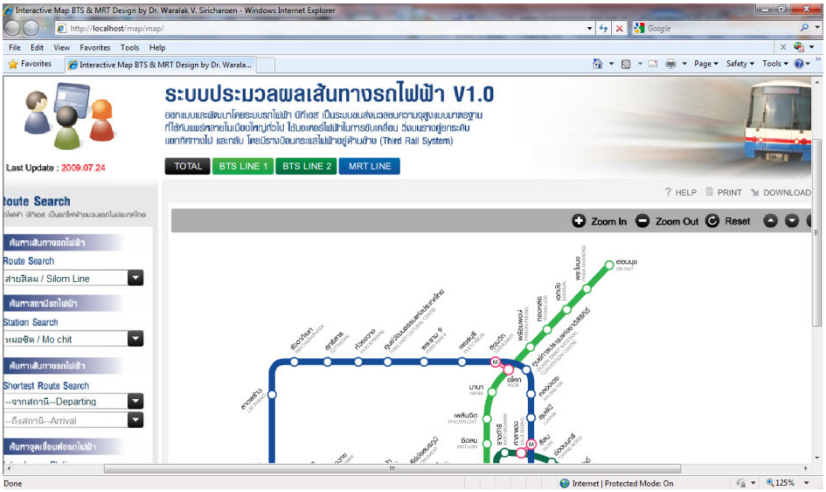


Fig. 7. The final web based interactive map

References

1. Smith-Atakan, S.: Human-Computer Interaction. In: Thomson Learning, ch. 1, 2, London, England (2006)
2. Sharp, H., Rogers, Y.: Preece J. Interaction Design: Beyond Human-Computer Interaction, 2nd edn. (2007), ISBN: 978-0-470-01866-8
3. iexplore.com. Where to go in Thailand (2009), http://www.iexplore.com/dmap/Thailand/Where+to+Go
4. Pew, R.W.: Evolution of Human-Computer Interaction: from Memex to Bluetooth and Beyond. In: The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, p. 15 (2003)
5. Waterworth, J.A.: Multimedia interaction with Computers: Human Factor Issues. Ellis Horwood Series in Information Technology, West Sussex, England (1992)
6. Weissenberger, U., Thompson, C.F.: User-Centered Design SAP User Experience
7. Usability Professionals' Association: what is User-Centered Design? (2009), http://www.upassoc.org/usability_resources/about_usability/what_is_ucd.html
8. UsabilityNet : ISO 13407, Human centered design processes for interactive systems (2006), http://www.usabilitynet.org/tools/13407stds.htm
9. Shneiderman, B., Plaisant, C.: Designing the user interface: strategies for effective human-computer interaction, 5th edn. Addison-Wesley, Boston (2010)

Development of Wearable Device by Kid's Friendly Design for Kid's Safety

SeungHee Lee[1], Jahee Sohn[1], Atsushi Usami[2], and Masatoshi Hamanaka[2]

[1] Graduate School of Comprehensive Human Sciences, University of Tsukuba
{lee,jahee80}@kansei.tsukuba.ac.jp
[2] Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Japan
{usami,hamanaka}@iit.tsukuba.ac.jp

Abstract. In this study, we develop a wearable device for kids under 6 years who are growing up so fast physically enough to face various experiences but low ability to describe the events by language exactly what they have experienced at nursery or kindergartens to share with their parents. This system will be linked with local safety network to let somebody react real time when a kid faced to any inexperienced events. We adapt biological information such as heart rates, physical information such as body movements, GPS and camera on the device. The data from the device could be shared to parents or teachers in the kindergarten afterwards or real time. We focus on designing the device has fascinated form giving, ease to wear and symbolic indication of "Protected" by wearing on. To keep kids enjoy wearing the device as like a pendant. In this paper, we introduce the user oriented development of the device which has kid's friendly design and useful function.

Keywords: Kid's friendly design, areal network, shared information.

1 Introduction

In increasing crimes or accidents on kids, parents need to consider the way to protect them and predict what are going to happen to them to know their habits of ordinary behaviors or related people around them. In Japan, it is now very general to protect the children from the crimes, most of the primary school children put security alarm buzzer on their school bags. Some schools hand it out to the students as like textbooks. But many of the children do not give attention to it and no useful function except to make noses surround. No more security functions with the alarm buzzer but a simple toy.

In many cities, increasing working mothers' kids are being taken care by nurseries or kindergartens on their working days. They spend less than 3-4 hours with their kids before and after working per day. And under 6 year kids they are growing faster in physically but sometimes cannot describe their experiences by words to their parents even though those were scary, happy, surprise or sad.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 80–85, 2010.

But the responsibility of kids' security is on their own parents wherever, whenever or whatever. In this system, we construct a system which can help the parents can confirm their kids' behavior or experiences through a small and charm wearable device connected to social network with local support. This device can detect heart rates and behavioral movements of kids using heart rates detector and three axis accelerometers. It also has a camera which turns on only when the kid got a big change of heart rates or movements such as falling down or jumping down. And the visual information on the camera can be browsed at home through a secured program.

2 Development of Wearable Device for Kids

The device is developed by reversed order compare to ordinary development process, such as, designing first and put the advanced technology in a box next. The reason why designing first than the technology is, to make kids wear the device as their own willing being fascinating by attractive form giving and ease to wear. We call it, 'Kids Friendly Design'.

For 'Kid's Friendly Design', we first concern the weight and size of the device, and secondly about how and where to wear on kid's body. Concerning preferred graphical design for kids would be the last step.

2.1 Weight and Size of the Device

To fix the size of the device, we surveyed Kids' body scale based on Research institute of Human Engineering for quality life in Japan. They've reported all the size of average body scale of ages from 0 to 12 from Japanese kids since 2005 every year. The chest size of 3 year old kids is 164mm and 5 year is 176mm. From the chest size, the device should not be exceeded 100mm in any direction.

About the weight, we surveyed the toys which can be attached on kid's neck, most of them are between 30 to 50 grams and which can be attached on their waist, are less than 100 grams. And we also surveyed about the weight to mothers in the kindergarten, 75 percent of them answered under 100 grams can be applied to their kids. For references, the alarm buzzer has 45~60 grams and kids mobile phones have 130~150 grams.

2.2 How and Where to Wear the Device

To attach on an appropriate position on kids' body, we should concern that it is not to bother kids' active movements. And at the same time, the camera on the device should be keep positioning well focused even they move around. Here we show examples of A and B on the figure 1 and 2.

2.3 Design Development of the Wearable Device

a. **Necklace Type**
   Based on example A, we suggest necklace type of ideas and mock-ups (Fig.3).

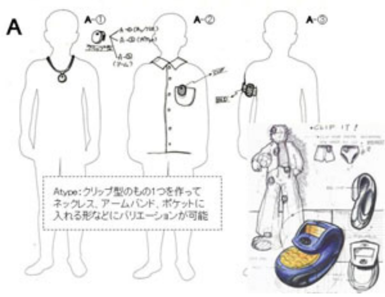


Fig. 1. Example A (Clip type)

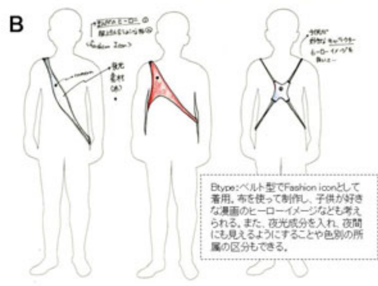


Fig. 2. Example B (Suspender type)

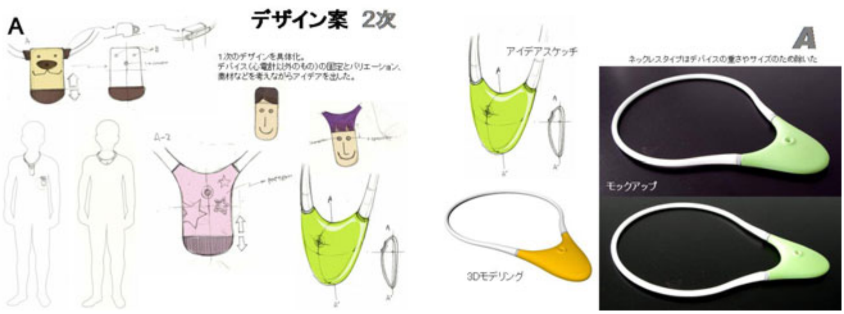


Fig. 3. Necklace Type Design

We build up the mock-ups by 3D Printer (Dimension 3D Printer BST). This necklace type is not able to fix the camera focused and have a possibility to be fastened on kids' neck by accident.

b. Shoulder Sac Type

Based on example B on the figure 2, we suggested four kinds of ideas and mock-ups (Fig. 4). This shoulder sac type is more stable to put a camera on the device and have a diversity of positioning the sensors. Those are well attached on the body, so children feel lighter than a real weight. B-1 could be made with fabrics or covered with soft rubbers. It fits on the body but the camera position could be lower than other types. A kid's chest is high as grownups' knee. If the camera placed lower position on the device, we need to do to control the direction of lens. B-3 is updated idea of B-1. A support belt was added on the other side to improve the stability. B-2 has less space attached on the body than B-3. It has three corners on the shape and each corner connected with belts. And B-4 is applied idea of B-2. It has four corners to spread sensible weight better than others. Figure 4 shows the various shape of device design on kids around age of 4-5 body sizes.

2.4 Build-in Mock-Ups

From the shoulder sac types of design, we finally build the sensors based on the idea B-3. The total weight of the sensors which should be built in, would be at least 90 grams, the necklace type was rejected from the candidates, because it will be a big burden on their neck. Among four ideas of shoulder types, it offers relatively high position of camera and merit of ease to wear.

In detail, to decide the size of device, we made updated sketch of B-3(Fig.6) and made a 3D mock-up by 3D printer. To put the sensors safely in a box, the layout was decided after trials. On figure 7, solving the way to fix on the body, small connection parts were created on the corners. We named it 'Omusbi' (Fig.10), meaning of 'Rice ball' or 'Connection' in Japanese. 'Rice ball' is traditional snack for kids and any ages of people in Japan. 'Connection' is also meaningful function with this study to share information not only between kids and parents but also with local supports.

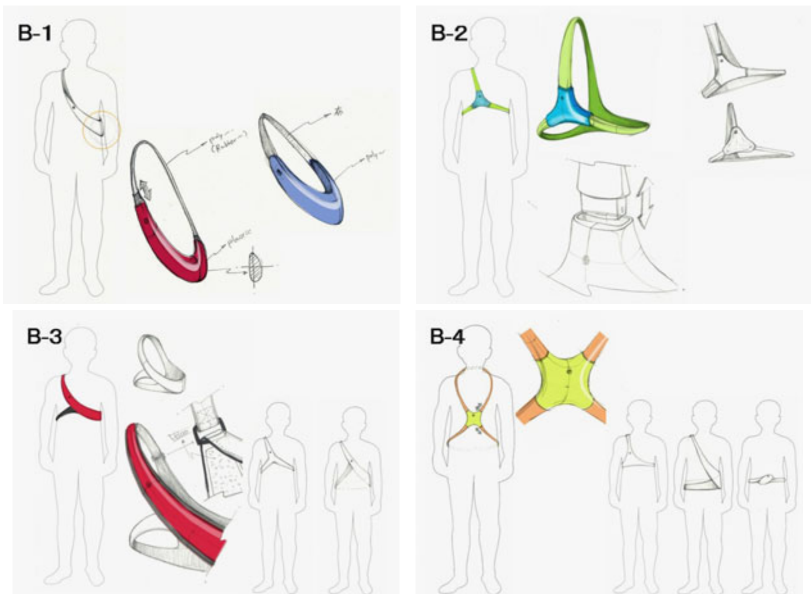


Fig. 4. Development of various shape of wearable device design

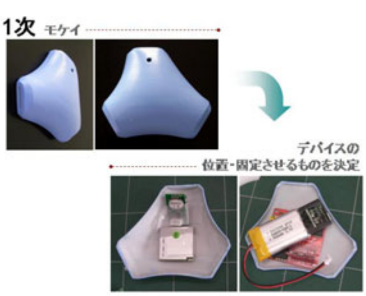


Fig. 5. Built-in Device

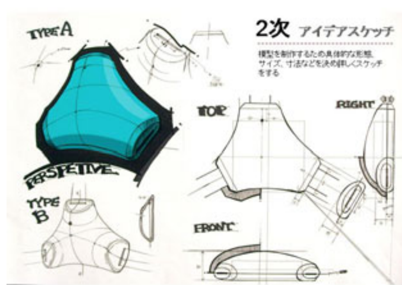


Fig. 6. Updated Sketch of B-3

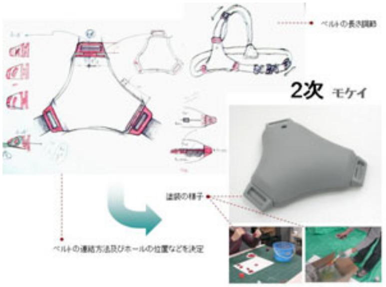


Fig.7. Design Sketches for final mock-up **Fig. 8.** Prototype of Device(Omusbi)

The on-off switch is on the back side in a tiny hole to turn with an originally designed stick only by their parents to prevent to be adjusted by somebody else.

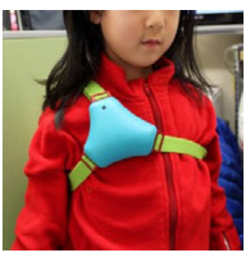
3 Discussion

3.1 Weight and Size of the Device

The goal of designing the devices in this paper, less than 100 grams and 100 mm in every direction for kids between ages of 2-6 was achieved. The model having three corners concluded as the final design for kids in this study. We adapted the sensors, such as heart rates detector, GPS, three axis accelerometers, camera and micro computer in the device and we still have a possibility that it could be lighter if the sensors we adapt will be lighter and smaller. The final prototype has 94 gram with out the belt.

3.2 Position of Wearing the Device

To take heart rates, kids will attach a set of wireless detectors on their left side of chest directly. Then the device should be positioned on the right side of chest on outside. The 'Omusbi' device has three corners to fix on the chest so, it does not bother to get signals each other. Figure 9 shows how kids wear the device.



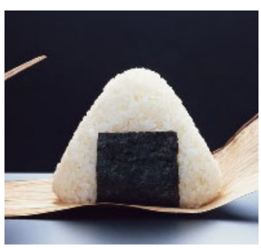


Fig. 9. Device attached On Kid's body **Fig. 10.** Omusbi

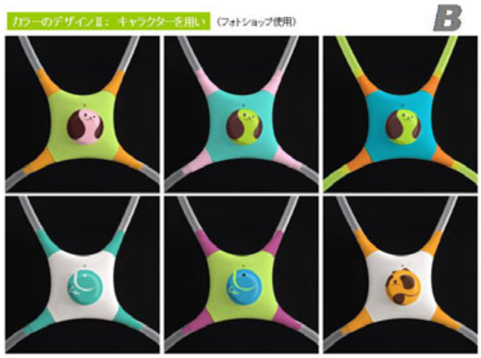


Fig. 11. Graphical package examples

3.3 Graphical Package for Kid's Friendly Design

Graphical package of device will be also very important for kids to make them want to wear the device by themselves. It will be also a symbolic indication of "Under Protected" when they wear the device. With friendly animal graphics or originally created hero characters would be preferred by kids. Figure 11 show the examples of graphic packages on the device.

4 Future Study

This study is under going to be linked with local safety network and usability test by kids in the kindergarten will be continued in the next stage. The design of Omusbi will be upgraded to be lighter and more fancy at anytime when the better sensors are developed. For user friendly design, various types of devices will be offered to parents. They can choose the functions on and off among the sensors to depends on their kids' condition. Each of them will have different experiences and different feeling on the events around them. The device will offer scenes of kids' daily experiences only when they were stimulated by unexpected events, to the network by pictures when parents access on secure system. And also staffs of kindergarten or nursery can also access when the kids got some changes on their physical conditions.

Acknowledgement

This study is a consignment project with SCOPE (**S**trategic Information and **Co**mmunications R&D **P**romotion Program) of Ministry of Internal affairs and Communications in Japan.

References

1. Farringdon, J., Moore, A.J., Tilbury, N., Church, J., Biemond, P.D.: Wearable Sensor Badge and Sensor Jacket for Context Awareness. In: ISWC, p. 107 (1999)
2. Kern, N., Shiele, B., Holger, J., Lukowicz, P., Troster, G.: Wearable sensing to annotate meeting recordings. In: Euro-Par 2006. LNCS, vol. 7(5), pp. 263–274. Springer, Heidelberg (2006)
3. Wiklund, M.: Usability in practice: how companies develop user-friendly products. Academic Press Professional, Inc., London (1994)

Towards a Usability Coaching Method for Institutionalizing Usability in Organisations

Åsa Cajander[1], Elina Eriksson[2], and Jan Gulliksen[2]

[1] Uppsala university, Dept. Of IT/HCI, PO Box 337,
75105 Uppsala, Sweden
[2] KTH (Royal Institute of Technology)
100 44 Stockholm, Sweden
Asa.Cajander@it.uu.se, elina@kth.se, gulliksen@csc.kth.se

Abstract. The ideas related to user centred systems design are difficult to implement in organisations and usability is given low priority in in-house systems development. This problem is multifaceted and complex and needs to be addressed with a method capable of adapting to situations, people and context. In this paper we outline a new method – usability coaching – that has the capability for dealing with the situated problems of introducing user centred systems development in an organisations. The method is the results of a larger action research case study in which 9 individuals in an organization received usability coaching. Our study indicates that the coaching program made coachees work more actively with usability activities; hence the coaching program had a substantial effect on their actions and contributed to the organizational change.

Keywords: Coaching, user centred-systems design, organizational change, usability, action research, qualitative research, learning theories, conceptual change, and threshold concept.

1 Introduction

"Usability and user centred systems design is like peace in the world. Everyone says that it is truly important, but no one really understands how they can contribute to it, and no one takes responsibility for it". This colorful quote from an experienced project manager in a large government organisation captures the difficulties achieving user centered systems design (UCSD), [1]. Our research group has extensive experiences from action research projects in close cooperation with large government organizations with the goal of contributing to a good work environment by promoting UCSD [2]. Previous research has shown that it is not sufficient to work with the systems development department in an organization – attitudes on all levels in the organization need to change [1, 2].

The following story about an IT manager illustrates how we have addressed this issue trough a new usability method that we call "Usability Coaching" and we will subsequently analyze different aspects of this story: John is a 50-year-old very experienced top level IT manager in a government organisation. He had prior to the usability coaching not paid much attention to UCSD aspects in development projects. However, he was very dedicated to the

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 86–97, 2010.

coaching program, where he took every opportunity to discuss usability and usability issues in relation to his daily work and his responsibilities.

During one of the first coaching sessions there was a discussion concerning the waterfall model in relation to iterative design where John saw no reason to work iteratively in their systems development projects. Among other things, he expressed that "Iterative design is relevant in companies that work with new product development. But business development through IT requires a waterfall model." We had an animated discussion, and John expressed a variety of reasons for not working iteratively. Many, but not all, of these had been voiced by him in an earlier workshop. A few weeks later we were invited to a workshop on usability work in the organisation. John presented a new software development model as a speaker at the workshop. It turned out that the new model he presented was iterative, and included several usability activities.

Our previous research has focused on understanding how to integrate usability in design and especially how basic values and business values affect this integration [3]. Our research indicates that strong business values such as automation, efficiency and customer satisfaction shape the development of new computer systems and ultimately work practices. Moreover, our previous research on manager's and their perspective on usability in the same organization has shown that even though formal usability training has been provided, and despite the organisations official focus on usability – managers in the organisation had only vague ideas about what usability is and their responsibility for integrating usability [4]. Most managers agreed that usability was something important in the organization, but they still expressed that they had limited responsibility for usability. Hence one can conclude that most of the managers agreed that usability is important, but that it is someone else's responsibility. Moreover, new methods are needed to impact values and the interpretation of what systems development is about, as well as what goals to strive for in an organisation. The discussion and interest in values and perspectives in systems development is not new in HCI, and the conflict between different perspectives has been extensively discussed, see for example [5-10]. Our research group has worked with informal coaching of professionals to impact values and promoting usability issues. This paper concludes these experiences into an outline of a more formal method for usability coaching. As a basis for this method one of the authors of the paper was enrolled as an external usability coach in a public authority with the dual aims of developing the usability coaching method and to understand how such a method affect the coachees and their actions to promote UCSD in the organisation. The experiences from this case study are discussed in relation to relevant learning theories.

2 Methodology and Method

The overarching methodology of this study is action research. This approach is *"unique in the way it associates research and practice, so research informs practice and practice informs research synergistically"* [11]. Hence, action research has dual aims and is intended to produce both action (change) and research (understanding) [12]. Action research is a participatory research methodology where researchers and practitioners work closely together. The research in this study has been planned and conducted together with participants from the organisation, with a project group within the organisation as the main collaboration partners. The research findings described in the case study are based primarily on participatory observations both during the coaching sessions and outside when working together in the project. One of the authors has been the coach in the usability coaching program. During the coaching sessions a handwritten diary was used by the researcher, with focus on the problems discussed. In between the coaching sessions a computer based research diary was used which contained more of a summary of the conversation including didactic and pedagogic notes on how the coaching sessions could have been improved to create a better learning environment. Moreover observations of behaviour by other researchers participating in the research project were also written down in the research diary. Furthermore an interview study was made as an evaluation of the effects of the entire action research project in the organisation. The interview study encompassed 36 semi-structured interviews and is described in more detail in [13]. However, in this article, only those interviews with relevance to the coaching sessions, that is the nine coachees, were analyzed. All interviews were audio recorded and notes were taken, which both were used in the analysis.

Analysis of data from the case study has been done through mind map sessions, where the research diary as well as other written material has supported memory recall. The interviews have been analyzed in two steps, first an analysis where general findings were reported in [13] and then a second analysis with particular focus on the coaching sessions. The notes taken from the interviews were used in order to see emerging themes and where necessary the recorded interviews have been carefully listened through a second time. All quotes in this paper are translated from Swedish and in some cases altered for readability reasons; furthermore, all names have been altered. Halfway through the usability coaching program some preliminary findings was presented and discussed at an INTERACT 2007 workshop [14]. The comments and reviews of that paper have contributed to the subsequent coaching work and also to this paper.

3 Research on Coaching

Coaching and mentoring have a very long history and some say that it origins in ancient Greece with links to the Socratic dialogue [15]. Recent studies on the discourses related to mentoring and coaching concludes that the meanings of the words have subtly altered over time and that they have become more or less interchangeable in the modern usage of the word [15]. The purpose of coaching historically has been to help a younger and less experienced person to master a profession or situation through discussions and reflections with a more experienced and older person. This still remains one of the core purposes of many coaching programs today. Despite the lack of a common (well established) definition [16], most researchers agree that the word coach generally indicates a person acting in a supportive relationship as mentor, teacher, leader or role model of some kind. There are some articles on coaching in the HCI area (see for example [17], but it seems to be rather unexplored. The article above describes mentors in en experimental course in HCI. There is however a large body of research on coaching in other research areas such as education, gender research, management theory and organisational theory. An extensive literature review made in 2003 [19] has studied mentoring research from many perspectives. Among other things they have looked at studies of mentoring outcomes, and conclude that most studies have focused on outcomes for protégés, rather than on outcomes for coaches or organisations. However, Zey has discussed possible organizational outcomes and identified seven outcomes of mentoring. These organizational outcomes include employee integration, reduction of turnover, organizational communication, management development, managerial succession, productivity and socialization to power. Moreover, there has been quite a few mentoring studies related to how personality traits are related to behaviours in a mentoring relationship [19]. Some of the hypothesis examined are if the drive to excel will lead those with type A personality to have a mentoring relationship [20], or if people with an internal locus of control are more likely to perceive that they can improve their skills [21]. Furthermore, there has been research on how mentoring relationships evolve over time [22] and how the coach and the coachee interact and influence each other. However, despite the fact that mentoring involves intense and interpersonal relationship, research on the how behaviour, perceptions and affect is limited [19].

4 Learning in a Coaching Situation

Usability coaching is based on constructivism as the underlying epistemology, i.e. there is no objective truth and that meaning is constructed in interaction between human beings and the world. This perspective implies that the usability coaching program did not work to establish pre-determined ideas where procedures are well defined for situations, or where there is a “right” or objective answer. However, the discussions have been inspired by the values and perspectives of UCSD. Moreover, we agree with Schön [23] that problems can be of different natures and that many problems that occur when trying to introduce new ideas and values such as UCSD in systems development are complex and multifaceted. This is also discussed in general by for example [24] who call these problems “wicked problems”. Schön describes these as belonging to the swampy lowland where predefined methods and techniques are of no use when trying to solve the problem. Schön takes this one step further and elaborates on how to educate people in addressing these complex and swampy problems to become reflective practitioners. The usability coaching method can help coachees to reflect on the particular problems of their swampy lowland. Hence the usability coaching we in this paper refers to the learning that Schön describes.

5 Case Study Setting and the Usability Coaching Program

The usability coaching was conducted within a large action research project together with the Swedish public authority that handles financial aids for students, CSN. The computer support are mainly developed in-house at headquarters, at the IT-division. The action research project were initiated at the Human Resource department [25]. The goal of the action research project was to increase knowledge about usability and a good computerized work environment.

The usability coaching program was initiated by the project manager at the Human Resource Department and one of the authors of this paper acted as a usability coach.

The coach was externally financed coach, and the coaching was on a voluntary basis as the coachee had the possibility not to participate in the program. Unlike many coaching programs that have personal competence and career as the focus – the usability coaching program had the aim of institutionalizing usability and UCSD through personal knowledge in the organisation. Hence this usability coaching program aimed at organisational change, and not merely personal knowledge. This usability coaching program was aligned with the long-term objectives and strategic positioning of the organisation as it aims at being the most user-centred authority in Sweden and usability is a part of the business goals of the authority. Another major difference from many other coaching programs is the relation between the coach and the coachee. The coach was only more experienced, being an expert, in one aspect of the profession, i.e. usability. For example, in this usability coaching program the coach is no expert of IT strategies and the strategic work in an authority which was the responsibility of one of the coachees – rather the coach and the coachee create a new understanding of IT strategy work and usability based on their previous experiences and knowledge. Nine key persons representing the main stakeholders in the systems development process were appointed by the organisation, and they had formal titles such as “IT Strategy Manager”, “Business Process Manager” and “Human Resource Manager”. Seven of the stakeholders also had the role of sub-project-managers in the project. All stakeholders had the same coach, and a possibility to meet with the coach every three weeks during one year. Some chose to meet the coach more seldom. The practical goal of the usability coaching program was to encourage people to reflect on their work and behaviour and to achieve a personal accountability of usability issues in the coachee and to influence the organisation to work with usability on all levels. During the coaching sessions the discussions included for example problems and conflicts when introducing the ideas and perspectives of UCSD. Some of these perspectives included views of humans, work, and efficiency as well as automation of services in relation to creating a good work environment. Some coaching sessions concerned the use of different usability methods, but these were rare.

In a retrospective reflection the project manager explained the reasons for launching the usability coaching program in this way: *“We had already much increased the level of knowledge, and now there was a need for more continuous usability coaching discussions. I thought the timing was good, and we needed a new approach.”* Furthermore, during the same interview she presented her expectations of the usability coaching program in the following way: *“The expectation I had was that the sub-project managers would work harder. That they would get some support in making this change.”*

6 Experiences from the Case Study

Following we will describe and discuss our experiences doing usability coaching. Furthermore it will be related to learning theories and other research in the analysis.

During the final interviews of the project most of the coachees expressed that they had not gained any use of the coaching sessions or that they did not understand the purpose of the coaching sessions. This can be linked to their expectations of the coaching sessions, which might have been somewhat different than the expectations of the coach and the project manager who initiated the usability coaching program. It can also be linked to the difficulty of reflecting on changes. Moreover, one can note that a few of the people engaged in the mentoring program were not interested in having a mentor. One reasons for this was that they experienced that they already knew all they need to know about usability and their work.

The action research project was in its last year when the usability coaching program was initiated and the situation and the roles were more stabilized than in the beginning of the project. We believe that if the usability coaching program had started earlier, perhaps even in the very beginning of the project, the benefit of the sessions would have been more apparent to the coachees. This is also something that they themselves express in the interviews, that they believe the program would have been more beneficial in the beginning of the project. However, we have seen in the coaching sessions as well as outside the sessions that the coachees have changed their behavior, John, described below, is one example. These changes are subtle and stretched over time, and in hindsight perhaps difficult to see.

Conceptual Changes and Threshold Concepts

Conceptual Change [29, 30] and Threshold Concepts [31] are learning theories relevant to the usability coaching situation. That is, consciously intending to create a discussion that exposes the coachee to challenges of their own conceptual views and where they will meet new ideas requiring overcoming of threshold concepts. Conceptual change as described by Entwistle is used in order to understand the learning of the coachee [30, 32].

Threshold concepts are much discussed in literature on learning and teaching. Concepts can be seen as building blocks of a discipline, and the threshold concepts are among these building blocks. The threshold concepts in this setting are unique from other concept in the way that they are *transformative* as they change the way the coachees look at their work. They are also *integrative* and tie concepts together in new ways and *irreversible* in that they are difficult to unlearn. However, they might also be *troublesome* for the coachee as they are perceived as alien, difficult or counter-intuitive. Finally these threshold concepts are often *boundary markers* as they can be seen as indicating the limits of the conceptual area [33].

In the coaching context, the significance of the framework provided by threshold concepts lies in its explanatory potential to locate troublesome aspects when integrating usability in the organisation. It is principally an analytical framework for trying to understand how coachees learn, where the barriers to their learning lie and how they can be helped to overcome the difficulties. We have come across a few threshold concepts during the coaching sessions. One of the most visible is that IT-systems development creates new work and is responsible for a future good work environment is a concept that many coachees have understood. The results from the study suggests that some coachees have come to recognise that their influence goes beyond making computer systems; they are also influencing and creating the future work environment and health of those who will work with the system. Another threshold concept is that models of systems development are not used as step-to-step procedural descriptions of how to work in a systems development project. Systems development is a complex activity, and it consists of many situated decisions and problems that are impossible to solve with the help of prevailing standards or methods. Work is situated, and models of work cannot capture this aspect of work. Finally, we have identified that iterative systems development as a concept seems to be a threshold concept difficult to incorporate and to fully understand. Many of the coachees expressed that it is impossible to work iteratively due to reasons such as the test environment, the project management methods and the business plan for the authority.

The most important finding from this study is the power of impact the usability coaching program has had on many of the coachees, and on their way of talking about usability and taking responsibility for usability issues in their professional role. This could be seen from their behaviour in other contexts where usability was mentioned. From the theories of conceptual change this can be seen as if they have transformed through their new understandings through discovering, applying and assimilating new knowledge. The discussions with John have made him consider and reconsider the way systems development and architecture of IT is organised, and he questions and discusses many aspect of usability work in relation to his work. One example is a discussion about usability aspects when automating case handling in the spirit of e-Government. John believes that from an automation point of view, the user is irrelevant and that the goal of automation in the organisation is contradictory to the goal of usable computer systems and a good work environment. Here John identified two contradictory perspectives on work within the organisation, and had difficulties in understanding how to move forward in his work with this new knowledge:

"If the automation is our focus, then our focus isn't on the user and how he is supported by the system. And that becomes contradictory, I think. And from this perspective the user is irrelevant, if you know what I mean".

Personal Conflicts

During the coaching sessions many coachees described conflicts and discussions they had with other people and/or departments of the organisation. Also in the interviews, conflicts were mentioned and one of the perceived benefits of the coaching program as a good and constructive environment to discuss the handling of personal conflicts and dilemmas. Some of the conflicts were highly relevant for the project goals for example difficulties in communication with people or departments, and critical statements and rumours about the achievements of the project which resulted in discussions and conflicts. Other examples were that personal conflicts between people in the organisation made it impossible for some parts of the organisation to cooperate. The handling of the discussion around these conflicts was indeed a core part of the coaching sessions, which might seem quite surprising given that the coaching focused on usability and related activities to the introduction of the ideas behind UCSD.

These results concur with recent research in HCI [34] that describes usability work as a human activity where for example learning, personality, conflicts are illuminated. When working with the introduction of new ideas and perspectives conflicts are likely to occur as people change their work and view of the responsibilities in the organisation when collaborating. Nørgaard describes the problems of cross professional collaboration in the following way: *"Our work shows that cross-professional* *collaboration is subject to challenges that arise from stakeholders having conflicting priorities, procedures and personalities."* The amount of conflicts can also be seen as a result of the organisational change that the project aimed for. During a change process there are often frustrating feelings when people try to understand and make sense of the new organisation.

Private vs. Official

One of the most difficult problems with usability coaching is to decide what parts of the sessions that are private, and what parts are official. In many of the usability coaching sessions the coach and the coachee tended to discuss conflicts and personal experiences working with others. Some of these discussions have true relevance to the success of the usability project, and if discussed openly the possibilities for the project to succeed would improve. Examples of such conflicts are conflicts between the usability professionals and different units in the organisation.

A discussion about the consequences of considering the coaching sessions as private or official is appropriate. If the coaching sessions are considered as private this might imply that the coachee feels free to express opinions and perspectives in a more open manner. John expresses his view of the coaching session in the following way: *"Coaching is a relationship with high integrity; it's a thing between you and me."* However, later on in the interview he also talks about how beneficial the coaching sessions could be in helping him handling conflicts with others. That the coach can see the difference in expectations that different people have and give advice on how to handle this. This suggests that some of the discussions during the coaching sessions are very relevant to other coachees as they might describe interpretations of situations and conflicts that explain why difficulties arise. The question of confidentiality is explored in literature on coaching [15] where confidentiality is seen as fundamental for the success of coaching.

7 Usability Coaching – Towards a More Formal Method

As a conclusion of the experience of this case and of previous coaching experiences we want to specify the usability coaching method as a way to capture and disseminate the insights and knowledge we gained during the cases. Our understanding of the usability coaching method is inspired by Schön's [35] view of reflection-in-action: most real life situations are confusing messes where the problem has to be understood and the solution to the problem must be defined according to the person's personal, tacit experience rather than to any abstract general knowledge. Hence a description of the usability coaching method should not work as a normative guide that describes how a problem should be solved. Rather, we believe that a method seen from this reflection-on-action approach can be useful in a number of ways, for example as a common language and description of what to aim for in the coaching situation. In this study the word coach was chosen since it was what the organisation used when talking about the program. Moreover, the following definition is used to describe the usability coaching situation:

Usability coaching is a method that, based on the basic values of usability and UCSD, supports people in reflecting on their views and actions and on their role for promoting usability in the organisation and in any ongoing development activities.

Preparation

When preparing the mentoring program the coach needs to discuss the expectations and goals of the program. The organisation and the coachee might have slightly different goals, but there must be a main theme. The goals and the expectations are best discussed in meetings where the frame of the coach program is set up. Moreover, practical matters such as time, length and duration should be set. Other issues that need to be addressed are the accessibility of the coach. Is it OK for the coachee to mail and call in-between meetings? How much time should the coach spend reading, commenting or reflecting on reports, strategy plans, design, methods etc in-between meetings? When preparing the coaching sessions it is a good idea to have a general picture of the organisation and the organisational goals to understand the context of the coaching session. Field studies or an interview study is recommended in order to get this knowledge. Furthermore we also recommend that the coach keeps a diary with relevant information and has a plan of how to organize the information gained. The aim of the diary is memory recall and a way to remember things to look up or learn more about between meetings. Either the coach has one diary for all coach sessions, or separate diaries for each person. To support the goal of increased awareness of usability and UCSD we believe that the coach needs to be a senior usability professional with practical experience from working with usability methods and with an education in HCI subjects. In discussions the coach needs to be a good listener who is interested in understanding the coachee and his/her work situation. Preferable personality traits according to descriptive research in coaching are open-mindedness, patience and honesty [36, 37].

When looking at the organizational outcomes of the coaching program one need to discuss whether one should have one coach or several. Having one coach leads to the fact that the coach had knowledge and information about situations from several perspectives. When conflicts or problems were described by several coachees the coach in the study presented often had the possibility to make the coachee see the problem from different perspectives as other coachees had already talked about the problem. Other positive outcomes for the organisation resulting from the fact that there was one coach was the communication in the organisation was improved due to the fact that the coach could give tips and point at other parts of the organisation that worked with similar issues. Drawbacks of having one coach can be found in the fact that a group of active coaches might impact the organisation better since they would provide a multitude of different perspectives and experiences.

The complex nature of the problems addressed in the mentoring sessions meant that it was not a question of applying some research based theory and technique to solve the problem. Hence, potential solutions to the problems are found through a discussion and analysis of the problems at hand. The coachee and the coach contribute with their view of the situation, and the solutions are found through mutual understanding and discussions.

Implementation and Winding Up

During the coaching meetings the problems experienced by the coachee should be the focus, and the coach helps the coachee to see and understand the problem from different perspectives as described by [35]. Handling of personal conflicts between the coachee and others in the organisation is one aspect that the coach needs to prepare for. Are these a natural part of a coaching session, or are they on the agenda somewhere else? If they are a part, how should they be treated?

When preparing each coaching session the coach should read through the diary, which includes things to remember. Such things might include reading material provided by the coachee or to look up relevant areas of research etc. Preferably this is done a few days in advance depending on the time that needs to be spent on for example copying or reading. It might be a good idea to have a rough meeting agenda when coming to the coaching session, as for example: 1. What has happened since last time? 2. Problems that need to be discussed? 3. Anything more that needs to be discussed?

When winding up the usability coaching program it is a good idea to evaluate the learning experience. The method for the evaluation must be chosen to fit the situation at hand. The evaluation gives the coach input on what to improve in the coaching, and the coachee has the possibility to reflect on his/her learning experience during the mentoring sessions. However, it is not always possible to know what one has learned and in what way a coaching program has affected my sense making of my work and usability.

8 Recommendations for Research and Practice

Further research is needed to explore how the coachees made sense of the usability coaching and their experience of the impact on their actions. Here, recent theory building in research about coaching offers useful guidance considering what coachees learn in a coaching relationship, and offers a taxonomy of learning outcomes [19]. These categories include *affective learning*, which consist of changes in attitudes and motivation. Moreover, it would be interesting to further explore the ethical aspect of usability coaching, as usability coaching contains a number of ethical issues to be considered [38]. Another area of interest would be to evaluate the coaching together with the coachees. Our conclusion from this study is that usability coaching has the potential of being a powerful method in user-centred systems design and that organizations might gain from introducing this method. Our findings show that the usability coaching program made coachees more aware of their responsibility for usability, and in what ways usability related activities are part of their professional role.

References

1. Gulliksen, J., et al.: Key principles for user-centred systems design. Behaviour & Information Technology 22(6), 397–409 (2003)
2. Gulliksen, J., et al.: User-Centred Systems Design as Organizational Change: A Longitudinal Action Research Project to Improve Usability and the Computerized Work Environment in a Public Authority. International Journal of Technology and Human Interaction 5(3), 13–53 (2009)

3. Cajander, Å.: Values and Perspectives Affecting IT Systems Development and Usability Work, Department of Information Technology. Uppsala University, Uppsala (2006)
4. Cajander, Å., Gulliksen, J., Boivie, I.: Management perspectives on usability in a public authority: a case study. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 38–47 (2006)
5. Sandblad, B., et al.: Work environment and computer systems development. Behaviour and Information Technology - BIT 22(6), 375–387 (2003)
6. Greenbaum, J., Kyng, M. (eds.): Design at work: cooperative design of computer systems. ed. G. Joan and K. Morten, p. 294. Lawrence Erlbaum Associates, Inc., Mahwah (1992)
7. Kammersgaard, J.: Four Different Perspectives on Human Computer Interaction. In: Preece, J., Keller, L. (eds.) Human-Computer Interaction. Prentice-Hall, Cambridge (1990)
8. Gulliksen, J., et al.: Making a difference: a survey of the usability profession in Sweden. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction, Tampere, Finland. ACM Press, New York (2004)
9. Boivie, I.: A Fine Balance: Addressing Usability and Users? In: Needs in the Development of IT Systems for the Workplace, p. 85. Acta Universitatis Upsaliensis, Uppsala (2005)
10. Orlikowski, W.J., Gash, D.C.: Technological frames: making sense of information technology in organizations. ACM Transactions on Information Systems (TOIS) 12(2), 174–207 (1994)
11. Rasmussen, L.B.: Action research—Scandinavian experiences. AI & Society 18(1), 21–43 (2004)
12. McKay, J., Marschall, P.: The Dual Imperatives of Action Research. Information Technology & People 14(1), 45–59 (2001)
13. Gulliksen, J., et al.: A longitudinal action research case on the introduction of usability and user-centred design into a public authority. The International Journal of Technology and Human Interaction, IJTHI (2009)
14. Cajander, Å.: Usability Mentoring - An Exploratory Study. In: INTERACT 2007. Rio de Janeiro, Brazil (2007)
15. Garvey, R., Stokes, P., Megginson, D.: Coaching and Mentoring: Theory and Practice. Sage Publilcations Ltd., Thousand Oaks (2009)
16. Mertz, N.T.: What's a Mentor, Anyway? Educational Administration Quarterly 40(4), 541 (2004)
17. Hartfield, B., Winograd, T., Bennett, J.: Learning HCI design: Mentoring project groups in a course on human-computer interaction. SIGCSE Bulletin (Association for Computing Machinery, Special Interest Group on Computer Science Education) 24(1), 246–251 (1992)
18. Allen, T.D., Eby, L.T.: Relationship Effectiveness for Mentors: Factors Associated with Learning and Quality. Journal of Management 29(4), 469 (2003)
19. Wanberg, C.R., Welsh, E.T., Hezlett, S.A.: Mentoring research: A review and dynamic process model. Research in Personnel and Human Resources Management 22, 39–124 (2003)
20. Aryee, S., Lo, S., Kang, I.L.: Antecedents of early career stage mentoring among Chinese employees. Journal of Organizational Behavior 20(5), 563–576 (1999)
21. Colarelli, S.M., Bishop, R.C.: Career commitment: Functions, correlates, and management. Group & Organization Management 15(2), 158 (1990)
22. Kram, K.E.: Phases of the Mentor Relationship. The Academy of Management Journal 26(4), 608–625 (1983)

23. Schön, D.A.: Educating the reflective practitioner (toward a new design for teaching and learning in the professions). Jossey-Bass, San Francisco (1987)
24. Rittel, H.W.J., Webber, M.M.: Dilemmas in a general theory of planning. Policy Sciences 4(2), 155–169 (1973)
25. Gullliksen, J., et al.: A longitudinal action research case on the introduction of usability and user-centred design into a public authority. The International Journal of Technology and Human Interaction, IJTHI (2009)
26. Kavathatzopoulos, I.: AvI-enkäten: Ett verktyg för att mäta användbarhet, stress och nytta av IT-stöd, It Department. Uppsala University (2007)
27. Johansson, N., et al.: A participatory Process Supporting Design of Future Work. In: Singh, S.K. (ed.) Ergonomics: An Introduction. Icfai University Press, India (2007)
28. Eriksson, E., Cajander, Å., Gulliksen, J.: Hello World! - System Developers Doing Field Studies. In: INTERACT 2009, Uppsala, Sweden (2009)
29. Posner, G.J., et al.: Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education 66(2), 211–227 (1982)
30. Entwistle, N.: Conceptions of learning and the experience of understanding: thresholds, contextual influences, and knowledge objects. In: Re-framing the Conceptual Change Approach in Learning and Instruction, p. 123 (2007)
31. Meyer, J.H.F., Land, R.: Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge. Routledge, New York (2006)
32. Entwistle, N.: Threshold Concepts and Transformative Ways of Thinking Within Research into Higher Education. In: Meyer, J.H.F., Land, R. (eds.) Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising within the Disciplines (2003)
33. Meyer, J.H.F., Land, R.: Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines. Occasional Report, 4 (2008)
34. Nørgaard, M.: Understanding Usability Work as a Human Activity. Københavns Universitet, Det Naturvidenskabelige Fakultet, Datalogisk Institut, Copenhaguen (2008)
35. Schön, D.: The Reflective Practitioner - How Professionals Think in Action. Ashgate Publishing, Aldershot (1983)
36. Allen, T.D., Poteet, M.L.: Developing effective mentoring relationships: Strategies from the mentor's viewpoint. Career Development Quarterly 48(1), 59–73 (1999)
37. Allen, T.D., Poteet, M.L., Burroughs, S.M.: The mentor's perspective: A qualitative inquiry and future research agenda. Journal of Vocational Behavior 51(1), 70–89 (1997)
38. McDonald, K.S., Hite, L.M.: Ethical Issues in Mentoring: The Role of HRD. Advances in Developing Human Resources 7(4), 569 (2005)

The Focus on Usability in Testing Practices in Industry

Marta Kristin Larusdottir[1,2], Emma Run Bjarnadottir[3], and Jan Gulliksen[2]

[1] Reykjavik University
Menntavegur 1
102 Reykjavik, Iceland
marta@ru.is
[2] The Royal Institute of Technology
SE-100 44 Stockholm, Sweden
gulliksen@kth.se
[3] Calidris Ltd.
Vesturhlid 7
105 Reykjavik, Iceland
emmarun1@gmail.com

Abstract. A study exploring the focus on usability in testing practices in software development teams in Iceland using the agile software process Scrum is described in this paper. A survey was conducted to describe how testing is conducted and to what extent testing techniques are used. The results show that unit, integration, system and acceptance testing are the most frequent testing techniques used, but usability testing is not that common. Neither are alpha, beta, performance/load and security testing. Interviews were conducted to exemplify how practitioners conduct usability testing and what they describe as the difference between usability and acceptance testing. Some examples from the interviews show that practitioners are willing to do formal usability testing on extensive parts of the system, but because the iterations in Scrum are short and the changes to the system in each iteration are small, formal usability testing does not fit into the project work.

Keywords: Usability, software testing, agile development, Scrum, practitioners.

1 Introduction

This paper describes and discusses the results of a study conducted in April though August 2009, on how usability is emphasized in the practices of software testing in software development teams in Iceland using the agile development process Scrum [1]. A survey was conducted to study how usability is emphasized in testing by development teams in comparison to other testing techniques. Furthermore the team members were asked to compare their experience of testing in Scrum to their experience of testing in any prior development process. Interviews were used to exemplify how practitioners conduct usability testing and especially what they describe as the difference between usability and acceptance testing.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 98–109, 2010.

The need for studying the complexity of usability evaluation in practice was stated by Wixon [2], where he suggests that the usability evaluation should be studied in its real context, not on simulated systems or hypothetical models. Many researchers share his opinion, for example Cockton and Woolrych [3] recently stated that knowledge from empirical studies is needed. They suggest that these studies could use questionnaires, interviews or controlled experiments as research methods.

Furthermore the need for empirical studies of testing practices and evaluation has been identified, to develop an understanding of the organizational rationale for how software testing is practiced, for example in [4] and [5]. Software testing research has focused on extending the quality of testing in technical terms, like improving the design of tests, designing tools to support testing and measuring test coverage and efficacy of fault finding techniques [6]. Fault removal is only one of the potential goals for software testing; other goals include the evaluation of specified quality, e.g. usability or performance testing. It is preferable to spend the test budget to apply a combination of different techniques [6], even if one technique is shown to be the most effective. The actual use of different testing practices and the combination of those in the software industry needs further observation.

Modern agile development processes and traditional processes are often perceived as being opposed [7]. Traditional approaches tend to emphasize specifications and verification. These processes focus on the abstract and on correctness. Agile processes emphasize test and rapid prototyping, with focus on the concrete and the convenient and on validation. Both approaches emphasize co-development of the program and its evaluation. More and better empirical studies of agile software development within a common research agenda is needed [8]. An interesting aspect is whether it is easier to process testing in agile development processes than traditional processes.

Scrum is one of the agile development software development processes [1] where the focal aspects are simplicity and speed [9]. It progresses via a series of iterations called sprints, typically 2 – 4 weeks long. At the end of each sprint, a potentially shippable product increment should exist. The key roles in Scrum are Product Owner that defines the features of the product and defines what should be in each sprint, the Scrum Master that represents management to the project and software team members [1]. In Scrum there are four main activities, the sprint planning, the sprint review, where the team present what was accomplished in the sprint, the sprint retrospective, where the whole team discusses what is and what is not working in the project work and finally the daily scrum meeting, where all the team members report on their tasks. There are three main artifacts in Scrum, the product backlog, containing all desired work in the project, the sprint backlog that contains what will be done in the particular sprint and the burndown charts that illustrate the progress in each sprint [1]. One of the key factors in Scrum is that the team members never get ahead of the testers, because a requirement is ideally not “done” until it has been tested [10].

The focus of the study described in this paper is to research the actual use of testing techniques used in the software industry where the Scrum process is used, particularly focusing on how usability testing is performed in comparison with other testing techniques.

The research questions explored in the study are:

1. How is testing practiced in Scrum projects in the industry?

2. To what extent is usability testing performed compared to other testing techniques?
3. How does usability testing differ from acceptance testing in Scrum projects?

2 Research Methods

There were two research methods used in this study, a survey and interviews. The survey was done by sending a web based questionnaire to practitioners in software development companies using Scrum as their development process in Iceland. The respondents were all asked to take part in interviews after the survey. Six interviews were conducted from the pool of 9 people that agreed to being interviewed.

2.1 The Survey

The survey included 26 questions, 5 open questions and 21 multiple choice questions. There were 5 main sections in the questionnaire: a) background and experience of the respondent, b) information on the company where the respondent works, c) the software development process(es) used in the company, d) to which extent and who is conducting different testing techniques and e) the change in conducting software testing when compared to previous/parallel software development process.

The survey was sent to 20 software companies that had confirmed using the Scrum process and doing software testing. The questionnaire was distributed to a single contact at each company, previously chosen to fit the target audience. The respondents were asked to send the questionnaire to another person in the company, if they thought that person fitted better the target audience. The results from the survey are in most cases based on 25 responds from 18 companies. In some cases one answer per company is used, e.g. when asking about the company's industrial section.

Most of the respondents, 76%, do have a degree in computer science or engineering. The majority of the respondents were male 68%, 20% female and 12% gave no reply. Nobody had been working for less than a year, 28% had been working for 1 – 3 years, 24% for 4 to 9 years and 28% for 10 to 14 years. Sixteen percent had been working for more than 15 years and one respondent did not reply.

Half of the companies are in the computer service or software development industry, 22% are in the banking, insurance or finance, 11% are in gaming, 11% in telecommunication and telephone, and 6% in the airline industry. The results on the size of the software companies show that 33% had up to 19 employees, 28% had 20 to 59 employees, and 33% over 60 employees. No data was given in 6% of the cases.

All the 18 companies use Scrum as a development process to some extent. Almost half of the respondents say that 81% – 100% of the software development department use Scrum, 34% said it is 21% - 80% and 22% were using it to the extent of 0% – 20%. In 61% of the cases another software development process was also used beside Scrum within the company, for example the agile process XP, Waterfall process and RUP. Interestingly 44% mentioned using their own process besides other variations. The respondents were asked about their primary role in relation to the Scrum development process. Almost half of them 44%, were Scrum Masters, 24% software testers, 20% Product Owners and 12% had other role.

2.2 The Interviews

To strengthen the results from the survey and to exemplify the reason for the results, six interviews were conducted, three with software testers and three with Scrum Masters. One of the goals of the interviews was to gather data to deepen the understanding of the results for the third research question on the difference between usability and acceptance testing.

The same person conducted all the interviews face to face and prepared a set of questions for the interviews. Each interview was recorded and transcribed afterwards. The interviews were semi structured, so the interviewer asked more questions than in the interview guide, if some issues were relevant in that particular case.

3 Results

The results are described in three subchapters. First the results on how testing is practiced in Scrum in the companies involved are described. Second, the results on how usability testing is practices compared to other testing techniques are represented and finally the results on the difference between usability and acceptance testing are explained.

3.1 Results on How Testing Is Practiced in Scrum

Table 1 describes the results on fundamental activities of the Scrum development process. Many of those activities are practiced up to 90% or more. The activity that is least practiced is playing planning poker, which should be used to estimate the effort needed to complete the requirements in the particular sprint. Potentially shippable product after each sprint has been one of the main selling points for Scrum. There are only 64% of the respondents that say that this is practiced.

Table 1. The practice of fundamental activities in Scrum development

Scrum activity	Percent of repondents
The role of Product Owner is used	88%
The role of Scrum Master is used	92%
Requirements/user stories are kept in a Product backlog	84%
Planning poker is played on Product backlog items	60%
A Sprint backlog is established at the beginning of a sprint	92%
Each sprint/iteration is 2 - 4 weeks long	96%
Potentially shippable product exists at the end of a sprint	64%
A sprint review meeting is held at the end of a sprint	80%
A sprint retrospective meeting is held at the end of a sprint	88%
Scrum (Agile) metrics are used, like burn down charts	76%

The respondents were asked to name at least three things they believed to be positive effects of the Scrum process and compare with parallel/prior development process. Eighty percent of the respondents gave some comments to this question. The comments indicate that the respondents are generally happy changing over to using Scrum as their development process. Some examples are “more productivity”, “QA involvement is a lot better”, “combined responsibility”, “better morale”, “relationship with client improved”, “customer in closer connection with the development and can change functions before final release”, “less disturbance”, “easier to handle changes in requirements/priorities” and “stand up meeting keep communication open”.

Respondents were also asked to comment on what could be improved in relation to Scrum. The results indicate that there is always a room for improvement. To give some examples, “don't over commit”, “more design documentation”, “internal marketing of Scrum projects” and “organizational changes” are all comments from the respondents.

Table 2. The applicability of testing practices in Scrum projects

Testing practice	Percent of repondents
Software testing falls within the frame of "done" in each sprint	64%
Software testing is squeezed into the end of each iteration	36%
Software testing is not well integrated with coding and ends up one sprint behind	20%
Software testing is performed in a separate test environment	44%
Good management of version control	60%
Before a major version release, there is a bug-fix sprint	40%
Software testing became easier than in a parallel/prior process	44%
Overall more software testing is done than in a parallel/prior process	44%
Overall less software testing is done than in a parallel/prior process	12%
Programmers started using more test-driven development/design	48%
Software testers became more involved throughout the whole development	72%

In Table 2 are results on how testing is practices in Scrum and how that compares to prior or parallel process. The positive results are that 60% say that they have good management of version control and 64% say that testing falls within the frame of “done”, which means that the required functionality is not delivered to the customer before it has been thoroughly tested. The testing does only end up one sprint behind in 20% of the cases and is squeezed into the end of each sprint in 36% of the cases.

When asked about the testing practice in comparison to prior or parallel process, 72% of the respondents say that testers became more involved throughout the whole development which is positive. The result show that 44% of the respondents say that software testing became easier than in parallel or prior process. Also 44% of the respondents say that more testing is done and only 12% say that less testing is done. These results show that Scrum has positive effects on testing practices.

3.2 Results on Usability Testing Compared to Other Testing Techniques

The results from the survey in this subchapter build up to answer the research question:

To what extent is usability testing performed in relation to other testing techniques?

Table 3. The description of each testing technique that the respondents got in the survey

Testing technique	Description
Unit/component testing	The testing of individual software components.
Integration testing	Testing performed to expose defects in the interfaces and in the interactions between integrated components or systems.
System testing	The process of testing an integrated system to verify that it meets specified requirements. This includes test design techniques like boundary valued analysis and is usually done by internal software testers.
Acceptance testing	Formal testing with respect to user needs, requirements and business processes conducted to determine whether or not a system satisfies the acceptance criteria and to enable the users, customers or other authorized entity to determine whether or not to accept the system.
Usability testing	Testing to determine the extent to which the software product is understood, easy to learn, easy to operate and attractive to the users under specified conditions.
Alpha testing	Simulated or actual operational testing by potential users/customers or an independent test team at the developer' site, but outside the development organization. Alpha testing is often employed for off-the-shelf software as a form of internal acceptance testing.
Beta testing	Operational testing by potential and/or existing users/customers at an external site not otherwise involved with the developers, to determine whether or not a component or system satisfies the user/customer needs and fits within the business processes. Beta testing is often employed as a form of external acceptance testing for off-the-shelf software in order to acquire feedback from the market.
Performance/load testing	The process of testing to determine the performance and/or measuring the behavior of a component or system with increasing load, e.g. the number of parallel users and/or numbers of transactions, to determine what load can be handled by the component or system.
Security testing	Testing to determine the security of the software product.

The respondents were asked to answer how much testing is done of each testing technique, who is performing the testing and give reasons for why particular testing technique is used less than other testing techniques. The results are described below. When respondents were asked about the testing techniques, these were explained as described in table 3, according to descriptions in [11].

The participants were asked to rate how much testing is done in their Scrum projects. The results are shown in table 4. The most common answer is showed in bold letters and the results on usability testing are shown in italics.

Table 4. The extent to which the testing techniques are used in Scrum projects, N = 23

Testing technique	Yes, a lot	Yes, some	So and so	Little	No, not at all
Unit/component testing	22%	**35%**	26%	13%	4%
Integration testing	17%	**35%**	31%	13%	4%
System testing	**39%**	30%	22%	9%	0%
Acceptance testing	30%	**44%**	13%	13%	0%
Usability testing	*4%*	*22%*	***35%***	***35%***	*4%*
Alpha testing	4%	13%	17%	17%	**48%**
Beta testing	9%	22%	9%	17%	**44%**
Performance/load testing	0%	26%	26%	**35%**	13%
Security testing	4%	22%	8%	**39%**	26%

Table 5. The person using each testing technique in Scrum projects

Testing technique	Programmer	Software tester	External Software tester	Customer	Others	N
Unit/component testing	**73%**	17%	3%	7%	0%	23
Integration testing	**44%**	36%	6%	11%	3%	22
System testing	21%	**50%**	13%	11%	5%	23
Acceptance testing	6%	33%	8%	**44%**	8%	21
Usability testing	*9%*	***43%***	*14%*	*26%*	*9%*	*22*
Alpha testing	13%	**35%**	13%	30%	9%	12
Beta testing	9%	32%	14%	**41%**	4%	13
Performance/load testing	**47%**	43%	3%	3%	3%	21
Security testing	**44%**	33%	4%	4%	15%	18

The results in table 4 show that system testing is the most common testing technique. Fairly common are also acceptance, unit and integration testing. Usability testing and performance/load testing have a similar use pattern, though more respondents chose the alternative "so and so" for usability testing than for performance/load testing. Security testing is not done as much as performance/load testing, but still there are 26% that do security testing to some extent or a lot. Almost half of the respondents say that alpha and beta testing are not used at all.

In Table 5 results on who is using each testing technique are listed. Since some answers were lacking, the actual N is labeled for each testing technique in the table. Software testers use all the testing techniques extensively as would be expected, 50% of the system testing is done by software testers, 43% of the usability testing and 43% of the performance testing. Programmers are mainly responsible for unit/component testing. Customers are taking considerable part in acceptance, beta and alpha testing and usability testing is done by customers in 26% of the cases. Performance/load and security testing are done jointly by software testers and programmers.

Table 6. The reasons for using some testing techniques less than others

Testing technique	Lack of training/ knowledge	Lack of budget	Lack of time	Other	N/A	N
Unit/component testing	**36**%	0%	32%	5%	27%	22
Integration testing	11%	0%	42%	0%	**47**%	19
System testing	7%	0%	**47**%	0%	**47%**	15
Acceptance testing	7%	0%	27%	7%	**60**%	15
Usability testing	*20%*	*15%*	***35**%*	*10%*	*20%*	*20*
Alpha testing	0%	11%	11%	10%	**68**%	19
Beta testing	0%	11%	17%	11%	**61**%	18
Performance/load testing	26%	11%	**32%**	0%	**32%**	19
Security testing	**47%**	5%	16%	0%	32%	19

Table 6 analyzes what causes some testing techniques to be used less than others. Again some answers were lacking, and the actual N is labeled for each testing technique. The main reason for why usability testing is not conducted is lack of time. Lack of training was the reason in 20% of the cases and lack of budget in 15% of the cases. In this particular question many of the respondents chose the alternative "not applicable", which probably means that they do not know the reason.

The respondents were finally asked if they were missing any kind of software testing. Only a few answers were given, mentioning exploratory and regression testing.

3.3 Some Examples for the Reasons for These Results

Six interviews were conducted to get some examples on the reasons why usability testing is not practiced to a wider extent, how the interviewees explain the importance of usability testing and if usability testing is something that could be ignored. Furthermore some examples on how practitioners explain the difference between acceptance testing and usability testing are described.

Three interviews were conducted with software testers and three with Scrum Masters. All the interviewees had answered the questionnaire three months earlier and the interviewer had their answers to refer to during the interview. The results are explained in the following.

3.3.1 The Importance of Usability Testing

When asked about the importance of usability testing all the respondents mentioned that if a project is big and many changes have been made, formal usability testing would definitely be important. One respondent explained that if the project is really agile, the changes are not that extensive each time and the importance of being quick to market is strong, so usability testing is really not needed, because a shippable product has been delivered and the customers can complain. Furthermore because the changes are small, extensive usability testing is not needed and is too expensive. "The main thing is to confess your fault and change quickly according to the customers complaints so you can be very quick in adjusting to their needs" one of the respondents remarked. This respondent explained that asking for usability testing was really the customer's responsibility.

Some respondents remarked that the users were not always willing to take part because they were busy doing their own work and did not want to be involved in software development and one respondent said that usability is fuzzy, hard to measure and usability requirements are always changing.

All the respondents were asked whether they would like to do more usability testing if they had time and money. All the respondents wanted that at least occasionally. They were also asked if usability testing was something that could be ignored. None of the respondents wanted to ignore it. One of the respondents specified that there is a need for a formal usability test of the system once a year or every second year, but during other testing usability issues are implicitly considered, "It is always on my mind" the respondent commented.

3.3.2 The Difference between Acceptance Testing and Usability Testing

All the respondents were asked about what the difference between acceptance testing and usability testing is. Many explained that acceptance testing is more structured that usability testing. During acceptance testing, the functionality of the new system is shown to the customer in predefined steps. The customer has to sign that he or she has accepted that the requirements are fulfilled.

The respondents all agreed that usability testing is more about testing how useful the system is for the particular users. Often users were involved in the testing, but in some cases the customers, the persons paying for the system, took part and not the actual users. In some cases the requirement analyzers were responsible for the usability testing. Some of the respondents also mentioned that they asked external usability professionals to do the usability testing. Their rational was that because knowledgeable external testers were available, training people within the company was not needed.

One form of testing mentioned by several of the interviewees was user acceptance. During user acceptance testing the users check if the system is developed according to what they asked for, check that all the calculations are right and accept the system. The users get test cases and go through those step by step, so the tasks that the users get are not as open as in usability testing. If the users see some usability issues they are asked to inform on that, but there is not a formal process of usability testing at this point. User acceptance is done in the end of the project, but sometimes it is done earlier to show the users what has been developed.

4 Discussion

4.1 Testing Practices in Scrum

Many of the studies on testing practices are analytical comparisons. One recent example is the study described in [12], where the expected number of failures detected by different testing techniques is compared. The analytical studies are really interesting from a theoretical viewpoint, but further investigation is needed before these can be applied in practice because the main focus in practice is to minimize the effort and time required to demonstrate that the software is good enough [4]. The negative consequences of this lack of information is that testing remains one of the most expensive activities of the development process and with delivered products of poor quality and 4 used less than other testing techniques.

The results from the interviews indicate that testing in Scrum is easier because the teams in Scrum are often arranged that way that one of team members has the testing role. The team discusses the ongoing tasks on their daily meetings so the testers know well the functionality that is ready for testing. Some of the respondents in the interviews indicated that all the team members know the progress of the tasks in Scrum, so they realize that some testing tasks are pending and acknowledge that.

4.2 Factors Affecting Testing Practices

Various factors that affect the testing processes are emphasized in current research. These include: the involvement of testing in the development process, communication and interaction between the development and the testing teams, management of the complexity of testing, risk-based testing among other factors [13]. Involvement of testing in the development processes is a complicated issue to study, because the software phases glide in parallel, as noticed in a study by Baskerville, et. al. [14]. Their main results show that developers run their testing or quality assurance in parallel with other development phases, which results in adjusted processes.

In a traditional software development process, testing is mainly conducted at the end of the project time. Testing in agile or Scrum development processes differs from testing in traditional processes, such that each increment of code is tested as soon as it is finished, e.g. the testing becomes dynamic almost from the beginning [10].

The results from the survey show that in 36% of the cases, testing is squeezed into the end off each sprint. This indicates that the opposite is true in 64% of the cases that is each increment of code is tested as soon as it is finished. Only 20% of the respondents say that testing is not well integrated with coding and ends up one sprint behind. Furthermore 72% of the respondents say that testers became more involved throughout the whole development.

Some of the respondents in the interviews mentioned that the testers working so closely with the teams had also some negative consequences. There aren't separate testing teams in all cases, so all the testing is done by team members that sometimes do not have the distance needed to test the important aspects. They are too much involved in the actual development to be able to test the system neutrally. Testing the increments done as soon as they are finished has also the side affect that formal testing of the whole system emphasizing some quality attribute like usability, is not done. One respondent mention that including testers in the team was positive, but separate testing teams were also needed to cover the more extensive testing approaches.

In this study exploring the difference of acceptance testing and usability testing was particularly emphasized because the goal of both of these testing techniques is to determine if the system fits the user's needs. One of the outcomes of this study is that practitioners view these testing types quite differently. Acceptance testing is commonly accepted and done in all cases, but usability testing is not as widely done. The most common reasons for this are lack of time, lack of training and lack of budget.

5 Conclusion

To summarize the results, the fundamental roles, activities and artifacts suggested in Scrum development process are used to great extent in the companies involved and the sprints are usually 2 – 4 weeks as recommended. Potentially shippable product after each sprint is only accomplished in around 2/3 of the cases though, which is less than expected because this is one of the fundamental issues in the Scrum process.

About half of the respondents said that testing became easier in Scrum than in prior or parallel process used and more testing is done. Only a few said less testing is done. The reason could be that one team member has often the testing role in Scrum and the daily Scrum meetings give good overview of what needs to be tested.

Usability testing and performance testing are practiced in a similar way, but unit, integration, system and acceptance testing are much more frequent. Some examples from the interviews show that practitioners would like to do formal usability testing on extensive parts of the system, but because the iterations in Scrum are short and the changes to the system after each iteration are small, formal usability tests do not fit into the project work. When the focus of testing is on quality attributes like usability, security or performance, the examples in the study show that the testers would prefer to carefully plan their tests. Planning and conducting the tests requires extensive workload from the testers, which is in contrast with the fundamental principles in Scrum, simplicity and speed. The implications for further development of usability testing in Scrum are to find ways of testing the usability on a smaller scale so it can be integrated into the testing activities in each sprint.

Acknowledgement

We thank all the respondents that took their time responding to the survey and particularly the interviewees. Special thanks to Ólöf Una Haraldsdóttir that conducted all the interviews and to Marjan Sirjani, associate professor at Reykjavik University, for her invaluable comments on several drafts of the paper.

References

1. Schwaber, K.: Agile Project Management with Scrum. Microsoft Press, Redmond (2004)
2. Wixon, D.: Evaluating Usability Methods: Why the Current Literature Fails the Practitioner. Interactions 10(4), 28–34 (2003)
3. Cockton, G., Woolrych, A.: Comparing Usability Evaluation Methods: Strategies and Implementation – Final report of COST294-MAUSE Working Group 2. In: Law, E.L., Scapin, D., Cockton, G., Springett, M., Stary, C., Winckler, M. (eds.) Maturation of Usability Evaluation methods: Retrospect and Prospect. IRIT Press, Toulouse (2009)
4. Martin, D., Rooksby, J., Rouncefield, M., Sommerville, I.: 'Good' Organisational Reasons for 'Bad' Software Testing: An Ethnographic Study of Testing in a Small Software Company. In: Proceedings of International Conference on Software Engineering, pp. 602–611 (2007)
5. Bertolino, A.: The (Im)maturity Level of Software Testing. SIGSOFT Softw. Eng. Notes 29(5), 1–4 (2004)
6. Littlewood, B., Popov, P., Strigini, L., Shryane, N.: Modelling the Effects of Combining Diverse Software Fault Detection Techniques. IEEE Transactions of Software Engineering 26(12) (2000)
7. Rayside, D., Milicevic, A., Yessenov, K., Dennis, G., Jackson, D.: Agile Specifications. In: Proceeding of the 24th ACM SIGPLAN Conf. Companion on Object Oriented Programming Systems Languages and Applications. ACM, New York (2009)
8. Dybå, T., Dingsøyr, T.: Empirical Studies of Agile Software Development: A Systematic Review. Information and Software Technology 50(9-10) (2008)
9. Abrahamsson, P., Warsta, J., Siponen, M.T., Ronkainen, J.: New Directions on Agile Methods: a Comparative Analysis. In: Proceedings of the 25th International Conference on Software Engineering. IEEE Computer Society, Los Alamitos (2003)
10. Crispin, L., Gregory, J.: Agile Testing: A Practical Guide for Testers and Agile Teams. Addison-Wesley, Reading (2009)
11. Graham, D., Veenedaal, E.V., Evans, I., Black, R.: Foundation of Software Testing: ISTQB Certification, Thomson (2007)
12. Morasca, S., Serra-Capizzano, S.: On the Analytical Comparison of Testing Techniques. In: Proceeding of ACM ISSTA. ACM Press, New York (2004)
13. Taipale, O., Smolander, K.: Improving Software Testing by Observing Practice. In: Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering, ISESE 2006. ACM, New York (2006)
14. Baskerville, R., Levine, L., Pries-Heje, J., Ramesh, B., Slaughter, S.: How Internet Software Companies Negotiate Quality. Computer 34(5) (2001)

Added Value of Eye Tracking in Usability Studies: Expert and Non-expert Participants

Marco C. Pretorius, Judy van Biljon, and Estelle de Kock

School of Computing
University of South Africa
Pretoria, South Africa
marco.pretorius@gmail.com, vbiljja@unisa.ac.za,
dkocke@unisa.ac.za

Abstract. This paper investigates the value of eye tracking in evaluating the usability of a Learning Management System, at an open distance learning university where the users' computer and Web skills vary significantly. Eye tracking utilize the users' eye movements, while doing a task, to provide information about the nature, sequence and timing of the cognitive operations that took place. This information supplements, but does not replace standard usability testing with observations. This forces the questions of when the added value of eye tracking justifies the added cost and resources. Existing research has indicated significant differences in the usability experienced by experts and non-experts on the same system. The aim of this paper is to go one step further and shed light on the type and severity of the usability problems experienced by non-expert users. Usability testing with eye tracking is a resource intensive method but our findings indicate that eye tracking adds concise, summarised evidence of usability problems that justifies the cost when testing special groups such as users deficient in Web and computer skills. The contribution of this paper is to highlight the added value of eye tracking as a usability evaluation method in working with Web non-expert users. Furthermore, the findings improve our understanding of the knowledge differences between expert and non-expert Web users and the practical challenges involved in working with non-expert users.

Keywords: Usability, Eye tracking, expert, non-expert.

1 Introduction

The Web is evolving into the preferred delivery mode for Learning Management Systems (LMS) which offer services such as student enrolment, coursework notice boards and assignment submission. A LMS is defined as a Web-based application used by institutions and companies that want to get involved in e-learning either for providing services to third parties, or for educating and training their own people [1]. In developing countries the students' computer and Web skills vary, some are experts with sophisticated skills while others lack the most basic computer skills. The problem is compounded by the fact that the system is used occasionally and under time

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 110–121, 2010.

constraints such as registering or submitting an assignment before a due date and even financial constraints where students use commercial internet facilities. When the Web interface of a LMS becomes the only access point, usability becomes critical. Designing for novice users has been well-researched. However, eye tracking as evaluation tool for evaluating non-expert web users is novel. This raised the question: What is the added value of eye tracking in evaluating non-expert web users?

We can predict that computer expertise will influence usability positively, but the use of eye tracking could possibly help to find the severity of the effect on non-expert users and if anything can be done to improve the usability for these users. To answer these questions we conducted a usability study on the assignment submission section of a university LMS purposively selecting expert and non-expert users respectively. In order to identify the added value of eye tracking, we captured data by different means including eye tracking, observation (direct observations and video recordings) and a post-test questionnaire. We analysed the data sets separately and compared the findings to see if eye tracking did contribute additional knowledge.

Eye tracking has been used to research the difference between experienced and less experienced users in information retrieval tasks [2] and different styles have been associated with experienced and less experienced users [3]. In general expert users perform faster, more accurate and have more defined search paths whereas non-experts waste time searching or looking at irrelevant information [4, 5]. However, eye tracking has not previously been widely applied to studies of behavioural aspects of students using a LMS. What is lacking is an explanation of the implications the difference between expert and non-experts could have for the usability of a LMS.

The main contribution of the paper is to identify the added value of eye tracking in investigating usability problems experienced by non-experts in using a LMS. The scope of the research is limited to the evaluation of one LMS but the data sets captured through video and audio recordings, eye tracking video recordings, eye tracking data files and post-test questionnaires are triangulated to provide more reliability.

2 Related Usability Studies

Any research comparing expert and non-expert users of an interactive system has to take cognisance of the fundamental distinction between these two, founded in their knowledge and experience and therefore we first review knowledge categories before discussing usability evaluation paradigms, focusing on heuristic evaluation and eye tracking.

There are different empirical definitions of experts and non-experts but two (strongly overlapping) criteria used for the differentiation between experts and non-experts are *knowledge* and the *time* spent working with a particular system [6]. Given this fact that expertise is related not only to time spent working with the system, we use the term non-expert rather than novice. Furthermore, the term non-expert is useful for grouping categories such as novices and intermediate users, both not expert. Focusing on the knowledge component, Prumper et al. [6] states that experts operate at higher cognitive levels, i.e. using knowledge from different categories than non-experts do.

Working with non-experts require more structure in terms of doing tasks, more assistance and more encouragement in reporting problems [7], but Fields, Keith et al. [8] emphasize the value of user groups with varying forms of expertise. Popovic [9] states that all interactive interface design should support and facilitate the transitional process from novice to expert user. Gorman [10] classifies four knowledge categories, i.e. declarative knowledge (what), procedural knowledge (how), judgement knowledge (when) and wisdom (why) that we will use as reference.

The way experts and non-expert users use interactive Web sites is differentiated by the difference in their knowledge (particularly domain-specific knowledge) [9] and also by the way they approach and solve problems [11]. Äijö and Mantere [7] conducted a study on the usefulness of non-experts in identifying the usability problems in an interactive Web service. They found that non-experts were capable of reporting usability problems on the interface but had difficulties in reporting usability problems related to interactivity with the system. This concurs with Kotze et al. [12] who state that although experts and non-experts may differ in all the categories, novices rely more on declarative and to a lesser extent on weak heuristics based on procedural knowledge while experts rely more on judgment knowledge.

We now review the use of eye tracking in usability studies. The human eye reads a line of text in discrete chunks through a series of fixations and fast eye movements to take in the next section [13]. Eye tracking is based on the fact that a record of a person's eye movements while doing a task provides information about the nature, sequence and timing of the cognitive operations that took place [14].

Eye tracking studies have been used in diagnosing the effectiveness of Website designs with point of interest detection and information transmission via eye movement (scan path) as two main indicators [15]. Based on this relation between cognition and eye behaviour the trace of navigation pathways and user attention patterns is used to study the cognitive processes involved in reading [3], picture perception [7, 13], visual search [16], problem solving [5], face perception [17] and many other tasks.

In HCI eye tracking has been used to study the usability of Web pages [18], menu searching, information searching from Web pages and search result evaluation [3]. Eye tracking studies on the difference between expert and non-expert users in HCI are difficult to find but some have been done in other fields. The eye movements of experts and novices were compared in a virtual aiming task done in a laparoscopic surgery training system. The results showed significant differences in performance as well as eye movement between the two groups [4]. A study in the field of aviation compared expert and novice pilots to find that experts have better defined, more consistent and efficient eye-scanning patterns [5].

Eye tracking has the advantage of providing visual evidence of usability problems but the method has been criticized for reliability and objectivity. Sutcliffe and Namoune [19] question the reliability of eye tracking for determining which features are attended because they found that users discovered areas of interest by short fixations or proximal fixations i.e., it is difficult to know exactly what the user was looking at. Further criticism on eye tracking is the objectivity: the analysis of patterns is mostly based on the opinion and interpretations of the individual evaluator [20]. Correlating eye-movement metrics with usability problems is proposed as a more rigorous, verifiable mapping of eye tracking patterns to usability problems [20].

3 Method

The method for this usability and eye tracking study was based on the methodology by Pretorius et al. [21]. Section 3.1 presents the participants' profile, section 3.2 discusses the procedure followed and section 3.3 provides more detail on how the data was captured.

3.1 Participant Profile

The intended user group for the LMS is students who have to submit assignments online. A screening questionnaire was used to screen the participants for this evaluation. This questionnaire reflected the possible participant's LMS experience, computer and Internet experience, culture, age and gender. Regarding expert and non-expert users, results showed that individuals' ratings of their overall knowledge were better predictors than were estimations of frequency of use [22].

The screening questionnaire asked participants to rate their experience level as an Internet user. The following ratings were available: *Never used the Web*; *Beginner* – have read pages on the Web; *Novice* – have entered addresses and used bookmarks; *Competent* – can use a search engine to find information; and *Proficient* – know way around and have done Web transactions like e-banking.

From the 23 questionnaires completed, we selected 10 students, five who rated themselves as proficient Internet users and another five who rated themselves between a beginner and competent level. The first group were referred to as expert participants and the latter group as non-expert participants.

The participants had all used some sections of the LMS, their experience ranged from less than one year; between one and two years; and two years or more.

Our participants included five male and five female students. One student was below 20; six between 21 and 25; one between 26 and 30; and two above 30. The mother tongue (first language they learned to speak) of the 10 participants was as follow (number of participants indicated in brackets): English (2); isiXhosa (2); Afrikaans (2); SiSwati (1); Sesotho (1); Xitsonga (1); Yoruba (1).

3.2 Conducting the Test

One participant was tested at a time. On arrival, the participant was briefed about the experiment, which was followed by an explanation of the equipment to be used. The details of the material to be recorded were explained and the participant was required to complete an informed consent form.

A variation of the think aloud protocol was used, where participants were asked to comment if they were looking for something and could not find it; if they liked something particular about the Website; and if they disliked something particular.

Participants completed three tasks available from a task list (discussed in Section 4). Following the tasks, a quick interview was conducted to ask the participants what they liked most about the system, what they did not like, as well as questions about issues that the administrator picked-up. This was followed by a post-test questionnaire and a debriefing where the participant was thanked and given the opportunity to see the data.

3.3 Data Collection Method

Data was collected and calculated as follows: live video recordings were captured, including the screen, participant's face and mouse/keyboard movements; notes were taken during the test as well as a full evaluation of the video at a later stage; audio in the form of the participant or the test administer speaking were included with the video files; eye tracking video recordings included a cursor which indicates the participant's eye movements; eye tracking data files; a post-test questionnaire was used to capture the participants' perception of the user interface and the system; and continuous monitoring of tasks.

The usability laboratory consists of an observer room and a participant room, separated by a one-way mirror. The participant room is equipped with a 17" TFT monitor with resolution of 1280x1024 and a Tobii 1750 eye tracker, allowing the eye movement of participants on the screen to be recorded. A 9-point eye tracking calibration was used at all times.

4 Results and Findings

This section discusses the results and findings obtained from the usability test.

4.1 Task 1

The participant had to submit a written assignment (a Word file) for this task. This was done by entering or browsing for the correct file, before selecting the appropriate file format from a drop-down box. The Word file format was the default choice.

The usability results, as depicted in Table 1, showed a clear distinction between expert participants and non-expert participants. The expert participants had no problems with this task with task completion time ranging between 27 and 50 seconds. Expert participants made no errors and no assistance were needed.

In contrast, the non-expert participants struggled a great deal with task completion times significantly longer. Four of these participants made errors during this task; three needed assistance; and one was not able to complete the task. The participant who could not complete the task received assistance, but this did not help; a complete run through of the task was needed. The task completion time of this participant was removed from the sample (exceeded nine minutes). This task had to be turned into a training task that would help the participant to complete the subsequent tasks in the task list. This non-expert participant did not understand computer terms such as "minimise" and "desktop".

None of the non-expert Web participants could comprehend the term "File Name", where the name of the task file is entered or selected. One participant asked what a

Table 1. Usability results for task 1

	Completed	Time (in seconds)	Errors	Assistance needed
Expert	5	27-50	0	0
Non-Expert	4	68-161	4	3

file name is and another entered the course code as the file name. The participant asked “how do I load a file from my PC to the LMS?” with a confused facial expression. After clicking “*continue*” without entering a file name, the participant received an error message which read: *“ERROR: The type of file does NOT match the selected file type. (xxxxxx!=DOC)”*. The participant did not understand the error message.

One participant clicked on “*continue*” without entering a file name; after the error message the participant entered the course code as the file name. Another participant entered his own name in the file name text box. The participant clicked ‘continue’ and received an error message; repeating the same mistake three times.

The eye tracking results provide clear evidence of the difference in usability between expert and non-expert participants. The scan paths for non-expert participants (Figure 1) are much longer and have more fixations scattered over the screen. This shows us that participants searched and did not know exactly what to do.

As stated in the usability results, all five non-expert Web participants could not comprehend the word “File Name”.

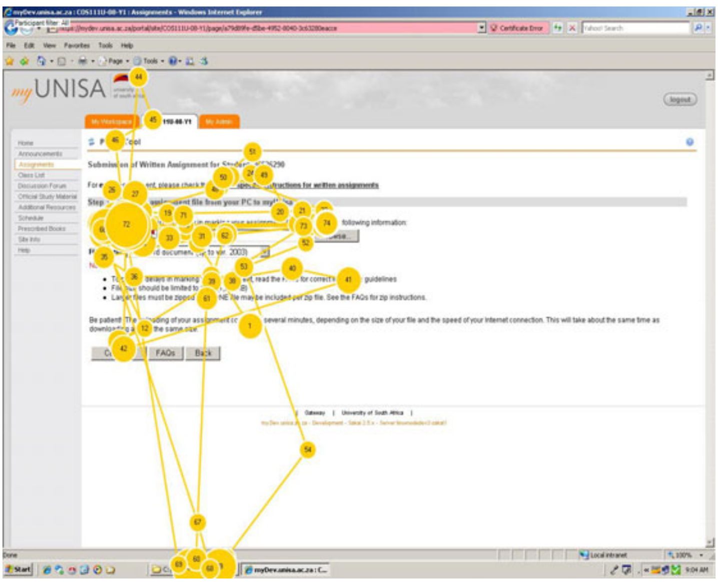


Fig. 1. Non-expert participant: long scan path with many fixations

The scan paths for expert users (Figure 2) are much shorter with fewer fixations. Eye tracking allows an evaluator to see what the participant was reading (and not reading) thereby identifying the problem areas on the screen as well as the information that goes unnoticed.

Throughout the task list, eye tracking data showed that the majority of participants did not read the instructions. Figure 3 demonstrates such an example where a participant did not read the instructions. The instructions are indicated by a rectangle. The phrase *“Here is your final chance to check that your assignment is correct”* is included in these instructions. Only one of ten participants read these instructions.

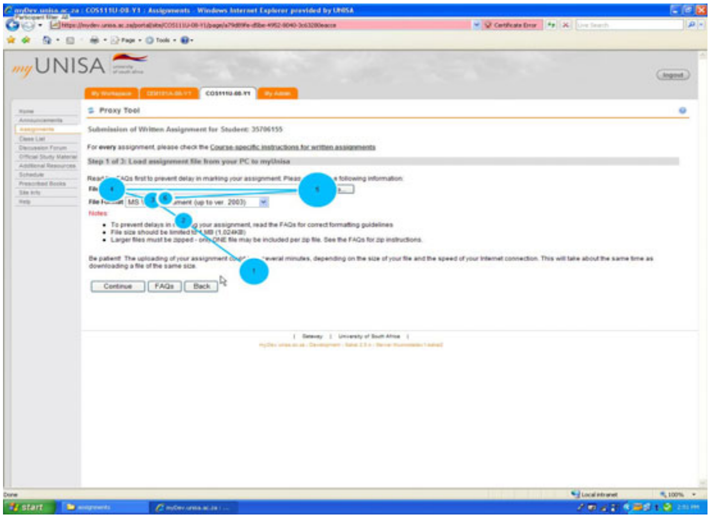


Fig. 2. Expert participant: short scan path with few fixations

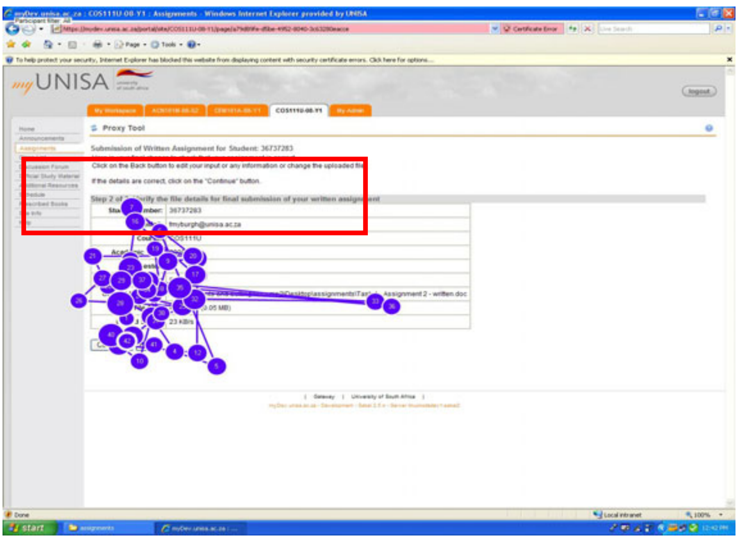


Fig. 3. Participant gaze plot: not reading instructions

4.2 Task 2

The participant had to submit a written assignment that included selecting a file in the correct (PDF) file format. The usability results showed a 70% completion rate but all participants made errors in not selecting the correct file format, with four participants needing assistance. Once the participants wanted to proceed with the task, without selecting the correct file format, the following error message was displayed: "ERROR: The type of file does NOT match the selected file type. (PDF!=DOC)".

All non-expert participants received this error at least twice. One participant received the error five times. The PDF file was opened, but the participant could not distinguish between a Word and PDF file. Consequently, this participant could not complete the task. One participant was given assistance after receiving the error four times. The assistance given was that the file is a PDF file and not a Word file. This did not help the participant, as the difference between PDF and Word was not understood. The participant shook her head in frustration during the task. Two participants submitted Word files instead of the PDF file after receiving the error twice.

Like in the first task, one non-expert participant entered his own name under file name again. After receiving the error six times, instructions were given as where to select the file format. The participant proceeded to guess the file format, by selecting each one available, until PDF was reached and accepted. This shows that the file format and PDF terminology is not understood. Three expert participants received the error message once only, correcting their mistake very quickly. One expert proceeded to submit the wrong file, a Word file. The last expert participant searched in Help and FAQ to find a solution.

The evidence from the eye tracking data supports the problem with understanding technological terms. Eye tracking data showed that all participants read the full error message; however only three participants (experts) continued to complete the task correctly after the first error message. None of the non-expert participants fixated on the word "File Format" or on the drop-down box arrow where the file format had to be selected. This indicates that the error message was not comprehended.

4.3 Task 3

Task 3 was the same as task one and would give an indication of the learnability of the Website. The usability results showed that eight of the participants completed the task faster than in task one. The average task time went from 68.67 to 34.70 seconds. Task completion was 100 %. This demonstrates that once a participant knows the basics of this Website, they use it without problems.

The expert participants had no problems with this task with task completion time ranging between 10 and 60 seconds. Expert participants made no errors and no assistance was needed. In this case, the two participants whose times did not improve were experts who double checked the files they were submitting.

All non-experts completed the task faster with task completion time ranging between 22 and 63 seconds. One non-expert participant struggled with the file type again, selecting PDF instead of Word. This participant learnt the process, but still struggled with the term *file type*, another non-expert struggled with the term *file name*.

Figure 4 demonstrates the same stage of the task as given in Task 1 for the same participant (a non-expert). Notice the improved shorter scan path and much less fixations. This demonstrates that the participants quickly learnt how to use this system. Research shows that a large number of fixations point to a less efficient search [17]. Once the participant knew what to do, the process was much faster.

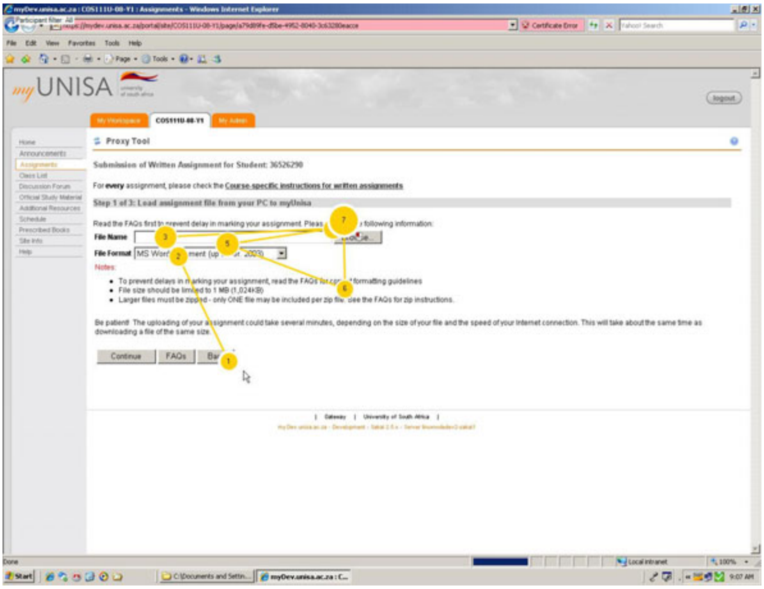


Fig. 4. Improved scan path and less fixations compared to same participant in Figure 1

4.4 Post-Test Questionnaire

All the participants were positive about the system despite the evidence that some of them found it difficult to use. Results showed that all questions had a median of 4 (5-point Lickert scale) or above. Only one question regarding error messages given by the Website had a median of 3. This discrepancy between their experience and response highlights the usefulness of eye tracking in identifying problems that would not have been reported. We can only speculate that this may be due to social desirability bias, quiescence or participants forgetting some of the difficulties they experienced. Participant comments in the post-test questionnaire included:

- *"Think it would be difficult for a person who rarely uses computers. They may have a problem understanding things like the file type."*
- *"It will be very useful if users could be shown how to use this Website. I had a friend who had to show me how to work it."*

These comments provide further evidence that it is necessary to distinguish between expert and non-expert users.

5 Discussion

The aim of this research was to examine the findings from the different usability methods namely, usability testing with observation, eye tracking and the post-test questionnaires. Looking at the findings for the three tasks evaluated, the eye tracking data supported the observations by providing concise, objective evidence of what was observed. Hence, the value of eye tracking lies in triangulating and corroborating the observations rather than in replacing the observations made on the user testing.

Different user's data can be overlaid to highlight specific problems. This summarised format is useful in providing the data to clients and managers who do not have the time to view all the recordings. In both cases the expertise of the observer or evaluator can influence the results but the scan paths and heat maps from eye tracking data provides a form of concise, summarised data that is difficult to match with other evaluation methods.

Eye tracking equipment is expensive and dedicated expertise is needed to use the equipment optimally as we have also experienced in our study. Therefore the overhead of cost and expertise needs to be carefully considered against the possible gains. Furthermore, our findings corroborate the fact that eye tracking is a tool to supplement usability evaluations and it does not replace standard usability testing with observations. This returns us to the questions of when the added value of eye tracking is justified.

The value of eye tracking, based on our findings is that it provides summarised, concise evidence. Furthermore, it is especially useful in capturing problems where users may be hesitant or unable to identify usability problems. For example, the disparity between the user-satisfaction reported in the post-test questionnaires, the user-satisfaction captured in the observations and the user-satisfaction captured in the eye tracking data. Despite the usability problems encountered, and captured by the objective measures, all the non-expert participants reported satisfaction with the system. In contrast the experts were more critical in the user satisfaction questionnaire while their efficiency and effectiveness usability measures rated higher. At this point we can only speculate about the reasons for the discrepancy between the objective and subjective results for non-experts. Data capturing errors such as the halo effect and social desirability bias could play a role since users have been found to make different judgments about the interfaces of scenarios for serious educational use [23]. This raises questions about the reliability of using subjective measures (post-test questionnaires in this study) for usability testing and emphasizes the importance of adding eye tracking data as objective evidence.

From a design perspective, error messages should be as short and simple as possible, when designing for non-expert users. From a practical perspective our study showed that participants had trouble with the most basic of terms e.g. *File type.* In designing a LMS it would be difficult to avoid using basic terms and therefore we recommend the use of a training video to provide a standardised, minimum body of knowledge before attempting any task. In the repeat task all non-experts' times improved; this means that the students learnt the system but they needed assistance in getting past the critical errors. From the design perspective, the negative effects of errors were compounded by the use of domain specific terms and the length of the error messages. It has been noted that non-experts provide valuable information about users' individual reactions and they are an important group in usability evaluations [7]. Our findings support the usefulness of non-expert input on the declarative (identifying problems with terminology) and procedural (identifying problems with instruction flow) knowledge levels.

6 Conclusion and Future Work

This study investigated the added value of eye tracking in usability testing of expert and non-expert Web users while doing LMS tasks. The usability and eye tracking data showed differences in terms of task time (considerably longer for non-experts), scan paths (longer for non-expert participants with more fixations scattered over the screen) and comprehension of terminology and error messages (non-experts had more difficulty understanding and kept repeating mistakes). Non-experts are known to have longer task times and scan paths than experts [4, 5], but what is surprising is the difficulty non-experts had in comprehending the most basic terminology and error messages to the point where it rendered the system unusable.

The added value of eye tracking was to pin-point the particular problem areas through users' scan paths, i.e., the problematic terms that caused many fixations as well as the instructions that were not read. Eye tracking data also provided visual evidence to back-up the observations made by the usability evaluator. In this study, where the findings from the post-test questionnaire and the usability testing observations were in conflict, eye tracking data provided evidence to allow resolution that the users were indeed experiencing usability problems.

We conclude that the additional cost of eye tracking is justified when working with users where usability problems may be purposely or unintentionally disguised. The number of participants used was satisfactory for a usability study, where five to eight participants are deemed sufficient [24, 25]. However, we plan to validate these findings further by having more participants completing a more comprehensive set of tasks. Future research will test these findings on different Learning Management Systems and also investigate the use of eye tracking with other user groups.

References

1. Avgeriou, P., Papasalouros, A., Retalis, S., Skordalakis, M.: Towards a Pattern Language for Learning Management Systems. Educational Technology & Society 6(2), 11–24 (2003)
2. Dillon, A., Song, M.: An empirical comparison of the usability for novice and expert searchers of a textual and a graphic interface to an art-resource database. Journal of Digital Information 1(1) (1997)
3. Aula, A., Majaranta, P., Räihä, K.-J.: Eye-Tracking Reveals the Personal Styles for Search Result Evaluation. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1058–1061. Springer, Heidelberg (2005)
4. Law, B., Atkins, M.S., Kirkpatrick, A.E., Lomax, A.J., Mackenzie, C.L.: Eye Gaze Patterns Differentiate Novice and Experts in a Virtual Laparoscopic Surgery Training Environment. Association for Computing Machinery, New York (2004)
5. Kasrskis, P., Stehwien, J., Hickox, J., Aretz, A.: Comparison of expert and novice scan behaviours during VFR flight. In: 11th International Symposium on Aviation Psychology Columbus. The Ohio State University, OH (2001)
6. Prumper, J., Frese, M., Zap, D., Brodeck, F.: Errors in computerized office work: Differences between novices and expert users. SIGCHI Bulletin 23(2), 63–66 (1991)
7. Äijö, R., Mantere, J.: Are non-expert usability evaluations valuable? In: 18th International Symposium on Human Factors in Telecommunications (HfT 2001), Bergen, Norway (2001)

8. Fields, B., Keith, S., Blandford, A.: Designing for Expert Information Finding Strategies. Technical Report: IDC-TR-2004-001 (January 2004)
9. Popovic, V.: Expert and Novice user differences and implications for product design and usability. Human Factors and Ergonomics Society Annual Meeting Proceedings 6(4), 933–936 (2007)
10. Gorman, M.E.: Types of Knowledge and Their Roles in Technology Transfer. Journal of Technology Transfer 27, 219–231 (2002)
11. Tabatabai, D., Luconi, F.: Expert-Novice Differences in Searching the WebAIS Electronic Library, AISeL (1998)
12. Kotze, P., Renaud, K., Van Biljon, J.: Don't do this - Pitfalls in using anti-patterns in teaching. Computers & Education (2006)
13. Beymer, D., Orton, P.Z., Russell, D.M.: An Eye Tracking Study of How Pictures Influence Online Reading. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 456–460. Springer, Heidelberg (2007)
14. Rudmann, D.S., McConkie, G.W., Zheng, X.S.: Eye tracking in Cognitive State Detection for HCI. In: ICMI 2003, Vancouver, British Columbia, Canada, November 5-7 (2003)
15. Yoneki, E.: Sentient Future Competition: Ambient Intelligence by Collaborative Eye Tracking. In: European Workshop on Wireless Sensor Networks (EWSN), Zurich, Switzerland (2006)
16. Bednarik, R., Tukiainen, M.: An eye-tracking methodology for characterizing program comprehension processes. In: ETRA 2006, San Diego, California, ACM, New York (2006), 1-59593-305-0/06/0003
17. Karn, K.S., Jacob, R.J.K.: Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In: The Mind's Eye, Cognitive and Applied Aspects of Eye Movement Research. Elsevier, Amsterdam (2003)
18. Jacob, R.J.K., Karn, K.S.: The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. In: Hyona, R.D. (ed.). Elsevier Science, Amsterdam (2003)
19. Sutcliffe, A., Namoune, A.: Investigating User Attention and Interest in Websites. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 88–101. Springer, Heidelberg (2007)
20. Ehmke, C., Wilson, S.: Identifying Web Usability Problems from Eye-Tracking Data. In: Proceedings of HCI: People and Computers XXI (2007)
21. Pretorius, M.C., Calitz, A.P., van Greunen, D.: The Added Value of Eye Tracking in the Usability Evaluation of a Network Management Tool. In: Proceedings of the 2005 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries. White River, South Africa (2005)
22. Vu, K.L., Hanley, G.L., Strybel, T.Z., Procto, R.W.: Metacognitive Processes in Human–Computer Interaction: Self-Assessments of Knowledge as Predictors of Computer Expertise. International Journal of Human–Computer Interaction 12(1), 43–71 (2000)
23. De Angeli, A., Sutcliffe, A., Hartmann, J.: Designing Interactive Systems archive. In: Proceedings of the 6th Conference on Designing Interactive Systems, pp. 271–280 (2006)
24. Barnum, C.: Usability Testing and Research. The Allyn and Bacon Series in Technical Communication (2002)
25. Nielsen, J.: Why You Only Need to Test With 5 Users. Alertbox 2000, `http://www.useit.com/alertbox/20000319.html` (cited 2009/01/12)

Supporting the Spreadsheet Idea for Interactive Database Applications

Mario Gleichmann[1], Thomas Hasart[1], Ilvio Bruder[1], and Peter Forbrig[2]

[1] IT Science Center Rügen gGmbH,
Circus 14,
D-18581 Putbus, Germany
Tel.: +49 38301 88290
{gleichmann,hasart, bruder}@it-science-center.de
[2] University of Rostock,
Albert-Einstein-Str. 21
D-18051 Rostock, Germany
Tel.: +49 381 498 7620
peter.forbrig@uni-rostock.de

Abstract. Data base applications allow the analysis of complex and large data. There are many analysis functions showing different relations between the data. End users have often new requirements to see data and relations which can not be shown by the existing analysis software. They need possibilities to create new user interfaces to fit their requirements. Generally, users don't have programming knowledge and cannot wait until the development department has specified the corresponding software. They need a tool which can easily and quick produce corresponding results. The tool must allow navigating via complex data structures of data bases.

This paper discusses a tool that allows end users to specify interactive applications like spreadsheets. The tool supports OLAP applications and is based on the Qt Designer.

Keywords: End user development, OLAP, Business Intelligence, Qt Designer.

1 Introduction

Our project on "Model-Driven Account Management in Data Warehouse Environments" (Monicca) aims at developing data base applications by end users. A tool for key account managers has been developed (key account management, see [9]). With this tool, such a manager can offer clients different views of aggregated data from a data warehouse via specific designed user interfaces. To generate the user views, a model language was developed which can describe necessary OLAP operations for the views, relations, the definition of the outputs and interactions.

This paper starts with a typical scenario of one of our industrial project partner. A part of a sales process is presented. It shows how an interactive application can be used to support business processes. For clarity, only a few exemplary data are presented.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 122–132, 2010.

The second part of the paper shows how end users can create the GUI without any programming knowledge. Afterwards, a short overview is given that discusses, which existing technologies were used and how these technologies had to be modified to support end user development on OLAP data.

Finally, some data from usability tests of the tool are presented.

2 Sales Talk Scenario

In this scenario, a salesperson of a bakery producer is negotiating with a customer on sales volumes. An essential part of the negotiations are the product prices and things like discounts. In the special case important aspects are the annual sales volumes of different types of bread.

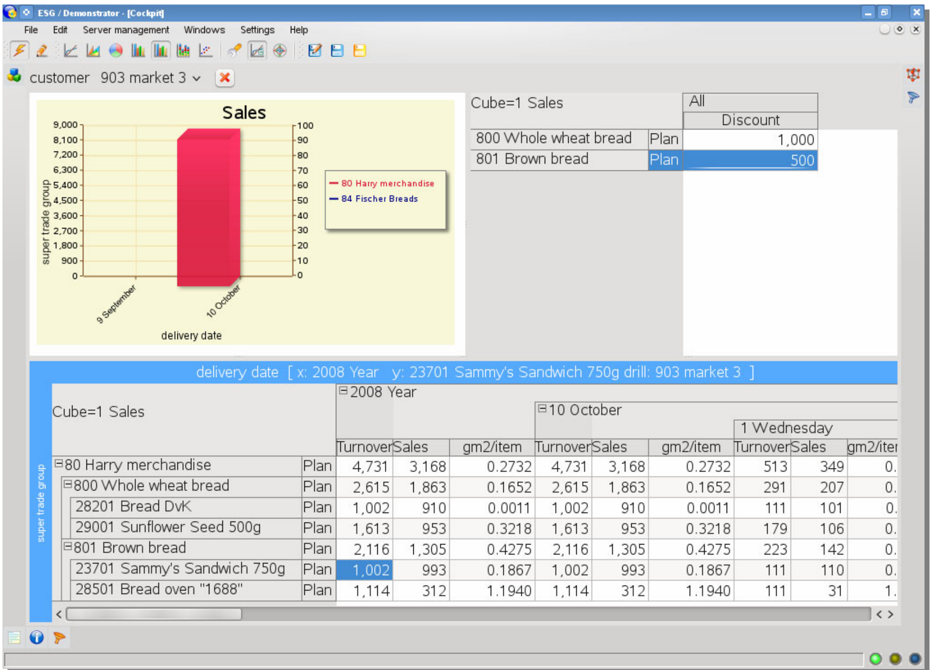


Fig. 1. Screenshot of a GUI giving an overview for planning

In Fig. 1., the data of the customer can be seen. In the groups, "Wholemeal" and "Wheat bread" that belong to the upper category of "Merchandise Harry", are the aggregate values of the "Turnover", "Sales" and "gm2 / item" (gross margin 2 per item) for the year 2008 and the October of that year. One can also see the aggregation of items into their group. The sales of the category "Whole wheat bread" are, e.g., the result of the sum of sales by its articles "Bread DvK" and "Sunflower Seed 500g". The same is true for the turnover.

delivery date

delivery date [x: 2008 Year y: 23701 Sammy's Sandwich 750g drill: 903 market 3]

Cube=1 Sales		⊟2008 Year							
					⊟10 October				
								1 Wednesday	
		Turnover	Sales	gm2/item	Turnover	Sales	gm2/item	Turnover	Sales
⊟80 Harry merchandise	Plan	4,731	3,168	0.1785	4,731	3,168	0.1785	513	349
⊟800 Whole wheat bread	Plan	2,615	1,863	0.1652	2,615	1,863	0.1652	291	207
28201 Bread DvK	Plan	1,002	910	0.0011	1,002	910	0.0011	111	101
29001 Sunflower Seed 500g	Plan	1,613	953	0.3218	1,613	953	0.3218	179	106
⊟801 Brown bread	Plan	2,116	1,305	0.1976	2,116	1,305	0.1976	223	142
23701 Sammy's Sandwich 750g	Plan	1,002	993	-0.0040	1,002	993	-0.0040	111	110
28501 Bread oven "1688"	Plan	1,114	312	0.8394	1,114	312	0.8394	111	31

Fig. 2. Negative gross margin in red (if not colorful darkest fields are red)

In contrast, the calculation of "gross margin 2 / item" is not as simple as the "Turnover". There are a variety of variable costs, such as discounts, rebates and discounts from the gross proceeds (gross revenues) deducted to get the net sales. The gross margin 2 is the difference of direct sales and marketing costs such as advertising subsidies and the like. In the upper right table, products and current discounts are shown. The top left graphics represents the sales date of the months September and October.

Assuming the customer wants a higher discount for "Brown bread", which is changed to 800 €. The system immediately performs all necessary calculations.

As shown in Fig. 2., this would cause a negative gross margin. Consequently, such a reduction would be unacceptable for the seller. To not refuse the higher rebate the manager may look for other compensations and changes the number of sales to 1.500.

delivery date (Sep 2008 - Oct 2008)

delivery date [x: 2008 Year y: 23701 Sammy's Sandwich 750g drill: 903 market 3]

Cube=1 Sales		⊟2008 Year						
					⊟10 October			
								1 W
		Turnover	Sales	gm2/item	Turnover	Sales	gm2/item	Turn
⊟80 Harry merchandise	Plan	4,731	3,675	0.1593	4,731	3,675	0.1593	
⊟800 Whole wheat bread	Plan	2,615	1,863	0.1652	2,615	1,863	0.1652	
28201 Bread DvK	Plan	1,002	910	0.0011	1,002	910	0.0011	
29001 Sunflower Seed 500g	Plan	1,613	953	0.3218	1,613	953	0.3218	
⊟801 Brown bread	Plan	2,116	1,812	0.1534	2,116	1,812	0.1534	
23701 Sammy's Sandwich 750g	Plan	1,002	1,500	0.0057	1,002	1,500	0.0057	
28501 Bread oven "1688"	Plan	1,114	312	0.8630	1,114	312	0.8630	

Fig. 3. Results of change the annual sales

This yields to a positive gross margin, which is even more than the old one. This is in spite of a higher discount (Fig. 3). This is a typical win-win situation that will contribute to the success of the sales talk.

In the following we will demonstrate how the discussed user interface can be designed using our tool.

3 End User Development

The result of a possible graphical user interface is illustrated in Fig. 1. With our tool it is possible to develop such an interactive system without programming knowledge.

This is as easy as the development of a spread sheet. There are a few steps necessary to perform this. First the tool has to be started. A blank widget will be created initially. Different visualization objects can be dropped on that widget. Fig. 4. shows the situation after an "ENodeSelectorWidget" (left) and an "ESumGrid" (right) were specified. (The visualisation objects are explained in detail in paragraph 3.1.)

On the right hand side the docked window "Cube Selection" can be seen, from which a cube can be selected and with drag and drop attached to the "ESumGrid". Afterwards specific date can be assigned to horizontal and vertical axis (Fig. 5.).

For other visualization objects like "ENodeSelectorWidget" other interactions are necessary. Each object needs to know to which cube it is assigned to and which details of the data have to be displayed.

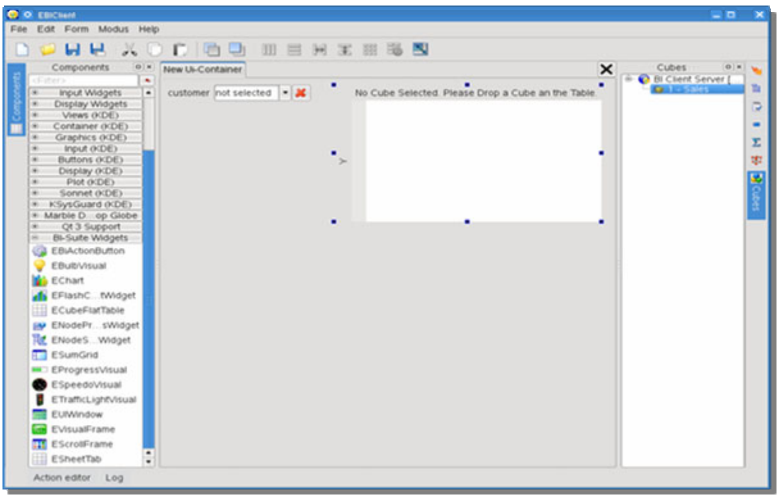


Fig. 4. "ENodeSelectorWidget" and "ESumGrid" dropped into the UI-Container

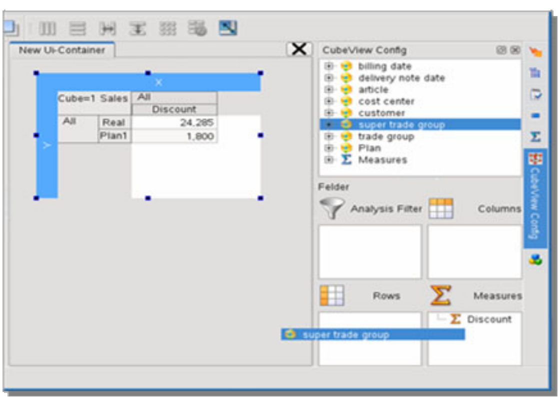


Fig. 5. Assignment of dimensions

In order specify connections between different objects in the user interface, their signals have to be linked. In this way changes in one object result in changes in the other objects.

Fig. 6. visualizes the dependency of "ESumGrid" from "ENodeSelectorWidget". The grid will be informed if another selection was made in "ENodeSelectorWidget". It will be updated and will display the corresponding data.

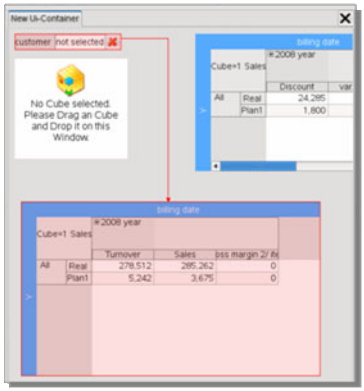


Fig. 6. Visualization of dependencies between object of the user interface

Finally, the individual widgets can be arranged by layouts. In this way they can respond to resizing. Therefore we select simultaneously the chart component and the upper right "ESumGrid" (Fig. 7.) and chose one of the different layout types. By default a horizontal layout is used that results in displaying all components with the same size. Thus the user could later change the sizes we use a horizontal splitter.

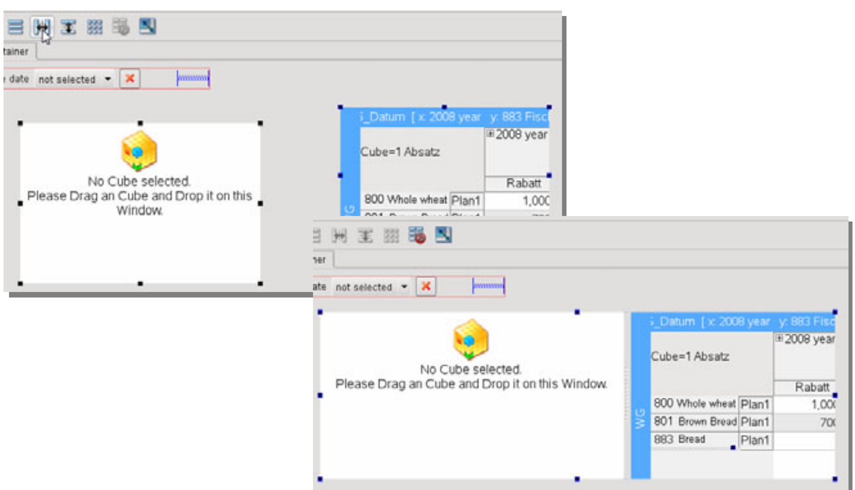


Fig. 7. Layout objects

As result we have a fully featured user interface which reacts on resizing and data changes. It can now be stored and can be used as a template for further developments.

We have demonstrated how our tool can be used to create graphical user interfaces. In this example specific visual objects were used. In the following we want to given an overview of existing objects and how they can be used.

3.1 Visualization Objects

The visualisation of data by different graphical object pays an important role for the usability of software. On the one hand it is very important to attract the attention of users on the other hand users should not pay too much attention to understand the visualisation of an application.

3.1.1 "ESumGrid" – OLAP Table

The "ESumGrid" (e.g. Fig. 3.) represents an OLAP table which can be used to display complex matters. In particular, this table is characterized by numerous and powerful navigation capabilities in multidimensional space.

Based on the example above on the x-axis is the dimension "billing date" and on the y-axis the dimension "super trade group" arranged. On the first start the aggregations of the chosen data, in our case "sales", "turnover" and "gross margin 2 / item", for the highest layers of the dimensions, "Year" and "super trade group" are shown. Now, if the column for the year is opened the months of the year appears and with it the values of data for the top category is shown. Because the column of the year remains visible it can be easily seen, that the addition of the total revenue columns for every month is the sum in the year column. Now, every month can be expanded again and releases the individual values of the days in that month.

Similarly, the y-axis can be expended. The super trade group releases by expanding the trade groups and the individual articles.

Fully expanded each individual sale of a certain article on a particular day can be found. It can also be analysed how it is part of the monthly and annual sum and of the trade and super trade group sum.

This type of analysis falls within the ambit of the "Online Analytical Processing", or short OLAP, and describes an entirely new approach to analyse economic data than the previous conventional approach of static snapshots of specific sizes [15].

3.1.2 "ETrafficLightVisual" - Traffic Lights

For easy monitoring of certain indicators thresholds can be defined. One way to visualize these are traffic lights. The thresholds can be shown on this object with certain colours. Red could be signalled that not enough pieces of an article are sold to justify the production or the purchase. Yellow could be used as a concern. Green visualises that there is no critical situation.

This traffic light object allows a quick overview of some specific data.

3.1.3 "ESpeedoVisual" – Tachometer

For the measures in relation to a particular end or maximum value tachometers are suitable.

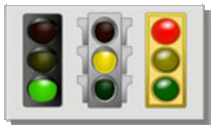


Fig. 8. Traffic lights in different styles

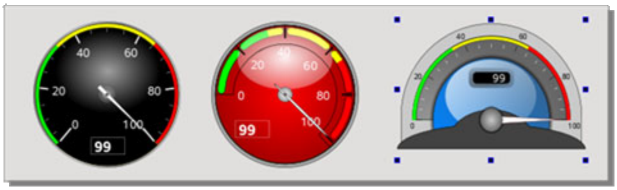


Fig. 9. Tachometers in different styles

3.1.4 "EFlashChartWidget" – Graph

As usual for chart components this widget allows the graphical representation of data. In addition to the shown bar charts (in Fig. 1.) are many other chart types possible (Fig. 10.). With this visualization object can show good temporal developments of certain data.

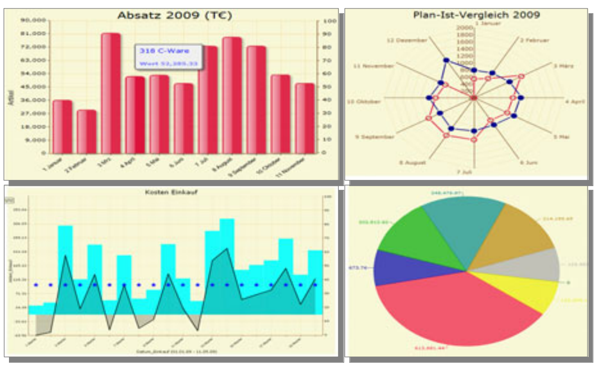


Fig. 10. Different chart types

3.1.5 "ENodeSelectorWidget" - Node Selection

This combo box allows selecting specific data in one dimension. If this object is used in conjunction with the "ESumGrid" it usually shows the data of a further (not in "ESumGrid" used) dimension. If the "ESumGrid" has the data of the dimensions "billing date" and "super trade group" the selector could have the data of the dimension "customer". With the selection of a certain customer the displayed data in the "ESumGrid" are restricted to that specific customer. There will be an aggregation of turnovers and sales of items that were sold to the selected customers only.

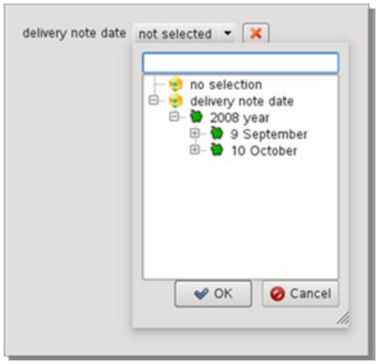


Fig. 11. A widget to select nodes

3.2 "EProgressVisual" - Progress Bar

This UI element is especially suitable for time-dependent quantities or values which achieve a fixed final value and are progressing.

Fig. 12. Progress bar

3.2.1 "EBIActionButton" - Command Button

For this item scripts for mouse actions can be defined. In this way, multiple analyses can be linked together or complete applications can be created.

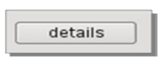


Fig. 13. An example for a button

Other functionalities are calling external programs, contact the properties or invoking actions on other elements of the current presentation.

3.2.2 "ESheet" - Spreadsheet

In these component formulas, pictures and charts can be combined in a flexible manner. The spreadsheet-like approach gives the user more freedom to change the appearance of the application. But for meaningful reports some knowledge is necessary to write the correct formulas.

We have seen how end user development for OLAP works. Special visual objects exist, which can be used for different aspects of analyses. The tool and the resulting application are easily and quick to use.

In the following the technology behind the tool will be shortly described.

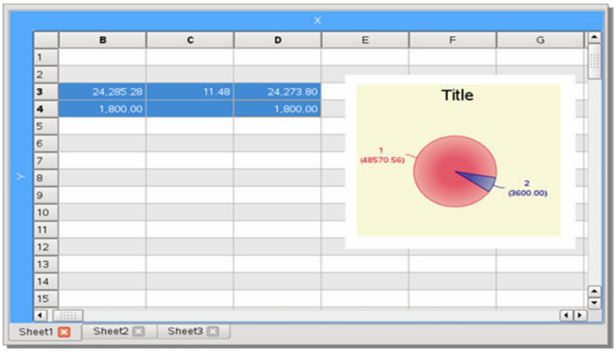


Fig. 14. A sheet with a chart

4 Technologies

To provide end user development we integrated an existing editor into our editor. We use the "Qt Designer", a part of the C++ framework "Qt" from "Nokia".

Usually this tool is used from C++ developer to arrange components to complex GUIs and save the settings in XML files.

We have extended it with dock widgets for the selection of data cubes, dimensions, and layers.

Additionally numerous plug-ins were developed which can be used as visualization objects.

As result a designer for end user development of OLAP applications exists. Like in a spreadsheet language [6] a user can specify the data that have to be displayed and he can use different visualization objects. The user can select a cube, the dimensions, layers, nodes and data. He can set relations between objects of the user interface. In some cases formulas can be specified to compute certain data.

5 Evaluation

To evaluate our tool we performed a usability test with 8 candidates. None of them were programmers. All of them work in a company as consultant. 4 of the candidates were already familiar with the idea of OLAP cubes and corresponding software. They already used the tool for adaptations in projects. The other 4 had to be introduced to the idea first.

All of them had to develop the user interface we discussed in this paper. First the scenario was explained and afterwards the candidates were asked to develop the presented user interface.

The ages ranged between 20 and 50. We observed that younger people were able to work with the tool better than older ones.

It was also recognized that it helped a lot if the candidates already knew the idea of OLAP cubes. More or less all candidates were able to specify the static user interface and to assign data to the tables.

50% of the candidates that already used our tool in projects had problems with the concepts of signals. They were not able to specify the dynamic behaviour of the user interface by specifying dependencies without help. None of the novices were able to perform this task.

In this respect training is necessary to be able to use our existing tool. As a result we have to think about a better user interface for this specific problem. Maybe it is not necessary to ask the user such detailed technical aspects like: "Which signal has to be propagated?" There might be better opportunities. We want to face this challenge in the future.

Our experiences with customers show some further interesting aspects. Even if we speak from end-user programming it is often not the end user who performs this work. The end user wants to have consultants that specify adaptations.

With our tool we are very close to the spreadsheet metaphor [3] for OLAP applications. But even for spreadsheets managers would like to perform the development task to be done by somebody else. For our tool the managers ask the consultants of our company to develop new user interfaces and to adapt existing user interfaces.

The good news is, even if we do not have reached end user programming, we reached programming by non-programmers. New user interfaces can be designed by consultants who are not involved in the development process of the tool and have no knowledge in programming. At the moment they are trained in specifying and we hope that by further improving our tool this training can be reduced. Maybe somewhere in the future no training will be necessary anymore.

6 Summary and Outlook

We have presented a tool for OLAP applications that allows non-programmers to develop own applications. In this paper, a scenario was discussed in which ad hoc reports on mass data in complex relationships are necessary. A GUI was described which makes these computations possible. It was shown how end-user can assemble themselves such a GUI using our tool that is based on the "Qt Designer" provided by "Nokia".

Additionally, some UI elements, developed by our group, were presented. These elements allow an adequate visualisation of specialized data.

In the future we work on a new interface based on the MDX standard replacing the current proprietary data connection interface for our visualization objects. This leads to a more flexible usage of the GUI on different OLAP servers.

Reasonable and reusable arrangements of GUI components developed by end users will be generalized and stored as patterns. These templates provide other users a faster GUI development for new OLAP applications. The variety of patterns will be stored in a database as a pattern library.

A first evaluation of our tool with different user groups was discussed and will support our next development steps. The results were promising. With further development of our tool the evaluation of the next prototype will be extended.

References

1. Ballinger, D., Bidle, R., Noble, K.: Spreadsheet Visualisation to Improve End-user Understanding. In: Australiean Symposium on Information Visualisation, Adelaide, Australia (2003)

2. Burnett, M., Cook, C., Pendes, O., Rothermel, G., Summet, J., Wallace, C.: End-User Software Engineering with Assertions in the Spreadsheet Paradigm. In: Proc. International Conference on Software Engineering, Portland, Oregon, USA, pp. 93–103 (2003)
3. Chitnis, S., Yennamani, M., Gupta, G.: ExSched: Solving Constraint Satisfaction Problems with the Spreadsheet Paradigm. In: CoRR. abs/cs/0701109, p. 1 (2007)
4. Dmitriev, S.: Language oriented programming: The next programming paradigm. JetBrains 'onBoard' Electronic Monthly Magazine (2004), `http://www.onboard.jetbrains.com/is1/articles/04/10/lop/`
5. Erwig, M., Abraham, R., Cooperstein, I., Kollmansberger, S.: Automatic Generation and Maintenance of Correct Spreadsheets. In: Proc. ICSE 2005, St. Louis, Missouri, USA, May 15-21, pp. 136–145 (2005)
6. Fisher II, M., Jin, D., Rothermel, G., Burnett, M.: Test Reuse in the Spreadsheet Paradigm. In: IEEE International Symposium on Software Reliability Engineering, p. 1 (2002)
7. Hodgins, J., Bruckman, A., Hemp, P., Ondrejka, C., Vinge, V.: The Potential of End-User Programmable Worlds: Present and Future. In: Panel SIGGRAPH 2007: ACM SIGGRAPH 2007 Panels (2007)
8. Ruthruff, J.R., Burnett, M.: Six challenges in supporting end-user debugging. ACM SIGSOFT Software Engineering Notes 30(4), 1–5 (2005)
9. McDonald, M., Rogers, B.: Key Account Management – Learning from supplier and customer perspectives. Butterworth Heinemann, Oxford (1998)
10. Meyer, R.M., Masterson, T.: Towards a better visual programming language: critiquing prograph's control structures. The Journal of Computing in Small Colleges 15(5), 181–193 (2000)
11. Millman, A.F., Wilson, K.J.: From Key Account Selling to Key Account Management. Journal of Marketing Practice: Applied Marketing Science 1(1), 9–21 (1995)
12. Mørch, A.I., Stevens, G., Won, M., Klann, M., Dittrich, Y., Wulf, V.: Component-Based Technologies for End-User Development. Communications of the ACM 47(9), 59–62 (2004)
13. Myers, B.A., Burnett, M.: End Users Creating Effective Software. In: CHI 2004 Special Interest Group, Vienna, Austria (2004)
14. Myers, B., Burnett, M.M., Wiedenbeck, S., Ko, A.J.: End User Software Engineering: CHI 2007 Special Interest Group Meeting. In: CHI 2007, San Jose, California, USA (2007)
15. Pendse, N.: What is OLAP? The OLAP Report (1998), `http://www.olapreport.com/fasmi.htm` (visited: 13.03.2008)
16. Scaffidi, C., Shaw, M., Myers, B.: An Approach for Categorizing End User Programmers to Guide Software Engineering Research. In: First Workshop on End User Software Engineering (WEUSE I), Saint Louis, Missouri, May 21 (2005)
17. Scaffidi, C.: A Data Model to Support End User Software Engineering. In: 29th International Conference on Software Engineering, ICSE 2007 Companion (2007)
18. Sidow, H.D.: Key Account Management. Landsberg am Lech: mi-Fachverlag (2007)
19. Nokia (2010), `http://qt.nokia.com/` (visited: 18.01.2010)
20. Winter, S.: Generierung von dynamischen Web-Anwendungen aus visuellen Spezifikationen. In: Schloss Birlinghoven - Sankt Augustin: Fachwissenschaftlicher Informatikkongress – Informatiktage 2005 (2005)

What Is Culture? Toward Common Understandings of Culture in HCI

Anette Löfstrom

ITC, Lägerhyddsvägen 2, Hus 1 75237 Uppsala
anette.lofstrom@it.uu.se

Abstract. What is culture in HCI? This article presents how the term culture has been used and discussed by authors in the discipline. Since culture is a complex concept, there is no aim to establish the term as a fixed definition in this paper. Instead the purpose is to enlighten some previous discussions and uses in order to ask questions, encourage new discussions and give motivation for searching common understandings of the term culture among HCI researchers. This article also includes a discussion of how such a change can be launched. Geert Hofstede's thoughts about culture are discussed out of a complementary perspective and a suggestion to use the dimensional thoughts on different cultural and societal levels is given. Notably, this paper does not answer questions. It asks them.

Keywords: Culture, HCI, Hofstede.

1 Introduction

The idea to this paper evolved during the Interact Conference in Uppsala, Sweden 2009. Throughout a two day long workshop "Culture and Technologies for Social Interaction" the participants discussed and elaborated on culture and technology for social interaction. Suggestions of what culture really is were conferred but there was great confusion, and common understandings were not even close to be established. Thus, in spite the number of cultural focused studies in HCI general interpretations still seems to be lacking. Therefore, the aim of this paper is to encourage discussions that focus on potential common understandings of culture in HCI. Further more, I argue for using eventual similarities and coherences between different scientific fields and on different analytical levels to increase cooperation. The contribution is questions that can provide fuel to those discussions. This is important since common understandings of culture would ensure similar intentions and interpretations among culturally focused researchers in HCI. Further more, suggested questions are based on previous HCI literature. Moreover, because the cultural thoughts of Geert Hofstede have had great impact on our field, this will be discussed among critical arguments against his cultural model. Adding this, a suggestion of using his thoughts about cultural dimensions on different societal and analytical levels will be presented. Choice of articles is due to their grade of inspiration for future discussions.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 133–141, 2010.

2 Previous Use of Culture in HCI Literature

2.1 Prior Overviews

When studying use of the term culture in HCI it is relevant to start with representing findings in previous literature overviews. Torkil Clemmensen and Kerstin Roese present a summary of publications in culture and HCI after 1998 in their article "An overview of a decade of Journal Publications about culture and Human Computer Interaction" [1]. This article follows an overview of earlier years, made by Pia Honold [2]. The purpose of Clemmensen´s and Roese´s paper is to review current practices related to how HCI cultural issues are studied and to analyze problems with the measures and interpretation of these studies. The authors present three important findings:

1. Most papers did not have any model of culture, merely used the word culture.
2. Most studies were quantitative studies with more than 20 participants.
3. In most of the studies, a major consideration in the choice of participants was that they could speak English.

Those findings inspire me to ask the following questions:

- How does the described way of using culture without reflection of the terms' immanent meaning affect potentially common understandings in HCI?
- How can we work in order to link together quantitative uses of culture, for example nationalistic grounded assumptions in large studies, with qualitative methods like deep interpretations of a specific cultural loaded phenomenon?

I am aware that the goal related to the second question above might be difficult to achieve. Still I think it is valuable to at least strive for, because it facilitates increasing cooperation between qualitative use of culture (deep interpretations of specific phenomena) versus quantitative uses of culture (studies including lots of people and with general assumptions). For example, if we find one frequent interpretation of a specific cultural loaded symbol on a basic level in different areas of HCI, maybe we can accept it as general? If so, this interpretation can be implemented on higher quantitative levels. This would ensure qualitative grounds in quantitative research and the other way around, qualitative focused researchers could strengthen their result by relating it to quantitative studies build on the same deep understandings as their own. Still this process should begin by searching common meanings in specific cultural loaded phenomena. Therefore, I suggest that linking qualitative interpretations with quantitative uses of culture should start in small scale and there after expand. These thoughts are similar to the ones discussed in an article presented by Singorini, Wiesemes and Murphy [3]. They advocate approaching culture by starting with examining micro- cultures in order to develop small models which can gradually be expanded in to larger models of culture. Their suggestion is a reaction against the national focused, criticized but often used cultural thoughts of Geert Hofstede.

2.2 Geert Hofstede, Critique and Possible Use

Singorini, Wiesemes and Murphy describe the five cultural dimensions used as analytical tools in Hofstede's work. They also discuss why they regard the dimensions as problematic and their opinion is quite extensive. One critique is that Hofstede overstates the effects of individualistic versus collectivistic nations and, following this, neglects other explanations for educational differences such as socio-economic factors, funding for educations etc. Moreover, the authors conclude that Hofstede uses the term in-groups without defining what it is; that he regards culture as something static; that differences more than similarities is focused and that his model can't be related to cultural complexity. However, in spite of the critique his model is frequently used. Myers and Tan listed 36 studies and 24 of those used Hofstede's dimensional thoughts [4].

Rachel F. Baskerville is another researcher offering critique against Hofstede. In an article called: "Hofstede never studied culture" [5] she identifies the following problems:

1. An assumption of equating nation with culture
2. The difficulties of, and limitations on, a quantification of culture represented by cultural dimensions and matrices
3. The status of the observer outside the culture

Hofstede made a reply on Baskerville's critique in a paper called: "What is culture? A reply to Baskerville" [6]. Here he claims that Baskerville did not read his work carefully and that she replicated parts of his work that failed to produce the same differences between countries. He also writes that his work has rarely been cited by sociologists and anthropologists who, he writes, are the ones that know about culture. At Baskervilles questioning if dimension indices that correlate with the result of other cross-national studies really are cultural dimensions he writes that whatever the dimensions are called they are the ones accounting research is interested in. He ends this argument with the following words:

"Most accounting researchers couldn't care less about what anthropologists call what- as little, by the way, as anthropologists care about accounting research" [6].

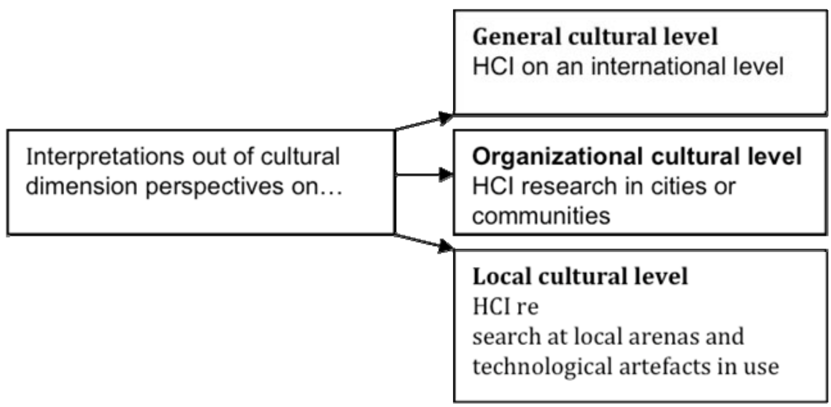
This quote motivates one of the meanings of this paper; to search for similarities and coherences between different scientific fields. I argue that searching cooperation's and mutual understandings between different research fields would increase scientific development. This leads to the following questions:

1. Is it possible that researchers who "couldn't care less" what researcher from other fields call what can meet and enrich each other scientifically? If so, how?
2. What would be the effects of such cooperation's?

However; in spite her critical standpoints Baskerville writes that: "It is not the wide application of Hofstede's framework that makes it good. Rather it is usefulness of the theories as mechanism of describing and understanding cultural diversity that makes them interesting and viable" [5]. Baskerville continues by claiming the need for developed theories and that Hofstede's work can facilitate this. I agree with this conclusion. Still I do not want to take a position in the debate "for or against" Hofstede since my personal scientific view is that quantitative and qualitative methods and their related theories are complementary and equally important. In quantitative cultural related studies it might be hard to avoid what Hofstede has been criticized for (static cultural view, national or general focused cultural assumptions etc.) Still such examinations are valuable since they provide knowledge on a general societal level. At the same time it is essential to analyze culture as dynamic and multi-facetted. Out of this reasoning and with Baskerville's claim of the value of providing theories as ground I suggest using Hofstede's cultural dimensions on different levels. However, I do not want my theoretical suggestion to be regarded as a way of "choosing side", because it is not. Rather it is a way of testing my view that quantitative versus qualitative focused scientific fields is mutually dependent and that they could enrich each other. Why not search for theoretical similarities and possibilities? The questions below could serve this purpose:

- Can Hofstede's thoughts of cultural dimensions be used as analytical tools on different levels and thereby work as glue between deep and general levels?
- How would such use of Hofstede's dimensions affect common understandings of culture in HCI?

I demonstrate my thoughts in the model below:



My view is that if interpretations out of the five dimensions on all three levels above reveal similarities, those could be used as one way to find common understandings of culture in HCI since the result would be usable for both quantitative and qualitative focused researchers.

After this discussion I will continue by discussing some examples of how culture has been used in HCI.

Examples of culture use in HCI

When looking back at previous studies of culture, it seems that the cultural perspective in HCI has increased. M., Kamppuri, R., Bednarik, M.,Tukiainen claim that articles in our field considering national/ethnic culture were rare in the first half of the 1990´s. They also write that:

Even though the total number of HCI articles has (…) grown during the later half, it is only about 1.2 times greater compared to the earlier half, whereas the number of culture related articles is six times greater compared to the earlier half [7].

Hence, it seems that the general interest of culture in the field has enlarged. Still there is a difference in the author's presentation above that needs to be noticed. In the first half of the 1990: s the authors speak about national/ethnic culture whereas they do not use this delimitation in the quote about the latter half of the 1990: s. Here they say "culture related articles" rendering possible for alternative interpretations of what culture really is. This could serve as an example of the immanent complexity in the term. It can also be used as a foundation for discussion, searching for an answer of the question:

- How can HCI researchers identify and avoid pitfalls that emanates from cultural complexity?

So far we have seen that many studies using cultural perspectives are quantitative. One example of this is a study of Ravi Vatrapu and Dan Suthers [8] who wrote an article where they aimed for a systematic examination of phenomena in the nexus of culture, technology and learning. Their assumption was that social affordances of technology might vary along cultural dimensions. They describe culture as patterns or schemas even if phenomena like symbols, rituals, norms, values and practices cover the most visible manifestation of culture. Those authors describe culture as something shared and as factors in individuals' social interactions. Culture is used as cognitive schemas that are formed from an individual's biography which includes interactive effects of geography of that individual's upbringing and the formative experiences of his/hers life. They argue that cultural models can be used to identify differences in culture that affect computer supported collaborative learning environments and they group the concept of culture in four of Hofstede's dimensions [8]. As shown Vatrapu and Suthers discuss cultural schemas (quantitative methods) while still including basic cultural issues such as rituals and symbols. This can serve as an example that linking quantitative uses with qualitative interpretations might be reachable. Still I want to point out the essential difference between this article and my scientific view. Vatrapu and Suthers argue that cultural models can be used to identify differences while I suggest using them to find similarities. Questions emanating out of this reasoning are:

- Would it affect common understandings of culture if we seek similarities instead of differences?
- What difficulties are we likely to face when analyzing basic phenomena, like for example effects of norms on technology use, and relate it to analysis on general levels?
- What are the potential benefits of such an approach?

Researchers claiming that approaches which take culture explicitly into account may add valuable contributions to the design and evaluation of effective online communities are Clarissa Barbosa, Carla Leitao and Clarisse de Souza [9]. They show how HCI researchers handle cultural dimensions in online communities in different ways. Focusing Brazilian people's way of living in informal social relationships and by observing that this people prefer to get information by exchanging messages with each other rather than searching it on the web these researchers studied how informality in daily life affected Brazilians online community activities.

My interpretation of this is that in the study above culture is used practically by studying daily life habits and transforming theme into online activities. The cultural perspective used is ethnicity (Brazilian). A common perspective among ethnologists is that ethnicity is based on identity, both by people belonging to a group to be an "us" and by other to group those people to "the others". Another ordinary view is that identifications phenomena such as loaded symbols, things and stories are used to shape a sense of fellowship with other members of the group. I suggest that this is important to regard when following the authors claims that: "cultural characteristics may be crucially important requirement for developing effective technology" [9]. They also emphasize that HCI professionals should: "investigate the cultural characteristics of communities' one design for, and comprehend their impacts both on the technological design and the subsequent success of online communities" [9]. I suggest that what these authors call cultural characteristic might be manifested in basic phenomena like symbols, heroes, rituals and values. Further more, in HCI I argue for regarding technological devices as cultural arenas as well since most technologies are designed with cultural loaded phenomena in the interfaces. Therefore, my next thought in this paper is that cultural focused researchers in our field might discover new cultural phenomena or even categories that are based in the technology. This reasoning leads to the questions:

- Are there any new cultural phenomena or categories in the technology, if we regard it as a cultural arena in it self? If so, how do we connect those categories with real life?
- Are such eventual phenomena or categories beneficial or do they make the concept of culture even more complex?

My view is that even if culture constitutes a central framework in the article above it is still difficult for a designer to know which identification is the main factor in the actual situations and tasks of the user. Therefore, I agree with the authors when they say that "shifting focus from capturing to understanding culture" [9] is important. This leads to the following questions:

- How do we know which cultural phenomena in real life are most essential to address for designers?
- What tools do we need for valuing the importance of different cultural elements for different kinds of designs?

However, if we manage to understand culture in a similar way in the future we face the next obstacle.

- How do we implement new cultural knowledge in design processes?

Ashwini Asokan writes that: "CHI community is still sparse on conversations and significance of world cultures on design and HCI" [10]. Despite the enlightened delimitation the author gives the following valuable description of culture as:

a lifestyle, a set of beliefs and value systems that shape every day life in countries around the world, is at the core of understanding what our community calls the `user´. It is in the context that explains all data and inspiration we use in our design and creative process [10].

Further more, Asokan's regards culture as highly context dependent. This assumption is important since every single person lives in different contexts and is affected of them all which mean that even if the specific context in the precise user situation is of extremely importance reality is much more complex than that when working with people and their interaction with technology. Let us suppose that the user had a bad morning with stress, sick children, broken car and delayed bus. It is likely that those circumstances lead to variations in contexts effects on the user and in the next step the user's perception and use of the technological device. This arguing leads us to another essential perspective. Among ethnologists humans are regarded and interpreted as "cultural beings" that both affects and is affected of tremendously many factors. Therefore I argue that the suggested contextual understanding of culture is extremely important but it should be complemented with the equally important view of humans as cultural beings since the "cultural human" affects technology use. In her conference contribution Asokan asks the question: "At an organizational and management level, what do we require to ensure integration of culture into our practice" [10]. Maybe interpreting users as cultural beings on a deep analytical level and relate the result to general (for example organizational and management) levels is one way to do this? However, a central perspective here is that Asokan understands culture as context dependent which leads to the following questions:

- What divides culture from context in HCI?
- If culture is context dependent, is context dependent of culture as well?
- How do these eventual dependencies affect technology designs and uses?

In their article Culturability: The Merging of Culture and Usability Wendy Barber and Albert Badre uses the term "cultural context". They write that: "As a consequence of existing international WWW users (…), usability takes on immediate and relevant cultural context" [11]. In this study the authors identifies what they call localisation elements and generalize them into "cultural markers" that are specific to a given culture and/or perhaps influenced by genre. They define cultural markers as design elements found in web pages and which prove highly prevalent within a particular cultural group and less prevalent or absent in another group. Usability issues are central in their study and cultural preferences and biases such as colours and spatial orientation among others impact what is deemed as user friendly [11].

Barber and Badre provide the reader with an explicit description of how they use the term culture. They write that: "We use the word "culture"- somewhat loosely- as a means of distinguish among the different countries and their respective websites" [11]. They use the term out of an ethnic perspective but with a consciousness of the terms complexity since they say that: "our use of the term is not intended to be inductive of all the nuances and properties frequently applied by the term" [11]. So, those authors make clear what culture is in this specific study. They want to permit a discourse that distinguishes one country or region of the world from another. In this case culture is used as boarder markers such as national symbols, colour or spatial organizations that are preferred by specific groups of users. Relating this study to Hofstede's cultural thoughts and my own suggestion to interpret culture in a similar way on different levels I ask the following question:

- How would cultural similarities on different levels affect designs?

Adding this I want to point out that what is obvious but implicit in this study is the great importance of symbols. The authors search "cultural markers" but if those were not loaded with symbolic meanings they would be useless in a cultural conscious design process, because they would not send a message to the user. They would just be there. For example, when designers use flags for choice of language. If the flags were not loaded with symbolic meaning the user would not know what flag to click for a certain language Therefore the article above is an example of research using cultural phenomena on a basic level (symbols) to distinguish cultures on a general level (nationality). However, there is a crucial difference between my theoretical suggestion in this paper and Barber's and Badre's article. They use basic symbols for distinguish on the general level while my suggestion is not to distinguish but instead search similarities and coherences.

3 Discussions

In this paper I have searched in HCI literature for presentations grounding questions about what culture really is in our research area. One common use of culture I have found is that the authors discuss phenomena typical regarded as culture such as symbols, values, rituals and norms. This inspires me to ask:

- What would be the result if designers used typical cultural phenomena as analytical tools for understanding technological devices in the same way that he or she interprets the user context in the iterations of the design process?

Ethnicity and nationality seems to be an often used cultural perspective. Still I suggest that HCI researchers should be open for other equally significant perspectives on general levels. Therefore I want to ask the following questions:

- How can we evaluate which characteristics on different cultural levels have the greatest significance when designing for a specific situation, task, context and user?
- How do we know what cultural issues are most important if the same technological device is to be used in different cultural contexts and by users with different cultural backgrounds and experiences?

Further more, in this paper I argue for striving towards common understandings of culture in HCI. Still, some readers of this paper might not agree with the need of such commonality. If so I suggest the following question.

- Are there any disadvantages with common understandings of culture in HCI? Is it worth striving for? Why or why not?

Finally, my personal point of view is that common understandings of culture definitely are worth striving for. I argue that if every author has at least some common understandings which are relevant for HCI it would ensure similar intentions when using the term culture and corresponding reader interpretations independent of the researcher's different educational backgrounds and choice of methods and theories. We would know what elements must be included when a conference topic contains "cultural issues". My suggestion to use Hofstede's cultural dimensions in different levels is one way to facilitate this. Beside these arguments there can be more benefits that my discussion in this article does not reveal. Therefore the last question I suggest is:

- What are the benefits of shaping common understandings of culture in HCI?

4 Conclusions and Future Work

In this paper I have raised questions to bring into the discussion about the concept of culture in the HCI discourse. I have also argued for cooperation's between different scientific fields in our research area.

This conference contribution shall be regarded as an inspiration for future work and as the beginning of an iterative process similar to those that are usual for technology development in HCI. The difference is of course that in this work the iterations occur between researchers developing a general "usability" of a complex but often used term. Therefore I plan to utilize the discussions this article grounds as empiric material in an iterative process, that is to say, to continue writing articles building on HCI researcher's understandings and uses of culture. Beside this, workshops and tutorials on conferences can be innovative for this purpose as well. Finally, this future studies and discussions of culture will be linked to my own research that is an examination of web based cooperation's in the City of Stockholm. My focus is on leader strategies and employees driving forces to use the web for communicating with leaders. One important question in this research is how culture affects such work.

References

1. Clemmensen, T., Roese, K.: An overview of a decade of Journal Publications about culture and Human Computer Interaction. Workingpaper nr. 03. Copenhagen Business School (2009)
2. Honold, P.: Learning how to use a cellular phone: Comparison between German and Chinese users. Technical Communication 46(2), 196–205 (1999)
3. Signorini, P., Wiesemes, R., Murphy, R.: Developing alternative frameworks for exploring intercultural learning: a critique of Hofstede's cultural difference model. Teaching in Higher Education 14(3), 253–264 (2009)
4. Myers, M.D., Tan, F.B.: Beyond Models of National Culture in Information System Research. In: Advanced Topics. Information Management, vol. 2 (2003)
5. Baskerville, R.F.: Hofstede never studied culture. Accounting, Organizations and Society 28, 1–14 (2003)
6. Hofstede, G.: What is culture? A reply to Baskerville. Accounting, Organizations and Society 28, 811–813 (2003)
7. Kamppuri, M., Bednarik, R., Tukiainen, M.: The expanding Focus of HCI: Case Culture. In: NordiCHI (2006)
8. Vatrapu, R., Suthers, D.: Culture and Computers: A review of the Concept of Culture and Implications for Intercultural Collaborative Online Learning (2007), `http://www.springerlink.com/ravivatrapu`
9. de A., Barbosa, C.M., Leitao, C.F., de Souza, C.S.: Why Understanding Culture is Important for Developing Effective Online Health Support: the Brazilian Context. In: International Conference on Human-Computing Interaction (2005)
10. Asokan, A.: Culture Calling: Where is HCI? In: CHI Proceedings (2008)
11. Barber, W., Badre, A.: Culturability: The Merging of Culture and Usability. In: 4th Conference on Human Factors and the Web (1998)

Personalized Support, Guidance, and Feedback by Embedded Assessment and Reasoning: What We Can Learn from Educational Computer Games

Michael D. Kickmeier-Rust and Dietrich Albert

Department of Psychology, University of Graz
Brueckenkopfgasse 1, 8020 Graz, Austria
{michael.kickmeier,dietrich.albert}@uni-graz.at

Abstract. Software that intelligently interprets the goals and needs of its users on the basis of their behaviors without interrupting the work flow and consequently disturbing concentration and software that can support the users in a personalized, smart, yet unostentatious way is a desirable vision, for sure. One attempt to such support system was Microsoft's famous paperclip. The underlying logic, unfortunately, was rather simple and the users did not accept the feature very well. This paper introduces a psychologically and formally sound approach to a non-invasive, hidden assessment of very specific needs of the users as well as their competencies and corresponding tailored support and feedback. The approach was developed in the context of adaptive digital educational games and is based on the concepts of Competence-based Knowledge Space Theory as well as that of Problem Spaces. The purpose of this paper is to broaden the concept and elucidate a possible bridge from computer games to regular software tools.

Keywords: Embedded Assessment, Micro Adaptation, Support Methods, Feedback, User Model.

1 Introduction

An important trend in the area of learning technologies is game-based learning. Computer games are a tremendously successful, popular, and attractive genre. A substantial number of young people spend many hours a week playing computer games, and these games are often the preferred play. Thus, using the motivational potential of computer games for educational purposes may open new horizons for educational technology and is challenging educators and developers. The very nature of utilising (computer) games for education is that playing games is one of the most natural forms of learning. Children learn to talk by playing with sounds and learn collaboration and strategic thinking when playing role playing games such as *Cowboys and Indians*. Immersive digital educational games (DEG) fall back to that origin and are a highly promising approach that makes learning more engaging, satisfying, inspiring, and maybe even more effective (see [1]). The major strengths of DEGs are generally observed at a high level of intrinsic motivation to play and proceed in the game, a

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 142–151, 2010.

meaningful yet rich and appealing learning context, immediate feedback, and a high level of interactivity, challenge, and competition. Some researchers even argue that the exposure to "twitch speed" computer games, MTV, and the Internet has altered cognitive processes, emphasising specific cognitive aspects while de-emphasising others [2]. Thus, the so-called "digital natives" may require different, possibly non-conventional, educational approaches. Current challenges in the area of DEGs are seen in, for example finding an appropriate balance between gaming and learning activities (e.g., [3]), finding an appropriate balance between challenges through the game and abilities of the learner (e.g., [4]), convincingly embedding educational objectives in a game scenario, particularly when declarative knowledge is concerned (cf. [5]), or managing the extensive costs of developing high quality games (e.g., [3]).

In closer examination, many of those challenges are very similar to the challenges of interaction principles and interaction design with "regular" software". The digital natives, Marc Prensky [2] was talking about in the context of DEGs, are also users of regular software products (office tools, web browsers, websites, communication tools, etc.) and demand very specific – maybe unusual – features. Moreover, we are facing a large group of users who are highly inexperienced and sceptical towards the use of computer devices of all kinds, for example the elder generation. Examples of requirements might be balanced and intuitive multi-tasking possibilities, an embedded and personalized support with complex functionalities, integrating various tools for various purposes in common and usable devices, or managing high development and support costs.

In the focus of the present paper is to address a major challenge for research in the context of translating ideas from game-based learning to the usability of regular software products. In DEGs it is all about an individual (learning) experience in order to reach educational effectiveness and maintain fun, immersion, flow experience[1], and the motivation to play and, therefore, to learn. Thus, meeting individual preferences, interests, abilities, and goals is key to successful game-based learning. Such claim, of course, is also a highly desirable asset of regular software. In the following sections we will present a formal cognitive approach for an non-invasive, embedded assessment of psycho-pedagogical aspects such s motivational state or learning progress and corresponding personalized support by subtle feedback and support, which is currently implemented in learning adventure games.

2 Intelligent Adaptation and Personalization in DEGs

To meet the aforementioned claims, DEGs are supposed autonomously adapt to the learner along a variety of axes, for example, prior knowledge, learning progress, motivational states, gaming preferences, and psycho-pedagogical implications. Generally speaking, adaptive approaches to technology-enhanced learning contest the one-size-fits-all approach of traditional learning environments and the attempt to tailor the learning environment according to individual needs and preferences. The spectrum of

[1] According to Mihaly Csikszentmihalyi [6], flow refers to a highly immersed experience when a person is engaged in a mental and/or physical activity to a level where this person loses track of time and the outside world and when performance in this activity is optimal.

approaches, methods, frameworks, and applications is quite broad [7] there are basically three major concepts:

- *Adaptive presentation*: adjusting the look and feel of a learning environment according to individual preferences or needs; for example, different colour schemes, layouts, or amount of functionality;
- *Adaptive curriculum sequencing:* providing the learner with learning tailored to individual preferences, goals, learning styles, or prior knowledge;
- *Adaptive problem solving support*: providing the learning with feedback, hints, or solutions in the course of problem solving processes.

Most of these methods and frameworks for adaptation and personalisation were developed in the context of conventional e-learning. The underlying concepts and ideas are currently extended and adjusted to the requirements of the rich virtual gaming worlds, particularly to maintain an immersive gaming experience and high levels of motivation, curiosity, and flow experience [8].

A method, which is highly interesting for interaction design in general, is an approach to non-invasive assessment of knowledge and learning progress in the open virtual worlds of computer games and a corresponding adaptation by personalised psycho-pedagogical interventions. The approach, labelled *micro adaptivity*, was developed in the context of *80Days* (www.eightydays.eu), a multidisciplinary research and development project. The project had the ambitious goal of utilising the advantages of computer games and their design fundamentals for educational purposes and addressing specific disadvantages of game-based learning. Within the project, a methodology for the successful design of educational games was established, and a game demonstrator was developed based on a state-of-the-art 3D adventure game (see Fig. 3 for some impressions of the game).

3 Adaptation on the Micro Level

A primary task for (game-based) adaptive mechanisms is to guide and support the user in accomplishing specific goals, for example, informing the user, intervening when misconceptions or errors occur or the work/learning progress is unsatisfactory, hinting, or providing the user with appropriate feedback. In addition, tasks are motivating, maintaining immersion, and personalising the environment according to the preferences and needs of the user. Accomplishing this goal requires a theoretical and technological approach that enables the system to assess cognitive states (e.g., *competence states* or *motivational states*), (learning) progress, possible misconceptions, or undirected/unsuccessful problem solving strategies. A distinct characteristic of adaptive DEGs is that gathering the necessary information from the user cannot occur in a conventional form (e.g., by test items, questions, or tasks). A DEG requires an assessment that does not harm motivation, immersion, flow experience, or the game's storyline [8]. The very basis of micro adaptivity is a formal psychological model for interpreting the behaviour within the virtual environment.

To realise the vision of a non-invasive support of users in DEGs (i.e., embedded assessment and subsequent educational interventions), we combine the *Competence-based Knowledge Space Theory* (CbKST), which is a formal set-theoretic psychological approach to human learning and development that has been successfully utilised in conventional adaptive e-learning, with cognitive theories of problem solving. This theory provides a detailed domain model that includes a set of meaningful *competence states* as well as a set of possible *learning paths*. Problem solving theories, in turn, provide a set of possible *problem solving states* as well as possible problem solving paths. While in an DEG (at least in most types of DEGs) complex problem solving is an important mechanisms of self-directed, constructivist learning, for regular software we can assume that the user's goal is to accomplish a specific task or to solve a specific problem that, in turn, requires performing a sequence of specific action and interactions with the software. To give a very simple example, the task might be to highlight a certain word of a text in an office software. The steps to solve this problem are (i) taking the mouse and highlighting to word by setting the cursor at the beginning or the end of the word, pushing the left mouse button and by moving the mouse with pressed mouse button along the word and , finally, by releasing the mouse button. When the word is selected, (ii) the user has to press the correct highlighting symbol in the tool bar with the left mouse button.

3.1 Competence-Based Knowledge Space Theory

In essence, CbKST [9, 10] originates from the *Knowledge Space Theory* (KST) established by Jean-Paul Doignon and Jean-Claude Falmagne [11, 12], which is a well-elaborated formal cognitive framework for covering the relationships between problems (e.g., tasks or test items). It provides a basis for structuring a domain of knowledge and for representing the knowledge based on *prerequisite relations* (Fig. 1, left panel), in the sense that one task is easier than another or that it is likely mastered before another. While KST focuses only on performance/behaviour (e.g., solving a test item), CbKST introduces a separation of observable performance and latent, unobservable competencies, which determine the performance. Essentially, CbKST assumes a finite set of competencies and a prerequisite relation between those competencies. As mentioned, a prerequisite relation states that a competency *a* (e.g., multiplying two positive integers) is a prerequisite to acquiring competency *b* (e.g., dividing two positive integers); if a person has *b*, we can assume that the person also has *a*. Due to the prerequisite relations between the competencies, not all subsets of competencies are possible *competence states* (e.g., it is highly unlikely that a person can multiply but not add numbers). The collection of possible competence states corresponding to a prerequisite relation is called *competence structure* (Fig. 1, right panel). While in the learning/development context we likely find logical prerequisites (e.g., to add two integers is a logical prerequisite of multiplying two integers), in the context of using software tools the prerequisite are likely not so obvious (e.g., to know how to change the font of a text can be considered a prerequisite for knowing how to make a table). Thus far, the structural model focuses on latent, unobservable competencies. By utilising mapping functions, the latent competencies are assigned to a set of tasks/test items/actions relevant for a given domain, which induces a *performance structure*, the collection of all possible *performance states*. Learning or development is not seen as a linear course, equal for all learners; rather, learning follows one of a set of individual learning paths.

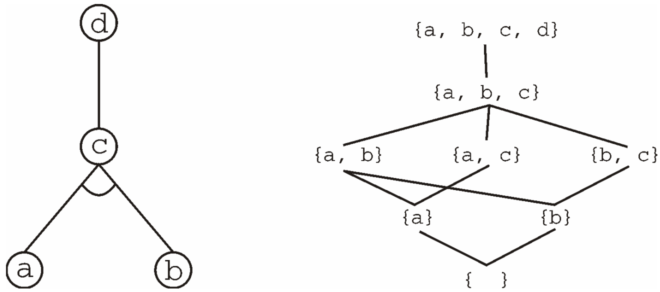


Fig. 1. The left panel shows a prototypical prerequisite function; the right panel illustrates the corresponding competence structure

Translated to the context of accomplishing tasks with a software tool, this approach enables to define the competencies necessary to use the tool in a very detailed and precise way. By the so-called competence states, the collection of competencies a specific user has, we can determine which task s/he is able to accomplish and what kind of information and support is required to solve other tasks. In addition to the structural hypothesis lying the competence structures, we need an equally formal idea of problem solving as a process. In this work, we rely on a formal theory of the human problem solving process, the theory of Newell and Simon ([13]; see also [14] for a conclusive overview).

3.2 A Theory of Human Problem Solving

[13] argued that problem solving is dynamic, rule driven process. The actions a person performs can be seen as a function of memory operations, control processes, and rules. The very basis of this approach is to decompose a problem or situation (you may think about all possible states of the *Tower of Hanoi* problem) into a *problem space*, a collection of all possible and meaningful *problem solution states*, objects relevant for a problem, and transition rules, which specify how admissible transitions from one to another problem solving state can occur. In so far, the approach of Newell and Simon is highly compatible with the structural approach to learning and development of CbKST. Consequently, the Tower of Hanoi's *problem space* would include all states where the rules "a larger disk cannot top a smaller" and "all disks must be on

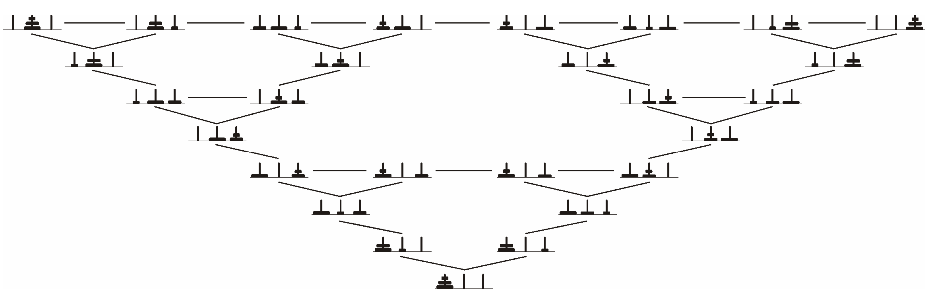


Fig. 2. Prototypical problem space for the "Tower of Hanoi" problem

one of the pegs" are not violated (see Fig. 2) Likewise, we establish a problem space for specific tasks to be performed with a software tool.

3.3 Merging Competence Structures and Problem Spaces

To provide the system with holistic information about the competence domain (i.e., the competence structure and the basis of a set of tasks to be accomplished with a given tool) and the possible actions within the problem solving processes (i.e., the problem spaces for the set of possible tasks), both must be linked together.

On a very fine grained level of competencies, with a large number of possible tasks, and with a large degree of freedom for the user, unfortunately, we would end up with vast competence structures and problem spaces, which could not be handled and computed in real time. In the context of games, we introduced the concept of *game states*, which we want to term now *system state*. These are meaningful accumulations of specific states in the virtual environment (e.g., when specific objects are assembled to a machine correctly or in terms of regular software, for example, when a set of options such as font, size or colour are set appropriately for a specific task). This accumulation is now a substrate of the problem space, which can be linked to the user's available and unavailable competencies. There are two options to do so; first, a *deterministic linkage*, and second, a *probabilistic linkage*. The former means that each system state is associated with a specific set of competencies that the user has and a specific set of competencies that the user lacks. The latter means that a numeric value that describes the strength of belief that a specific set of competencies is available or lacking is assigned to each position category. Independent of the linkage type, a *utility value* is assigned to each state to provide the game with information about "how appropriate" or "how inappropriate" a system state is for a specific task.

3.4 Updating Competence State Probabilities

In a next step it is necessary to transfer the assumptions of available and lacking competencies to entire competence states and the likelihood of those states. To identify a user's initial (i.e., at first use of the software) competence state, we assume an initial probability distribution over the competence structure. There exist an almost arbitrary number of possibilities to obtain such initial probability distribution. The simplest form would be a uniform distribution in which each competence state has the same initial probability. The selection of the right method strongly depends on the given information about the user's competencies and prior knowledge as well as the user's goals.

With each action the user performs, the system updates the probability distribution over the competence states, where updating rules define the way in which the probabilities are updated in a specific situation [12]. A simple method is to increase the probabilities of all competence states that include competencies that are (either definitely or likely) available when realising the corresponding position category. Conversely, we can decrease the probabilities of those competence states that include competencies that are unavailable. On this basis, we are continuously approaching every action and every realised position category with a clearer interpretation of the learner's competence state. Although single interpretations may not be convincing, with an increasing number of actions, certain and most often similar competence states become increasingly clear. These procedures are now the basis for supporting the user in a personalized, suitable, and hopefully in a non annoying manner.

4 Adaptive Support and Feedback

The basic idea of the micro adaptivity concept is to support the user with suitable guidance and support and with informative feedback. We have described so far, how we think that the system can gather a holistic picture of what the user knows, in terms of which competencies s/he has, and at which point of the problem space the user is on his/her way to accomplish a specific task (having the idea in mind that accomplishing a task can be interpreted as solving a problem in a given, well-defined and rule-based setting). The next logical step is to equip the system with a set of responses that can be triggered if necessary. Bearing in mind the game-based learning context, we can distinguish following types of response:

- *Competence activation interventions* may be applied if a user is stuck in some area of the problem space and some competencies are not used even though the system assumes that the user possesses them. For example, the system can make a specific function (e.g., a tool bar button) more visible.
- *Competence acquisition interventions* may be applied in situations when the system concludes that the user lacks certain competencies. For example, the system informs the learner about specific features of the software and how to use them.
- *Feedback* may be utilised to provide the user with information about the work progress. For example, the software could inform the user if s/he has used the software in an efficient way.
- *Assessment clarification interventions* may be applied, for example, in the form of a query if the user's actions provide contradicting support for the assumption of a certain competence state. In addition, such type of intervention maybe used to determine which task a user wants to accomplish.

Such kind of support and guidance proved to be effective in the learning context [15], more importantly in the context, recent research showed that such interventions could improve immediate goal achievement [16].

5 Leering at the Games: 80Days

5.1 Architecture

The concept of micro adaptive assessment and interventions has been developed in the context of the ELEKTRA project and its successor 80Days. In 80Days, prototype game was developed to demonstrate and evaluate the concept of micro adaptivity. The architecture consists of four modules, (1) the game itself, quite traditionally, is created using a high-end gaming engine, (2) a *skill assessment engine* updates the competence state probabilities, and the resulting information regarding the learner's competence state and its changes are then forwarded to (3) an *educational reasoner*, the pedagogical part of micro adaptivity. Based on a set of pedagogical rules and meta-rules as well as learning objectives, this engine provides recommendations on adaptive interventions to the adaptation realisation module. This, in turn, maps the abstractly formulated educational recommendations onto more concrete game recommendations. In this mapping process, data on game elements and information on previously given recommendations are considered. The necessary information for the assessment-intervention loop is stored (4) in an *OWL* [17], which allows the aforementioned engines to extract not only information, but also the relationships among the information from the ontology.

5.2 The Game

We have developed a learning adventure that is supposed to teach geography for the age group of 12 to 14 years. The curriculum includes, for example, knowledge about the planet Earth such countries or cities but also aspects such as longitude or latitude. In the game the learner takes the role of a boy or a girl (depending on the learners' gender) at the age of 14. The story starts from an extraordinary event; a space ship is landing in the backyard and an alien named Feon appears. Feon turns out to be a friendly alien, being an alien scout who has to collect information about foreign planets, in this case planet Earth. The learner accompanies Feon and is having fun with flying a space ship and exploring interesting places on Earth. Feon creates a report about the Earth and its geographical features. This is accomplished by the player by means of flying to different destinations on Earth, exploring them, and collecting and acquiring geographical knowledge. The goal is to send the Earth report as a sort of travelogue about Earth to Feon's mother ship. At a certain point in the game, however, the player makes a horrible discovery; the aliens are not really friendly but collect all the information about Earth to conquer the planet, lately. This discovery reveals the "real" goal of the game: The player has to save the planet and the only way to do it is to draw the right conclusion from the traitorous Earth report. The subject matter is embedded in hat story and learning occurs at various events in the game. From a pedagogical point of view, learning occurs by receiving information (e.g., seeing/reading something in the game or hearing something from Feon or other game characters), problem solving (e.g., reducing the negative impacts of a flood by appropriate "terra-forming"), or imitation (e.g., watching other game characters and learning from their behaviours). Screenshots are shown in Fig. 3.

Fig. 3. Screenshots from 80Days' game demonstrator

6 Conclusion and Outlook

The idea of adaptation on the micro level is to monitor a learner's behaviour within an adaptive DEG and to provide him/her with appropriate and tailored support and feedback. In the context of the 80Days project we implement the corresponding functionalities in a compelling demonstrator game. This game was subject of experimental evaluations which showed that the envisaged kind of "subliminal" or "hidden" assessment not only works in terms of assessment accuracy but also that the corresponding interventions lead to superior learning performance, motivation, and satisfaction with the game in comparison to no adaptive interventions or inappropriate interventions [18].

In the present paper we raised the idea that such non-invasive assessment method and especially the underlying logic might be successfully applied ion the context of regular software as well. While the famous and unsuccessful even annoying paper clip assistant of Microsoft's Office was a good idea in its essence, the underlying logic was not smart enough to convince most of its users. If we can transfer the psychologically sound frameworks of CBKST in combination with problem spaces we can equip a software system with a "deeper understanding of the user and his/her needs – at least in comparison to other support methods.

Of course, this approach is still in its infancy and must be elaborated in future work. Exactly here lies the major purpose of this paper, that is, introducing the concept to a community broader than that of game-based learning, to encourage discussions and to broader the concept for a variety of applications.

Acknowledgments

The research and development introduced in this work is funded by the European Commission under the seventh framework programme in the ICT research priority, contract number 215918 (80Days, www.eightydays.eu).

References

1. de Freitas, S.: Learning in immersive worlds. A review of game-based learning (2006), `http://www.jisc.ac.uk/media/documents/programmes/elearning_innovation/gaming%20report_v3.3.pdf` (retrieved August 28, 2007)
2. Prensky, M.: Digital game-based learning. McGraw-Hill, New York (2006)
3. Van Eck, R.: Digital game-based learning. It's not just the digital natives who are restless. Educause Review (March/April 2006)
4. Kickmeier-Rust, M.D., Peirce, N., Conlan, O., Schwarz, D., Verpoorten, D., Albert, D.: Immersive digital games: Interfaces for next-generation e-learning. In: Stephanidis, C. (ed.) HCI 2007. LNCS, vol. 4556, pp. 647–656. Springer, Heidelberg (2007)
5. Kafai, Y.B.: Playing and making games for learning. Games and Culture 1(1), 36–40 (2006)
6. Csikszentmihalyi, M.: Flow: The psychology of optimal experience. Harper and Row, New York (1990)

7. De Bra, P.M.E.: Adaptive hypermedia. In: Adelsberger, H.H., Kinshuk, J.M.P., Sampson, D. (eds.) Handbook on Information Technologies for Education and Training, pp. 29–46. Springer, Berlin (2008)
8. Kickmeier-Rust, M.D., Albert, D., Hockemeyer, C., Augustin, T.: Not breaking the narrative: Individualized Competence Assessment in Educational Games. In: Proceedings of the European Conference on Games-based Learning (ECGBL), Paisley, Scotland, October 25-26 (2007)
9. Albert, D., Lukas, J. (eds.): Knowledge spaces: Theories, empirical research, and applications. Lawrence Erlbaum Associates, Mahwah (1999)
10. Korossy, K.: Extending the theory of knowledge spaces: A competence-performance-approach. Zeitschrift für Psychologie 205, 53–82 (1997)
11. Doignon, J.-P., Falmagne, J.-C.: Spaces for the assessment of knowledge. International Journal of Man-Machine Studies 23, 175–196 (1985)
12. Doignon, J.-P., Falmagne, J.-C.: Knowledge spaces. Springer, Berlin (1999)
13. Newell, A., Simon, H.: Human Problem Solving. Prentice-Hall, Englewood Cliffs (1972)
14. Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge (1990)
15. Azevedo, R., Bernard, R.M.: A meta-analysis of the effects of feedback on computer-based instruction. Journal of Educational Computing Research 13, 111–127 (1995)
16. Tan, J., Biswas, G., Schwartz, D.: Feedback for metacognitive support in learning by teaching environments. In: Proceedings of the 28th Annual Meeting of the Cognitive Science Society, Vancouver, Canada, pp. 828–833 (2006)
17. Kickmeier-Rust, M.D., Albert, D.: The ELEKTRA ontology model: A learner-centered approach to resource description. In: Leung, H., Li, F., Lau, R., Li, Q. (eds.) ICWL 2007. LNCS, vol. 4823, pp. 78–89. Springer, Heidelberg (2008)
18. Kickmeier-Rust, M.D., Marte, B., Linek, S.B., Lalonde, T., Albert, D.: The effects of individualized feedback in digital educational games. In: Conolly, T., Stansfield, M. (eds.) Proceedings of the 2nd European Conference on Games Based Learning, Barcelona, Spain, October 16-17, pp. 227–236. Academic Publishing Limited, Reading (2008)

Investigating Sociability and Affective Responses of Elderly Users through Digitally-Mediated Exercises: A Case of the Nintendo Wii

Yin-Leng Theng, Pei Foon Teo, and Phuong Huynh Truc

Wee Kim Wee School of Communication and Information
Nanyang Technological University, Singappore 637718

Abstract. Worldwide, the growing number of senior citizens is a major source of concern in many countries. Recent developments in digitally mediated games aim to encourage exercise among users. It is generally reported in the well-established Technology Acceptance Model (TAM) that users' positive perceptions of ease of use and usefulness with IT systems will be sufficient to predict users' behavioural intention to use the systems. This paper describes a study to explore effects of digitally-mediated exercises such as those in Nintendo Wii on elderly users. We first developed a theoretical model based on TAM linking the antecedents into the key factors using information acquired from an extended literature review. We then undertook a survey of thirty elderly participants at a senior citizens centre in Singapore. Data collection was carried out via a self-reported questionnaire and video observation. Through the use of statistical tests, our findings showed that perceived ease of use and perceived health value were determinants of perceived usefulness. Perceived affective response mediates the influence of perceived usefulness on satisfaction, leading to elderly users' behavioural intention to use. The paper concludes with a discussion on the design and impact of digitally mediated games for elderly users.

Keywords: Senior citizens, elderly users, Wii, Technology Acceptance Model, sociability, emotive responses, healthcare, perceived usability, perceived usefulness.

1 Introduction

Worldwide, the growing number of senior citizens is a major source of concern in many countries. Recent developments in digitally mediated games (e.g. [11]; etc.) aim to encourage exercise and healthy food consumption among the elderly. Cheok, Lee, Kodagoda, Khoo and Le [2] in the design of inter-generational game for elderly and children demonstrate that exercise is important for both physical as well as mental health. Um [20] claims that regular exercise could reduce anxiety and depression while increasing self-esteem and satisfaction among the elderly. McGuire shows that social interaction between humans is essential for the enjoyment of life [15]. Although Keyani, Hsieh, Mutlu, Easterday, Forlizzi [14] designed a tool that was meant to provide entertainment and exercise for the elderly, the tool also enables positive social engagement.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 152–162, 2010.

Traditionally, digital games have long been considered something for children or young people to be played during their leisure times. Surprisingly, with several new developments in digital games, more games are targeted at elderly users (e.g., [1]; [12]; etc.).

The recently release of the Nintendo Wii games has rapidly become the market leader in gaming since the first launch in November 2006. The proven successful selling record was easily over 20 million units worldwide after a year time. With new interaction devices providing multimodality user experiences, computer games provide active way of engagement with elderly players. Using the WiiRemote, users can experience a greater freedom of body movement than mouse interaction in front of a computer screen. Furthermore, it offers multi-user game experiences with the help of its wireless physical interfaces [13]. This significantly helps to improve the social interaction among the elderly during gameplay.

The Nintendo Wii has created a phenomenon among senior citizens especially those in the United States and Japan to improve health and social interaction among the elderly. The positive implications of playing the Wii seem plenty and they range from improving health of elderly through exercise, to enhancing social interactions among family members and so on. The Wii seems like a fun way for seniors to get some exercise, stimulate their minds, and have a jolly good time. In addition, mental stimulation postpones the onset of dementia and might actually reverse the process. New evidence from the Alzheimer's Society suggests that the progression of Alzheimer's can be slowed by the use of computer-based puzzles. The Wii could also be an ideal toy for the elderly because its controller tracks spatial movement, which allows gameplay with normal human movements.

2 Case Study: Singapore

In the recent report on Committee on Ageing Issues by the Ministry of Community Development, Youth and Sports in Singapore (2006) (retrieved on May 22, 2010, http://www.mcys.gov.sg/successful_ageing/report/CAI_report.pdf), it is reported that Singapore will witness an unprecedented profound age shift. The number of residents aged 65 years or older will multiply threefold from current 300,000 to 900,000 in 2030.

In Singapore, the population is also ageing rapidly. For example, the middle ages of 50 to 59 years old for 2007 have superceded those of 1997 by 50 percent. The study on the local ageing population discussed different issues which faced by the elderly in Singapore, and these issues range from housing, accessibility of amenities such as transportation, affordable healthcare, and active lifestyles. It is also recognised that active ageing will lead to social integration and healthy living among the elderly.

Aging of population was defined as a summary term for shifts in the age distribution of a population toward older ages. It is expected to be the most prominent and significant global demographic trends of the 21st century [7]. Indeed, it demonstrates several challenges for public health and economic development towards the society.

Wii has definitely taken off in Unites States especially among the elderly. The responses and reported observations and news in the Internet have made quite a headline that more elderly homes are looking into Wii gaming seriously. In Singapore, a community centre in North Bridge Road has embraced and accepted Wii as a way to create social cohesion for the elderly within the area [6]. Everyday about 30 to 40 elderly residents from the Peace-Connect community center participate or watch elderly play a game of tennis or bowling with the Wii. The article also noted that having fun with Wii has helped to improve the resident's motor skills and hand-eye coordination for about 700 out of the 900 members aged 60 and above.

Based on the success from our European and US counterparts [8], more senior centres are interested to obtain digitally mediated games for the elderly. However, there is little research conducted locally to analyze the significance level of effectiveness and acceptance of these games among elderly in Singapore.

We conducted a pilot study in 2008/9 and found positive results [19]. This paper presents a follow-up study to further investigate the determinants of senior citizens' pre-adoption attitudes towards digitally mediated games for promotion of exercise and health consumption.

An underlying hypothesis of such fitness computer games is that it can be fun for the elderly to play and exercise at the same time; consequently it will improve their physical fitness. In this paper, we describe a second study conducted to examine the impact of digitally-mediated games on sociability and affective responses of elderly gamers in Singapore. Specifically, the main objectives of this paper are:

- To investigate the extent of affective responses and satisfaction of elderly users when using the Nintendo Wii.
- To examine whether the Nintendo Wii provides a platform for sociability for elderly users.

Section 3 describes the theoretical background and research model. Sections 4 and 5 discuss results and findings. Section 6 concludes with implications of game design for elderly users.

3 The Study

3.1 Research Model

Fishbein and Ajzen [5] developed Theory of Reasoned Action (TRA) Model that defines the connectivity between beliefs, attitudes, norms, intentions and behaviors of individuals in the field of social psychology. Hartwick and Barki [9] explain the role of user participation, involvement, and attitude during system development in IS implementation. Davis [3] continued with the formulation of the Technology Acceptance Model (TAM) to measure the factors that influence the acceptance and adoption of information systems that have integrated into our everyday lives.

TAM, derived from TRA [5, 9], studies the effect of user participation in systems development on IS use. While TRA is a general theory of human behavior, TAM is specific to IS usage. TAM has been proven robust in predicting user acceptance of IT, and has been applied widely in understanding the motivational issues in computer and software adoption, as well as usage of information systems [10]. There are various reasons to adopt TAM model as our project research methodology. One of the crucial reasons is because TAM is based on behavioral intention of people, hence it can be useful in determining whether the technology is suitable for the elderly from the psychological perspective. Emotional responses towards the interaction with the technology can be constructive as it has the direct connectivity in terms of usefulness and ease of use. TAM has been proven robust and applied widely in identifying and understanding the motivational issues in term of software adoption and usage of information systems [10].

Based on approach, the model helps to predict the acceptability of an information system; which determined by two main factors: perceived usefulness (PU); and perceived ease of use (PEOU). PU and PEOU are hypothesized to jointly decide users' intention to accept an IT application [3]. The behavioral elements assume that when a person forms an intention to act, one would be free to use it without any limitation.

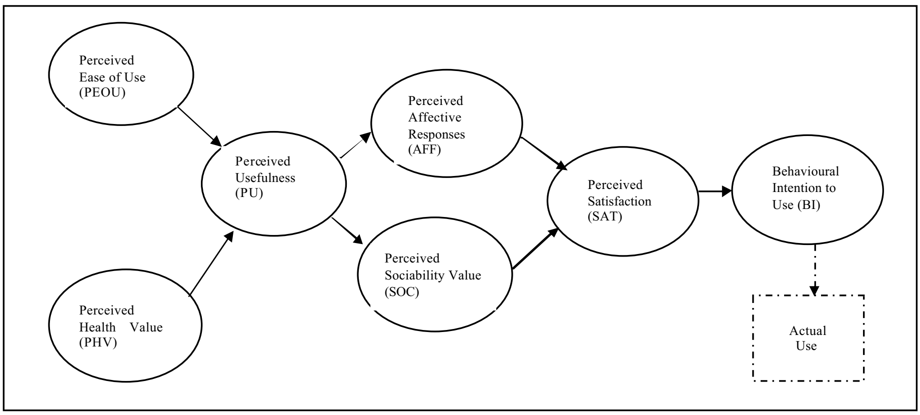


Fig. 1. Research Model

Figure 1 shows our research model adapted from TAM, and the factors explored for this study are as follows:

- Perceived Usefulness (PU) is defined as "the prospective user's subjective probability that using a specific application system will increase his or her job performance within an organizational context" [3].
- Perceived Ease of Use (PEOU) is defined as "the degree to which a person believes that using a particular system would be free of effort" [3].
- Behavioral Intention to Use (BI) is defined as "a measure of the strength of one's intention to perform a specific behavior" [5].
- Perceived Health Value (PHV). Regular physical exercise is always found to be difficult for elderly although it would greatly improve their physical health condition [4]. PHV is the perception of the value of the Wii in promoting exercise and healthy living [20]

- Perceived Affective Responses (AFF). Affective response refers to the emotional responses perceived by the elderly while interacting with Wii game. This measure is closely related to the satisfaction factor. By investigating the consistency of the emotional responses among the target respondents during the research study, we can simply identify the attitude of elderly when interacting with the Wii game. Measuring this factor is good to evaluate whether a wide range of cognitive issues [17] has been well-considered in designing the Wii games especially for elderly.
- Perceived Sociability Value (SOC). The Wii game offers opportunities for elderly to play with other people, expanding their social circle of friends [8, 15].
- Perceived Satisfaction (SAT). In this measure, we are attempting to understand the satisfaction level of the elderly when integrating with Wii. The satisfaction level towards Wii could reflect directly the pleasure, enjoyment and the actual engagement of Wii game with the elderly. Importantly, it was tailored to examine the satisfaction level which was identified to be significant to increases elderly users' self-confidence, ability to learn, and memo-retention [16]. As it is proven from other studies [18], the increase of self-esteem and life satisfaction has positive influence on mental health of aged people.

3.2 Measurement Items in the Survey Instrument

The survey instrument contained measures of the characteristics including health behaviors; emotion and sociability. The constructs employed in the research model was measured using multi-item scales. The survey instrument was formulated to include all the items written in the form of statements with 7-point Likert scales. We conducted a pilot testing of the survey instrument and revisions were made accordingly. Table 1 shows the measurement items (predictor latent construct items) used in the survey instrument.

3.3 Protocol

We contacted a few elderly centers to find out if they were interested in the study. The centers contacted were Pertapis, CareCorner, Retired and Senior Volunteer Programme (RSVP) and lastly PeaceConnect. After much deliberation, we concluded that RSVP would the primary centre of interest. RSVP is a non-profit organization focusing on providing community-based services to the elderly, particularly in the areas of education, health, IT and inter-generation integration. Its mission statements are to be a premier entity of active seniors serving and enriching their lives through volunteerism (retrieved on 22 May, 2010, http://www.csv-rsvp.org.uk/site/home.htm).

Similar to most developed countries, in Singapore, chronological age of "65 years and above" is defined as "elderly" or "older persons". This is different from the United Nations (UN) criterion for older population which is someone "60 years and above". Some scientific publications categorize the senior citizens into the "young-old" (aged 65 to 74), "old-old" (aged 75 to 84), and "oldest-old" (85 and over), or "octogenarians" (aged 80-89), "nonagenarians" (aged 90-99), and "centenarians" (over 100), for the purpose of this study, such distinct classification is not relevant. The terms *senior citizens*, *elder(s)*, *elderly*, *senior(s)*, *aged*, and *older person(s)* would be used interchangeably.

Table 1. Measurement Items: Predictor Latent Construct Items

Perceived Ease of Use (PEOU)	
PEOU1	Learning to Wii is easy for me.
PEOU2	Using Wii would not require a lot of my mental effort.
PEOU3	Overall, I find that Wii is easy to use.
Perceived Health Value (PHV)	
PHV1	Using Wii helps me to avoid tension.
PHV2	Using Wii helps me to stay healthy longer.
PHV3	Using Wii helps me to enjoy life more.
PHV4	Using Wii helps me to stay fit.
PHV5	Using Wii helps me to look younger.
Perceived Usefulness (PU)	
PU1	Using Wii would improve my health
PU2	Using Wii would improve my social interaction.
PU3	Using Wii is a good entertainment option.
Perceived Affective Responses (AFF)	
AFF1	When interacting with Wii, I feel delighted.
AFF 2	When interacting with Wii, I feel happy.
AFF 3	When interacting with Wii, I feel cheerful.
AFF 4	When interacting with Wii, I feel pleased.
AFF 5	When interacting with Wii, I feel eager to play.
Perceived Sociability Value (SOC)	
SOC1	When interacting with Wii, I feel friendly towards others.
SOC2	When interacting with Wii, I feel cooperative towards others.
SOC3	When interacting with Wii, I feel tolerant towards others.
SOC4	When interacting with Wii, I feel attentive towards others.
SOC5	When interacting with Wii, I feel patient towards others.
Perceived Satisfaction (SAT)	
SAT1	After interacting with Wii, I am interested in Wii,
SAT2	After interacting with Wii, I feel a sense of engagement.
SAT3	After interacting with Wii, I feel a sense of surprise.
Behavioural Intent to Use (BI)	
BI1	I am willing to use Wii to improve my health.
BI2	I am willing to use Wii to improve my social interaction.
BI3	I am willing to use Wii for entertainment purpose.

However, this study uses "50 years of age or older" as the general definition of senior citizens. Although it was recognized that difficulties may emerge while comparing results with other studies using a different definition of senior citizens, it was deemed necessary to solicit opinions from people in their 50's to gain a wider perspective on both *retirees as well as pre-retirees* as they are approaching retirement soon and their responses will reflect future needs of senior citizens.

The study was conducted in March 2009. Subjects aged "50 and above" were invited to take part in the study. There were not too many participants in the "above 65" years old age group at the RSVP centre. The participants at RSVP were paid S$10 for taking part in the 45-minute session.

First, the participants were instructed to play the different Wii games using the basic set: tennis, bowling, boxing, etc. After 10-15 minutes of interacting with the Wii games, they were then instructed to complete the survey instrument.

4 Findings and Analysis

4.1 Descriptive Statistics

Thirty-one participants, aged 50 years and above, took part in the study. Table 2 shows the demographics of the participants. More than 50% of the participants were above 60 years old. Two-thirds of the participants were females. About 32.3% of participants went to RSVP three to four times a week and 45.2% less than 2 times a week, indicating active engagement of these participants with activities at the RSVP.

Table 2. Demographics of Participants

Demographic variables		Frequency	%
Gender	Male	9	29.0
	Female	21	67.7
Age	50 years	1	3.2
	51-60 years	13	41.9
	60-70 years	14	45.2
	Above 70 years	3	9.7
Education	Primary	3	9.7
	Secondary	19	61.3
	Vocational	3	9.7
	Diploma	3	9.7
	Degree	3	9.7
First Time User of Wii	First time user	15	48.4
	Have used Wii before	16	51.6
Frequency of Visit to RSVP	Less than 2 times a week	14	45.2
	3-4 times a week	10	32.3
	4-5 times a week	3	9.7
	5-6 times a week	1	3.2

Table 3. Descriptive Statistics

Construct	Code	# of Indicators	Min	Max	Mean	SD
Perceived Usefulness	PU	3	4	7	5.63	0.956
Perceived Ease of Use	PEOU	3	2	7	5.25	1.018
Behavioural Intention	BI	3	2	7	5.45	1.185
Perceived Health Value	PHV	5	3	7	5.45	0.973
Perceived Affective Responses	AFF	5	4	7	5.83	0.737
Perceived Sociability Value	SOC	5	4	7	5.76	0.874
Perceived Satisfaction	SAT	3	5	7	5.78	0.737

Of the thirty-one participants, 61.3% attained secondary school education, confirming our initial observation that this group of elderly was fairly well-educated and able to understand and communicate in English. In terms of Wii awareness, about 51.6% of participants used Wii at least once.

Means and standard deviations of the model variables are shown in Table 3. On a scale from 1 to 7, from “strongly disagree” to “strongly agree” respectively, the range of means of the constructs ranged from 5.25 to 5.83, all above the “neutral” or middle of the Likert scale. Due to space constraints, we report only some of our findings in this paper.

4.2 Test of Measurement Model

Tables 4 and 5 show the reliability and discriminant validity scores of the measurement items. The measurement model links each construct in the theoretical model to indicators of the constructs. In our research model, we were interested in the indicators of the PU and PEOU constructs. The strength of measurement model is determined by conducting convergent and discriminant validity tests of the instrument items.

Table 4. Measurement Model – Reliability

Variable Constructs	Cronbach’s Alpha	The Composite Reliability (Internal Consistency Reliability)	Average Variance Extracted/Explained (AVE)
1. Perceived Usefulness	0.9526	0.9694	0.9136
2. Perceived Ease of Use	0.8716	0.9205	0.7945
3. Behavioural Intention to Use	0.9638	0.9764	0.9324
4. Perceived Health Value	0.9633	0.9714	0.8716
5. Perceived Affective Responses	0.9859	0.9889	0.9469
6. Perceived Sociability Value	0.9751	0.9805	0.9097
7. Perceived Satisfaction	0.9807	0.9873	0.9629

Table 5. Measurement Model - Discriminant Validity

Latent Variables	AFF	BI	PEOU	PHV	PU	SAT	SOC
AFF	**0.9731**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BI	0.7254	**0.9656**	0.0000	0.0000	0.0000	0.0000	0.0000
PEOU	0.7628	0.5593	**0.8913**	0.0000	0.0000	0.0000	0.0000
PHV	0.8450	0.6495	0.6393	**0.9192**	0.0000	0.0000	0.0000
PU	0.9117	0.6540	0.7293	0.8809	**0.9548**	0.0000	0.0000
SAT	0.9359	0.6958	0.7202	0.7909	0.8927	**0.9674**	0.0000
SOC	0.9624	0.6955	0.7573	0.8291	0.8913	0.9111	**0.9810**

AFF=perceived affective responses; BI=behavioural intention; PEOU=perceived ease of use; PHV=perceived health value; PU=perceived usefulness; SAT=perceived satisfaction; SOC=perceived social value

Convergent validity reflects the extent to which the indicators of a construct are similar to the other indicators of the same construct. The convergent validity of each construct was assessed by computing Cronbach's alpha. Cronbach's alpha for the all the constructs were above 0.7, which is a commonly acceptable level for judging the reliability, indicating that more than 70% of the variance has been accounted for.

4.3 Structural Model

After validating the measurement model, we used SmartPLS to test the structural model. The results of the analysis are depicted in Figure 2.

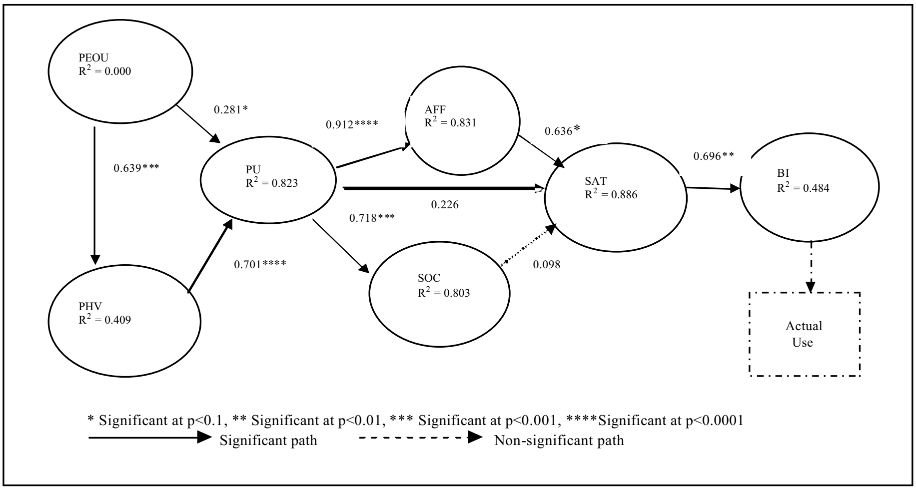


Fig. 2. Research Model Results

Assessing the two antecedents of perceived usefulness, we find that perceptions about the 'perceived ease of use' has a significant relationship with 'perceived usefulness' (path=0.281, t=2.78, p<0.1). The relationship between perceptions of 'perceived health value' and 'perceived usefulness' was found to be significant (path=0.701, t=7.37, p<0.0001). We also examined the relationship between the 'perceived ease of use' and 'perceived health value' and found a significant positive relationship between the two constructs (path=-0.639, t=3.82, p<0.001).

The two 'perceived ease of use' antecedents explain 82% of the variance in 'perceived usefulness'. This suggests the high explanatory power of the theorized antecedents of perceived usefulness, providing empirical validation for the proposed research model.

From the results in the 'consequences' part of the research model, we observe that 'perceived usefulness' in elderly perception has a weak significant relationship with 'perceived satisfaction' (path=0.226, t=1.74). However, the study shows a strong significant relationship between 'perceived usefulness' and 'perceived affective responses' (path=0.912, t=14.13, p<0.0001), and a moderate significant relationship between 'perceived affective responses' and 'perceived satisfaction" (path=0.636, t=2.424, p<0.1). A strong support is also shown specifying a significant relationship between 'perceived usefulness' and 'perceived social value' (path=0.718, t=3.708, p<0.001). But, there is no significant relationship between 'perceived social value' and 'perceived satisfaction (path=0.098, t=0.504). As shown in Figure 2, we observed that 'perceived affective responses' mediate the influence of 'perceived usefulness' on 'perceived satisfaction'.

Next, we observe that 'perceived satisfaction' is a significant predictor of 'behavioral intention' for adopting the Nintendo Wii (path=0.696, t=3.268, p< 0.01). Although, Davis [3] argued that ease of use may act indirectly on intentions to use through usefulness, this is indeed an anomalous result and requires deeper investigation to understand the reasons for the differential importance of 'perceived usefulness' and 'perceived ease of use' in the context of the Nintindo Wii providing digitally-mediated exercises for elderly users.

5 Conclusion and On-Going Work

This study examined the motivations that might influence the intentions of using the Wii among the elderly. Through the use of statistical tests, our findings showed that perceived ease of use and perceived health value were determinants of perceived usefulness. Perceived affective response mediates the influence of perceived usefulness on satisfaction, leading to elderly users' behavioural intention to use. However, there is no significant relationship between perceived social value and perceived satisfaction, implying that perhaps personal enjoyment with Wii is more important than having the opportunities playing with others.

Apart from the entertainment value, the study shows that the Wii games invoked positive affective responses. The involvement with the Wii games could provide alternative ways to keep the elderly active. Although not investigated in this project, the contribution of regular physical exercises through playing Wii games might help elderly stay healthy while stimulating their mental abilities through problem solving and cognitive skills through such gameplay.

Certainly, further studies could be conducted with wider sampling frame and bigger sampling size to achieve greater reliability and generalisability. Future studies could also look at hands-on experiments to identify usability problems encountered by senior citizens when using the Web. Nevertheless, the implications of the study were valuable. Firstly, this is a localised study that looked into the perceptions of senior citizens in Singapore and their perceptions towards digitally mediated games. Secondly, it provided the policy makers and game designers/developers in Singapore a better understanding of the elderly's perceptions and needs, thus narrowing the digital divide in Singapore among the young and old in terms of gameplay. Lastly, it could serve as a reference for future studies for gameplay as applied to the elderly in addressing active ageing.

Acknowledgements. The authors would like to thank the subjects who participated in the study, and Amirrudin Bin Dahlan, Meutia Latifah Akmal, Thant Zin Myint for conducting the pilot study. The project is supported by the Mitsui Sumitomo Insurance Welfare Foundation (MSIWF).

References

1. Abdullah Al, M., Omar, M., Suleman, S., Jean-Bernard, M.: Designing and evaluating the tabletop game experience for senior citizens. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges, vol. 358, pp. 403–406 (2008)
2. Cheok, A., Lee, S.P., Kodagoda, S., Khoo, E.K., Le, T.: A Social and Physical Inter-Generational Computer Game for the Elderly and Children: Age Invaders. In: Proceedings of the Ninth IEEE International Symposium on Wearable Computers, pp. 202–203 (2005)
3. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use and User Acceptance of InformationTechnology. MIS Quarterly 13, 319–340 (1989)
4. Dishman, R.K.: Exercise adherence: Its impact on public health, pp. 445–447. Human Kinetics Publishers, Champaign (1998)
5. Fishbein, M., Ajzen, I. (eds.): Belief, attitude, intention, and behavior: an introduction to theory and research. Addison-Wesley Pub., Reading (1975)
6. GameFrontz, Wii a hit with elderly residents in Singapore too (May 13, 2008), `http://gamefrontz2.wordpress.com/2008/05/14/wii-a-hit-with-elderly-residents-in-singapore-too/` (retrieved August 16, 2008)
7. Gavrilov, L.A., Heuveline, P.: Aging of Population. In: Demeny, P., McNicoll, G. (eds.) The Encyclopedia of Population, vol. 1, pp. 32–37. Macmillan Reference, USA (2003)
8. Geeksugar, Seniors Have Some Swing Batter Batter Wii Fun (August 16, 2007), `http://www.geeksugar.com/533365` (retrieved August 16, 2008)
9. Hartwick, J., Barki, H.: Explaining the role of user participation in information system use. Management Science 40(4), 440–465 (1994)
10. Igbaria, M., Zinatelli, N., Cragg, P., Cavaye, A.: Personal computing acceptance factors in small firms: a structural equation model. MIS Quarterly 21(3), 279–305 (1997)
11. IJsselsteijn, W.A., de Kort, Y.A.W., Westerink, J., de Jager, M., Bonants, R.: Virtual Fitness: Stimulating Exercise Behaviour through Media Technology. Presence: Teleoperators and Virtual Environments 15(6), 688–698 (2006)
12. IJsselsteijn, W.A., Nap, H.H., de Kort, Y.A.W., Poels, K.: Digital Game Design for Elderly Users. In: Proceedings of Futureplay 2007, pp. 17–22 (2007)
13. Lee, H.J., Kim, H., Gupta, G., Mazalek, A.: WiiArts: Creating collaborative art experience with WiiRemote interaction. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, pp. 33–36 (2008)
14. Keyani, P., Hsieh, G., Mutlu, B., Easterday, M., Forlizzi, J.: DanceAlong: Supporting Positive Social Exchange and Exercise for the Elderly through Dance. In: CHI 2005: ACM Conference on Human Factors in Computing Systems, pp. 1541–1544 (2005)
15. McGuire, F.A.: Improving the quality of life for residents of long term care facilities through video games. Activities, Adaptation & Aging 6, 1–7 (1984)
16. Ogozalek, V.Z.: A Comparison of the Use of Text and Multimedia Interfaces to Provide Information to the Elderly. In: CHI 1994 Conference Proceedings, Human Factors in Computing Systems, pp. 65–71 (1994)
17. Shneiderman, B.: Universal Usability. Communications of the ACM 43(5), 84–91 (2000)
18. Tad, H., Jodi, F., Elaine, H., Jennifer, G., Jacey, S., Chris, K.: The elder project – social, emotional and environmental factors in the design of eldercare technologies. In: ACM Conference on Universal Usability, pp. 72–79 (2000)
19. Theng, Y.L., Dahlan, A., Akmal, M., Myint, T.: An Exploratory Study on Senior Citizens' Perceptions of the Nintendo Wii: The Case of Singapore. In: Third International Convention on Rehabilitation Engineering and Assistive Technology, Singapore (2009)
20. Um, S.H.: The effect of a regular exercise on mental health on aged people. Science and Technology 3(26), 284–286 (2004)

Acquaintances Clustering for Social Relationship-Based Indexing of Digital Photos

Jonghak Kim, Taekwon Jang, Joonhyuk Yang, and Jung-hee Ryu

Graduate School of Culture Technology,
Korea Advanced Institute of Science and Technology (KAIST),
373-1 Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea
{airjonghak,taekwon,joony.yang,junghee.ryu}@kaist.ac.kr

Abstract. One of the effective ways to manage large collections of digital photos is to tag names of people appearing in those photos. However, the number of people appearing in photo collections may range in the hundreds and the names of tagged people are usually presented in alphabetical order or on a first-tag first-place basis. As a result, it is difficult to quickly search for name tags that a user wishes to find. In order to solve this problem, we developed a digital photo management system that automatically groups the name tags based on their social relationships. This system was tested on users' own photos against three other comparison interfaces. The average searching time for name tags was significantly faster with our system. Also, the user satisfaction was higher than the others.

Keywords: Acquaintances clustering, episodic memory, group of people, name tag, photo indexing, photo management, tag clustering.

1 Introduction

As the use of digital cameras became popular, personal digital photo collections quickly increased in number and size. The increasing number of photos makes it difficult for users to manage and retrieve their collections. As a result, there is a growing demand for tools to help managing, organizing, and browsing those large photo collections.

Various researches were conducted to develop an effective photo management system. Many of them focused on human episodic memory because it is important for people to recall past experiences [7, 32]. Therefore, researchers tried to develop a photo management system that fits with contents and structures of episodic memory. Episodic memory means that memories may be organized by events (episodes). The events include information such as the location of an event, the persons who were present, and what occurred before, during, and after the event. Among this information, the core information of episodic memory is a person [15, 28]. When people described about their photos, names of people in the photos were the most frequently

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 163–174, 2010.

mentioned information [23]. Also, people's ability to recall the participants of an event decreases quite gradually compared to other information [31]. In addition, a person enabled faster access to past experiences and helped to recall a greater number of total events than the others [3, 25, 26]. Finally, the most frequent content type of photos was a person [13]. Therefore, managing personal photo collections by person information is very important to develop an effective photo management system.

1.1 Related Work

There were many approaches to use person information for managing digital photos. One of the ways to use person information was to tag names of people appearing in photos. Name tagging could significantly improve the usefulness of photo collections [18, 19]. The most commonly-cited benefits were to help recall and support search. However, people have mostly avoided tagging, because it was laborious and the future benefits were unanticipated. In order to support the creation of name tags, various face detection and recognition technologies were used [1, 2, 9, 11, 14]. Consequently, name tagging activities became much easier and faster than before.

In many of the previous works, the name tags were usually presented in alphabetical order [1, 9, 11, 14] or on a first-tag first-place basis [2]. However, the number of people appearing in users' photo collections could range in the hundreds and the span of human memory imposed severe limitations [22]. As a result, it was difficult to manage and search the name tags that a user wished to find. Even worse, there were no relationships between closely placed name tags. Whenever users attempted to search for the name tags of related people, they needed to search the whole list repeatedly. As the human brain operates by association [5], the searching for the related name tags could occur frequently and become one of the reasons to increase the searching time.

In order to solve this problem, we focused on the cognitive strategies of people to reduce the complexity of memory and searching tasks. These are 'chunking' and 'narrowing down'. First, the term chunking means the formation of stimulus elements into subgroups in order to facilitate the assimilation of total information [22]. It is required because the memory span of human has a fixed number of chunks (7 ± 2). In order to break this information bottleneck, people are able to increase the number of information that it contains by building larger chunks. Second, the narrowing down is to use special landmarks or anchors for guiding recall and search [4, 6, 7, 12, 16, 17, 20, 21, 27, 30]. This was frequently observed when people wanted to search for a particular photo. For example, they would first think of an event that was relevant and use it as a guideline to dip into the collection and then move backwards or forwards. Through this process, the actual search-space or browse-space could be relatively limited. By using these two strategies, people could reduce the time for searching and retrieving information.

1.2 Research Problem

Based on our related works, we hypothesized that if a system provides the chunking and narrowing down, it could reduce the time for searching name tags. In order to achieve these, we focused on the intuition that groups of people exist and they are usually exclusive to each other. A group is defined as two or more people who for longer than a few moments interact with and influence one another [29]. We expected that the group could be an effective chunking and narrowing down criteria. We established two main objectives. The first was to develop a name tag clustering method that classifies each name tag into chunks that only hold socially related name tags. The other was to verify whether arranging name tags by their groups could facilitate the task for searching name tags and increase the user satisfaction.

2 System Development

We considered two or more people who meet and interact together as groups based on the Shaw's definition. Our input data were time-stamped photos with name tags of people appearing in them. By using this data, we developed the name tag clustering method that automatically groups the name tags on the basis of their co-attendance at events.

2.1 Step 1: Name Tag Clustering by Co-occurrence

First, we counted the number of co-occurrences (name tags that appear in the same photo) of every pair of name tags. After that, we classified each name tag into groups according to the rule that the top three name tags having the highest co-occurrence frequency be included in the same group. We repeated this process until every name tags satisfied this condition. We could then extract the primary groups.

After the above process, originally two or more groups could be merged into one group because a name tag was able to co-occur with members of several other groups. For example, a user's girl friend could appear in the same photo with his family or co-workers. In this case, his family, co-workers and his girl friend could be merged into a group as she took photos with them. In order to divide this heterogeneous group, we extracted the key name tag which co-occurred most frequently with its group members. Then, we made a list of name tags co-occurring with the key name tag. Next, we computed co-occurrences of each name tag in the list with the heterogeneous group members. Then, if more than 49% of the name tags in the list co-occurred with less than 13% of their group members (the percentages were chosen empirically), we grouped the name tags again by the same rule without the key name tag. If the heterogeneous group was divided into two or more groups, the key name tag was inserted into the group that held a name tag which co-occurred most frequently with the key name tag. Pseudo code was described in Fig 1.

```
Procedure Group classification (total name list, total photo list) {
   Co-table ← Build Co-occurrence table (total name list, total photo list)
   Group list ← Grouping (total name list, Co-table)
   Final Group list ← Regrouping (Group list, total photo list)
}

Sub-Procedure Build Co-occurrence table (total name list, total photo list) {
   For each name from total name list
   Choose photos which contain current name from given total photo list

   For each chosen photo
   Extract tags (=names) without duplication from photo

   For each extracted tag
   Update co-occurrence frequency of [name, tag] pair
}

Sub-Procedure Grouping (total name list, co-occurrence table) {
   For each name from total name list
   Choose the top three names having the highest co-occurrence frequency with name

   For each chosen candidate
   Make group (name, candidate)
}

Sub-Procedure Make group (name, candidate) {
   If name and candidate are not group members then
   Make new group then add name and candidate as member

   If name and candidate are belong to different groups then
   Move one of them to bigger group

   If one of (name and candidate) is belong to certain group then
   Move non-member to certain group
}

Sub-Procedure Regrouping (total group list, total photo list) {
   For each group from total group list
   Choose group-related photo list from total photo list
   Build Co-occurrence table(group members, chosen photo list) to make local table

   If group is separable then
   Remove key name from group who has the highest co-occurrence with its members

   Grouping (current group members, local table)
   Insert 'key name' into the group that includes the highest number of names co-
   occurred with it
}
```

Fig. 1. Pseudo code of our name tag clustering method (Step 1)

2.2 Step 2: Name Tag Clustering by Photo Creation Time

After the group refinement, there could be name tags that do not belong to any group, because they did not have co-occurrence with other name tags. In order to classify these name tags into groups, time-based clustering was conducted. Time clustering detects noticeable time gaps in the creation time of digital photos for identification of events [8]. If a gap is much longer than the local average gap, it is considered a change of event. When an independent name tag appeared in a time cluster containing other name tags, we assumed that the independent name tag attended an event together with them. We then inserted the independent name tag into the group that held the highest number of name tags appearing in the same time cluster. As time gaps had a very wide range, we used Platt et al. [10, 24]'s time clustering algorithm, because this algorithm adaptively determines the gap between events. After time clustering, there could still be name tags that weren't belonged to any group. It was because they didn't have co-occurrence and didn't appear in a time cluster containing other name tags. These name tags were classified into groups alone.

2.3 Step 3: Arrangement of Groups and Its Members

After name tag clustering, we computed the time that the most recent photo was taken for each group member. We then arranged the most recently taken group at the first place because the most important use of personal photos is looking at recently taken photos and sharing those with friends and family [6]. In each group, members were placed in alphabetical order. This was provided as another means of narrowing down to limit the actual search area.

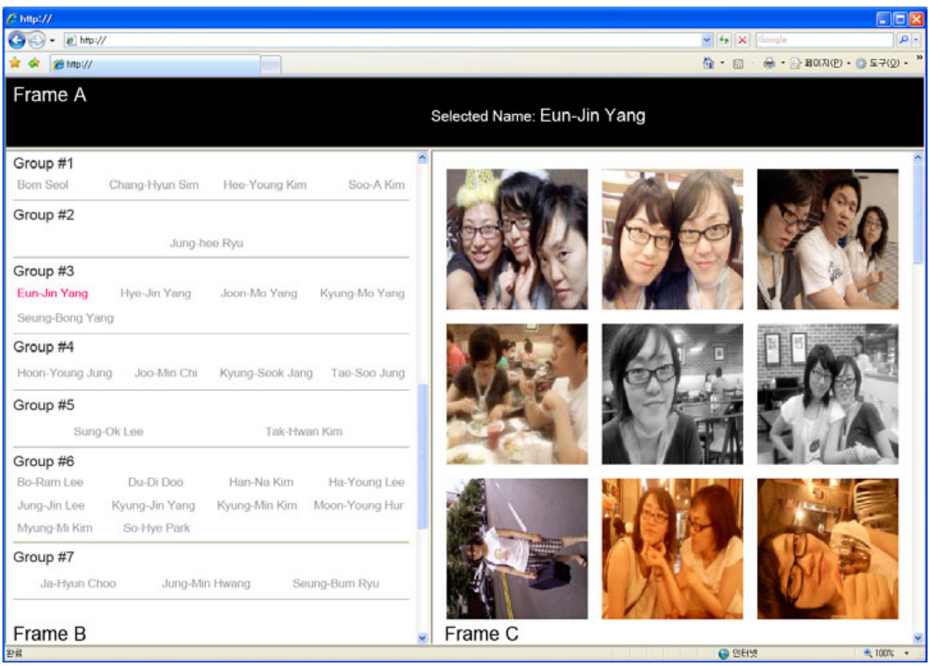


Fig. 2. Prototype photo management system applied our clustering method

2.4 Prototype Photo Management System

We developed a prototype photo management system applied our name tag clustering, as shown in Fig 2. The system was implemented by using C, HTML and PHP languages. Our user interface consisted of 3 frames. In frame B, the name tags were arranged according to their groups. Groups were distinguished by horizontal lines. Users could navigate the arranged name tags by using the scroll bar. When users selected a name tag, its color was changed from gray to red and it was displayed in frame A. Then, photo thumbnails related to the selected name tag were displayed in frame C. After that, users could browse the extracted photo collections. If users clicked a photo, it was presented by its original size.

3 User Study

We conducted user studies to evaluate the performance of our name tag clustering method and the value of arranging the name tags according to their social relationships (groups of people).

3.1 Subjects

We recruited fourteen subjects who had their own digital cameras and had taken photos for more than four years. These included 6 men and 8 women. Their ages ranged between 25 and 54, with an average age of 30.5 (SD = 7.3). All subjects had normal or corrected-to-normal vision, and had experiences in managing their own digital photo collections.

3.2 Data Acquisition

Before the experiments, each subject was asked to provide their personal digital photos taken during the period from January 2005 to September 2008. Among the vast amount of photo collections, there were some photos that didn’t have time stamps. We excluded them from our data set. After that, the number of time-stamped photos ranged from 247 to 5468 photos, the average size being 1739. The names of people appearing in each photo were manually tagged by each subject. In order to help the creation of name tags, we provided a web site. Log-in IDs were allotted to each subject. When they logged-in, their personal photos were presented in chronological order. The creation time of each photo was presented together. In each photo, it had a button for inputting name tags. The number of people appearing in their photo collections ranged from 9 to 132, the average being 55.

3.3 Measures

According to ISO 9241-11[1], usability is defined as effectiveness, efficiency and satisfaction. Effectiveness can be evaluated by accuracy and efficiency by time. At the beginning, in order to evaluate our clustering method by its accuracy, we conducted a questionnaire about how subjects classify their name tags into groups. However, their grouping criteria or levels were very diverse. For example, some subjects classified their father, mother and relatives into one group, ‘family’. On the other hand, others classified them into two or more groups, ‘family’, ‘relatives’, etc. Therefore, it was

[1] ISO 9241 is a multi-part standard covering a number of aspects for people working with computers.

hard to evaluate our system with the criterion of accuracy. As a result, we used time and user satisfaction to evaluate our system.

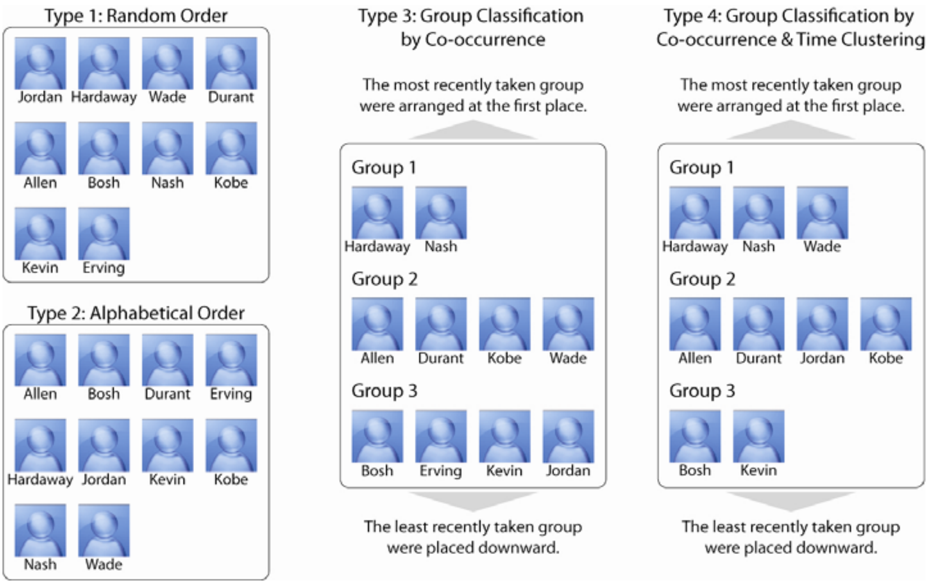


Fig. 3. Differences of each comparison interface

3.4 Comparison Interfaces

We made three comparison interfaces to verify the performance of our system by measuring time for searching each name tag, as shown in Fig 3. In Type 1 interface, name tags were arranged in a random basis without name tag clustering. Type 2 employed alphabetical order, also without name tag clustering. Type 3 employed a name tag clustering method based only on the co-occurrence of each name tag. For example, if name tags A, B appeared in photo 1 and B, C in photo 2, this classified A, B, and C as belonging to the same group. Type 4 employed our proposed name tag clustering method (Co-occurrence + Time Clustering).

There were major differences between Type 3 and 4. Type 4 classified name tags into groups by using both co-occurrence and time clustering. By time clustering, therefore, Type 4 was able to classify name tags that always occurred alone in photos into groups. In addition, Type 4 also considered the frequency of co-occurrence. As a result, if name tags of other groups co-occurred in a photo, this could classify them into different groups. On the other hand, Type 3 classified name tags by using only co-occurrence without considering frequencies. Therefore, Type 3 couldn't classify the independent name tags into groups and distinguish other group members co-occurring in a photo. We made Type 3 to compare with our proposed method (Type 4). In both Types 3 and 4, the most recently taken group was placed in the first place and the members of each group were arranged in alphabetical order.

3.5 Experiments

Each subject completed a series of name tag searching tasks on four different interfaces. When one of their name tags was randomly presented on the screen, the subjects navigated the interface, found the indicated name tag and clicked on it. Then, the searching time was recorded and the next name tag was automatically presented. This task was repeated ten times in each interface. The conditions were counter-balanced to avoid learning effects, so that half of the subjects experienced the non-group condition (Type 1, 2) before the grouped condition (Type 3, 4), and the other half experienced the conditions in the reverse order. In order to avoid ordering effects, the sequence of name tags was randomly changed for every pair of subjects.

After completing the task of searching name tags, the subjects were asked to use each interface for some time. Then, we conducted a questionnaire and the subjects rated their degree of satisfaction with each interface on a 7-point Likert scale ranging from 1, very dissatisfied, to 7, extremely satisfied.

4 Result

4.1 Searching Time

The average times for searching name tags were 7.54 sec (Type 1: random order), 4.12 sec (Type 2: alphabetical order), 4.20 sec (Type 3: grouped by only co-occurrence) and 3.37 sec (Type 4: grouped by our proposed method, using co-occurrence and time clustering), as shown in Fig 4. We used an RM-ANOVA on the interfaces and then performed post-hoc analysis using least-significant difference (LSD) to see the difference between each interface. Our system (Type 4) was significantly different with Type 1 ($p<0.005$) and Type 3 ($p<0.028$). However, it slightly failed to reach statistical significance compared to Type 2 ($p<0.065$).

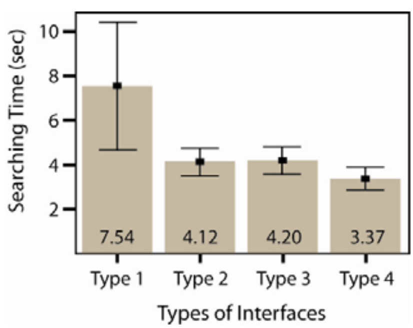


Fig. 4. The average times for searching name tags of each interface. Error bars represent standard errors.

We anticipated that our system would be more effective when the number of name-tagged people was increased. For further investigation, we classified the subjects into two groups (A and B). Group A consisted of the subjects whose collections contained more than 55 people (the average number of people appearing in subject's photo collections). Group B was less than 55. As shown in the Fig 5, when the number of name-tagged people was below 55, there were no significant time differences for searching name tags between each interface. However, when the number was more than 55, the average time was significantly increased in Type 1, 2 and 3. In contrast, there was no significant time increasing in Type 4. The average times for searching name tags were 11.35 sec (Type 1), 5.33 sec (Type 2), 5.35 sec (Type 3) and 3.31 sec (Type 4).

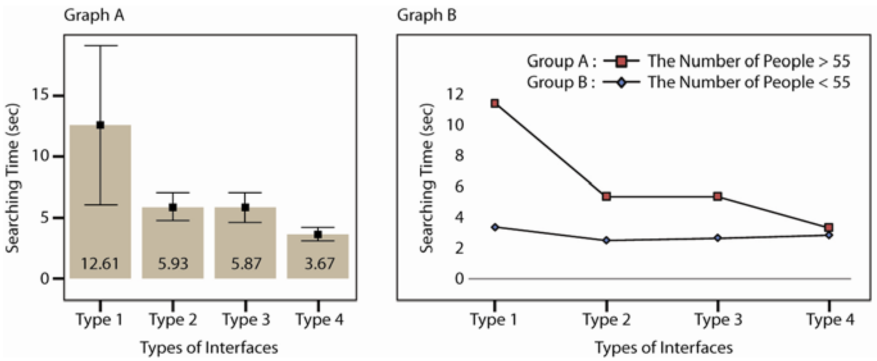


Fig. 5. Graph A - The average times for searching name tags of Group A. Graph B - Time differences for searching name tags between Group A and B.

A repeated measured analysis of variance (RM-ANOVA) with one within-subject factor (Interfaces) and one between-subject factor (Number-of-People) indicated a significant difference between group A and B (Wilks' lambda = 0.853, F=7.011, df=3.000, p-value < 0.05). The RM-ANOVA and LSD was performed only with the data of Group A. In this condition, our system (Type 4) showed significant differences with the other interfaces (Type 1: p<0.007, Type 2: p<0.001, Type 3: p<0.002). As a result, we could find that users were able to find their name tags significantly faster with our system. Also, it showed a stable searching time regardless of the number of people appearing in user's photo collections.

4.2 User Satisfaction

One-way ANOVA analysis was performed on the satisfaction for each interface. The result was statistically significant with F-value=12.52, p<0.0001. The average user satisfaction was 2.64 (Type 1), 4.43 (Type 2), 5.71 (Type3) and 5.86 (Type 4), as shown in Fig 6. The degree of user satisfaction with Type 2, 3 and 4 was significantly higher than Type 1. Especially, it was much higher in Type 3 and 4.

In addition, our subjects were satisfied with the clustering results. Fig 2 showed the name tag clustering results of our subject 3 (female and 28 years old). For example, her co-workers were classified into group 1. Group 2 was her graduate school professor. Group 3 was her family and relatives. Group 4 was her juniors in design school.

Group 5 was administrative staffs of her graduate school. Group 6 was friends of her university. Group 7 was Korean movie stars. She was very satisfied with this clustering result.

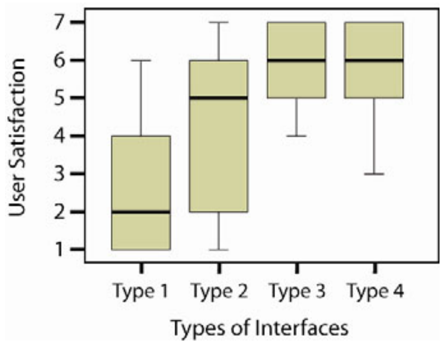


Fig. 6. The degree of user satisfaction with each interface on a 7-point Likert scale

5 Conclusion

In order to reduce the complexity of tasks for searching name tags in photo management system, we developed the name tag clustering method that automatically groups those tags based on their social relationships. We also implemented the prototype photo management system applied our clustering method.

After the implementation, we conducted user studies to evaluate our clustering method and the effect of arranging the name tags according to their social relationships. Our system was tested on subjects' own photo collections against three other comparison interfaces. The results showed that subjects could find their name tags significantly faster with our system. Especially, it showed a stable searching time regardless of the number of people appearing in user's photo collections. In addition, the user satisfaction was higher than the others.

6 Future Work

Events include information such as what, where, when and who. Among them, we firstly used the 'who' information because it was considered the most effective information in photo management. In the future, we are going to improve our system by applying other information (e.g. categorized by location or time) and develop a more refined photo management system.

Acknowledgement

We are deeply appreciative of our subjects who provided their personal photo collections and spent much time for our research.

References

1. Girgensohn, A., Adcock, J., Wilcox, L.: Leveraging face recognition technology to find and organize photos. In: MIR 2004: 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (2004)
2. Apple iPhoto, `http://www.apple.com/ilife/iphoto/`
3. Barsalou, L.W.: The content and organization of autobiographical memories. In: Neisser, U., Winograd, E. (eds.) Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory, pp. 193–243. Cambridge University Press, New York (1988)
4. Brown, N.R., Shevell, S.K., Rips, L.J.: Public memories and their personal context. In: Rubin, D.C. (ed.) Autobiographical Memory. Cambridge University Press, Cambridge (1986)
5. Bush, V.: As we may think. Atlantic Mon. 176(1), 101–108 (1945)
6. Kirk, D., Sellen, A., Rother, C., Wood, K.: Understanding photowork. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, April 22-27 (2006)
7. Horvitz, E., Dumais, S., Koch, P.: Learning predictive models of memory landmarks. In: CogSci 2004: 26th Annual Meeting of the Cognitive Science Society, pp. 1–6 (2004)
8. Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T.: Time as essence for photo browsing through personal digital libraries. In: Proceedings of JCDL 2002, pp. 326–335. ACM, New York (2002)
9. Cui, J., Wen, F., Xiao, R., Tian, Y., Tang, X.: EasyAlbum: An interactive photo annotation system based on face clustering and re-ranking. In: Proc. CHI 2007. ACM Press, New York (2007)
10. Platt, J.: Autoalbum: Clustering digital photographs using probabilistic model merging (2000)
11. Kang, H., Shneiderman, B.: Visualization Methods for Personal Photo Collections Browsing and Searching in the PhotoFinder. In: Proceedings of IEEE International Conference on Multimedia and Expo. (ICME 2000), pp. 1539–1542. IEEE, New York (2000)
12. Kemp, S.: An associative theory of estimating past dates and past prices. Psychonomic Bulletin and Review 6, 41–56 (1999)
13. Kindberg, T., Spasojevic, M., Fleck, R., Sellen, A.: The Ubiquitous Camera: An In-Depth Study of Camera Phone Use. IEEE Pervasive Computing 4(2), 42–50 (2005)
14. Kuchinsky, A., Creech, M., Freeze, D., Gwizdka, J., Pering, C., Serra, B.: FotoFile: A consumer multimedia organization and retrieval system. In: Proceedings of CHI 1999, pp. 496–503. ACM Press, New York (1999)
15. Lancaster, J.S., Barsalou, L.W.: Multiple organisations of events in memory. Memory 5, 569–599 (1997)
16. Lansdale, M., Edmonds, E.: Using memory for events in the design of personal filing systems. International Journal of Man-Machine Studies 36(1), 97–126 (1992)
17. Loftus, E.F., Marburger, W.: Since the eruption of Mount St. Helens has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. Memor. and Congnition 11, 114–120 (1983)
18. Ames, M., Naaman, M.: Why We Tag: Motivations for Annotation in Mobile and Online Media. In: Proc. ACM CHI, San Jose, USA (May 2007)
19. Marlow, C., Naaman, M., boyd, d., Davis, M.: Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. In: Proc. Hypertext 2006. ACM Press, New York (2006)

20. Means, B., Mingay, D.J., Nigam, A., Zarrow, M.: A cognitive approach to enhancing health survey reports of medical visits. In: Gruneberg, M.M., Morris, P.E., Sykes, R.N. (eds.) Practival Aspects of Memory: Current Research and Issues, vol. 1. Wiley, Chichester (1988)
21. Merriam-Webster, Merriam-Webster's Online Dictionary, Merriam-Webster, Inc. (2008)
22. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. The Psychological Review 63, 81–97 (1956)
23. Naaman, M., et al.: Context Data in Geo-Referenced Digital Photo Collections. In: Proceedings, Twelfth ACM International Conference on Multimedia (2004)
24. Platt, J.C., Czerwinski, M., Field, B.A.: PhotoTOC: Automatic clustering for browsing personal photographs. Technical Report MSR-TR-2002-17, Microsoft Research (2002)
25. Reiser, B.J., Black, J.B., Abelson, R.P.: Knowledge structures in the organisation and retrieval of autobiographical memories. Cognitive Psychology 17, 89–137 (1985)
26. Ringel, M., Cutrell, E., Dumais, S., Horvitz, E.: Milestones in time: The value of landmarks in retrieving information from personal stores. In: Proceedings of Interact 2003 (2003) (to appear)
27. Rodden, K., Wood, K.: How do people manage their digital photographs? In: Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 409–416. ACM Press, New York (2003)
28. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D.: Stuff I've seen: a system for personal information retrieval and re-use. In: Proceedings of SIGIR 2003, Toronto. ACM Press, New York (2003)
29. Shaw, M.E.: Group dynamics: The psychology of small group behavior. McGraw-Hill, New York (1976)
30. Shum, M.S.: The role of temporal landmarks in autobiographical memory process. Psychol. Bull. 124, 423–442 (1998)
31. Thompson, C.P., Skowronski, J.J., Larsen, S.F., Betz, A.: Autobiographical memory: Remembering what and remembering when. Erlbaum, Mahwah (1996)
32. Tulving, E.: Episodic memory: from mind to brain. Annu. Rev. Psychol. 53, 1–25 (2002)

Generating Exploratory Search Interfaces for the Semantic Web

Michal Tvarožek and Mária Bieliková

Institute of Informatics and Software Engineering,
Faculty of Informatics and Information technologies,
Slovak University of Technology,
Ilkovičova 3, 842 16 Bratislava, Slovakia
{tvarozek,bielik}@fiit.stuba.sk

Abstract. At present, the promise of the Semantic Web has yet to be realized, partly because there are few real-world applications that allow end-users to access, view and process Semantic Web information. We aim to facilitate Semantic Web adoption by providing users with advanced exploratory search capability over Semantic Web data by providing a faceted exploratory search interface for arbitrary Semantic Web repositories. Our approach takes advantage of metadata describing the structure of the respective information spaces to construct facets that can be used to visually construct semantic queries. Subsequently, we generate result overviews to display individual search results and lastly generate an incremental graph-based view for individual resource exploration. We performed proof of concept validation of our interface generation approach and present the lessons learned by a small scale feedback gathering user study.

Keywords: exploratory search, facet generation, faceted browsing, personalization, Semantic Web, user interface generation, graph exploration.

1 Introduction

Although the Web has become an almost ubiquitous virtual information space providing information, services and communication tools, there are still (large) parts of the Web that are not generally accessible to end-users. The Deep Web which accounted for most of the data on the Web in 2005 [5] consists of many databases which can be accessed over the Web by end-users through querying interfaces. Typical search engines cannot index the Deep Web as its contents are hidden in databases and exist only temporarily in the form of web pages when a query is made.

The Semantic Web aims to provide better search and browsing capabilities by enabling machine readability of information on the Web taking advantage of ontologies and metadata [9]. Most Semantic Web content is part of the Deep Web and stored in publicly accessible semantic repositories (e.g., accessible via SPARQL endpoints), in the form of distributed Linked data or as metadata associated with legacy web documents. Despite continuous progress in semantic search engines such as Sindice.com,

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 175–186, 2010.

the original promise of the Semantic Web still remains unrealized [9]. The main challenges for Semantic Web adoption lie in its:

- *Visualization* – Semantic Web contains raw information without any associated presentation templates thus offering no default way to render it in human readable form making end-user grade visualization difficult. Furthermore, resources can be associated with legacy web content (web pages, images, videos, etc.) or have many attributes and relations to other resources causing information overload.
- *Querying* – semantic queries resemble relational database queries rather than typical keyword queries used in web search engines making them impractical for most users. Manual construction of semantic queries (e.g., in SPARQL) is a complex task, which in addition to query language proficiency requires prior knowledge of the respective information space.
- *Exploration* – the Semantic Web is essentially a graph of resources and their attributes and relations, and also associated legacy web documents (e.g., web pages or multimedia). Exploratory search principles [7] stress open ended tasks, learning and understanding of information in context, not just finding a specific resource e.g., with a traditional search engine. Here orientation support, multiple navigation and/or visualization options and the ability to move towards a goal from different directions are needed to provide satisfactory user experience.

Our aim is to *facilitate Semantic Web adoption*, which is now seriously hindered by the lack of end-user grade search and exploration tools, by *providing an exploratory browser for the Semantic Web*. To achieve this goal we also have to address already existing problems associated with the present Web, such as information overload, the navigation problem or web dynamics (i.e., constant, uncontrolled changes).

In our previous work, we devised a faceted semantic exploratory browser taking advantage of Adaptive Web [2] and Social Web [11] approaches to provide personalized visual query construction support and address guidance and information overload [13]. In this paper, we *extend our browser with user interface generation* using metadata describing the presented information spaces to provide users with a smooth user experience accounting for dynamic changes in the information space.

We present an overview of related work in section 2, and *describe the novel generation of the respective parts of the browser's exploratory search interface* –result overviews, graph view and facets in sections 3 and 4 respectively. Next, in section 5 we present proof of concept validation of our approach in the domain of digital images, the feedback gathered in a small scale user study and the lessons learned so far. Lastly, we conclude and summarize our contribution in section 6.

2 Related Work

Our work has a strong multidisciplinary background ranging from Information retrieval to Adaptive web-based systems and HCI with focus on faceted browsers and facet generation, exploratory search, information visualization and the Semantic Web.

Wilson and schraefel performed a study comparing three prominent exploratory browsers – Flamenco, mSpace and RelationBrowser++ [15]. While Flamenco and RelationBrowser++ are more traditional faceted browsers, mSpace takes advantage of

RDF data (native to Semantic Web) to provide users with a set of customizable filters that can be used to visualize a subspace of a high dimensional information space. The RelationBrowser++ is tailored to exploration of large statistical data and persistently displays all facets at the top unlike Flamenco, which hides exhausted facets [16].

Unlike all previous browsers, the faceted browser Factic stresses personalization as a vital feature enabled by user action tracking and evaluation [13]. In order to better understand user behavior in faceted browsers, Kules et al. performed a user study examining how searchers interact with individual parts of a faceted browser. The study discovered that users primarily explore the result list and the facets, while mostly ignoring the current query. Kules also argued that the design of exploratory search tasks as well as methodologies for evaluation of exploratory browsers were still in an early stage of development making thorough evaluation difficult [6].

The BrowseRDF faceted browser provides elementary facet generation capability over simple RDF data [8]. BrowseRDF automatically identifies facets in source data based on several statistical measures, but offers only very limited interaction options and does not consider semantic metadata provided in the more expressive RDFS and OWL formats. Other approaches include automatic multifaceted hierarchy generation from textual collections [3], and middleware solutions posing as proxies between databases and users providing a faceted interface by dynamically suggesting a number of facets using precomputed decision trees [1].

Neither of these approaches can be effectively used for complex interactive exploration of Semantic Web content, which in addition to faceted querying needs to support interactive information visualization and exploration of graphs (the Semantic Web being a graph). The VisGets interface allows users to perform interactive/exploratory web search by querying it in three dimensions – time, location and topic, while also providing advanced visualization of search results [4]. However, VisGets uses its own crawling and indexing engine and thus cannot be effectively used for general web search or Semantic Web exploration. Here, also graph visualization and interaction approaches must be considered as described in [10].

3 Exploratory Search Interface Generation

We previously devised a personalized faceted browser for semantically enriched information spaces [13], which used personalization to improve overall user experience. In order to facilitate exploratory search experience, we extended the base browser with support for multi-paradigm querying and exploration including keyword-based and content-based search (query-by-example), adaptive result overviews and incremental graph-based resource exploration [12]. This however was still not enough due to the dynamic nature of the information spaces (e.g., users constantly adding or modifying information) where the ability to automatically adapt to changes, e.g. by (automatically) generating user interfaces, is crucial.

Thus, our focus lies with the generation of exploratory search interfaces for the Semantic Web environment, although this can be somewhat generalized towards the Deep Web and even legacy Web environment.

In order to support exploratory search and achieve these goals, we need to support three parts of user experience:

- *Query construction*, which includes the initial construction of an exploratory search query, its modification and execution; to support multi-paradigm search and exploration based on our previous work, we need to support keyword-based, view-based (faceted) and content-based (query-by-example) query construction.
- *Result browsing*, which includes the rendering of suitable result overviews, selection of result ordering and the displayed result attributes, and support for effective selection of individual results for further exploration.
- *Resource exploration*, which includes the detailed presentation of individual resources, their attributes and relationships with other existing resources.

We address these issues by generating a set of user interfaces each supporting the individual stages of the exploratory search process. We generate:

- *Faceted browser interfaces* for advanced query construction and modification.
- *Result overviews* for effective presentation of selected result attributes.
- *Graph-based exploration views* for incremental horizontal exploration of semantic resources and their relations with other resources.

3.1 Information Space Representation

We work with semantically enriched information spaces, e.g. an ontological repository, where both metadata describing the structure of the information space and data are represented by ontologies (e.g., in RDFS or OWL). Thus our approach assumes a description of classes, individuals, relations and attributes describing a particular domain. For example, in the digital image domain, *di:Author* and *di:Photo* are classes; *di:Author_1* and *di:Photo_1* are individuals, while *di:createdBy* is a relation between *di:Photo_1* and *di:Author_1*. Similarly, *di:viewedCount* equaling *10* is an attribute of *di:Photo_1*, also defining the domain of the attribute as the class *di:Photo* and its range as an *xsd:int* (see Fig. 1).

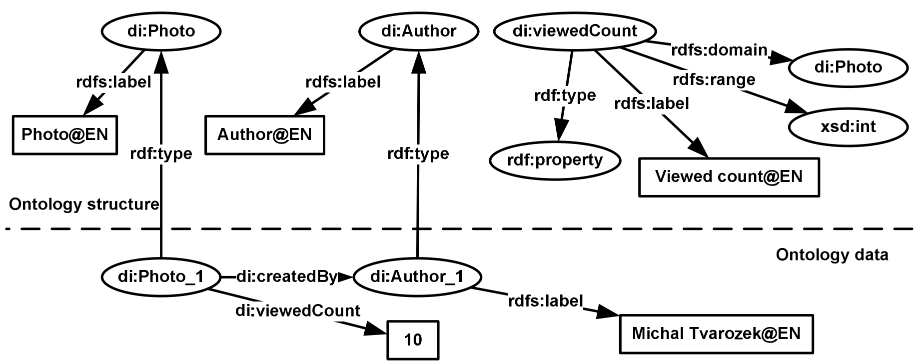


Fig. 1. Example of a simple domain ontology for the digital image domain. Metadata describing the domain model are shown at the top; individuals representing data are shown below. Round nodes denote complex resources with URIs, rectangular nodes denote literals.

As shown in the above example, a domain ontology as defined by W3C contains a detailed standardized description of classes, properties (relations and attributes) and the used data types, effectively defining a data model. Ontologies can also be populated with individuals, which conform to the specified domain model and materialize it in instances of classes and properties. Note that the ontology is in fact an oriented graph where nodes represent individual resources.

3.2 Result Overview Generation

We generate two result overviews – the *ListView* shows thumbnails and properties of individual results (see Fig. 2), while the *MatrixView* shows thumbnails, provides additional information in tooltips, and offers a generated editing pane for (batch) modification of individual result attributes (see Fig. 3).

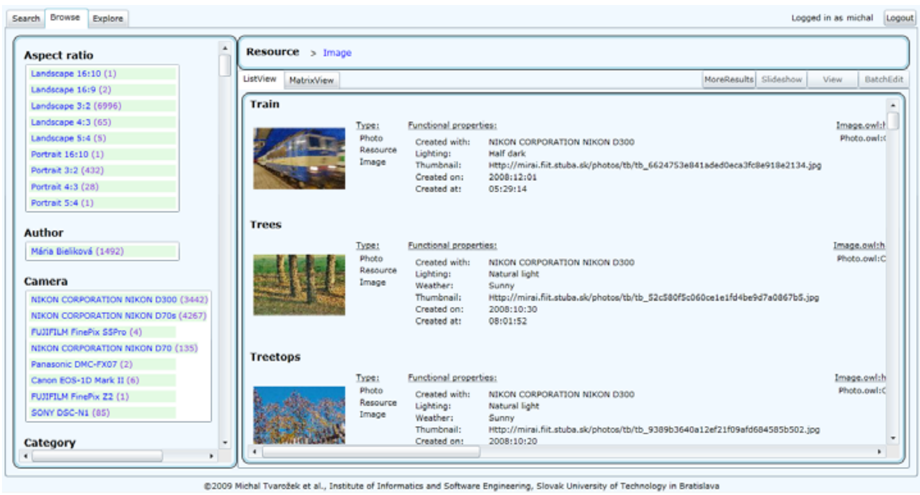


Fig. 2. Example of generated facets shown in the browser GUI (left) along with a list-based result overview showing all result properties (right). The Author facet corresponds to a direct object facet, while the Aspect ratio and Camera facets are indirect object facets associated via an EXIF helper object with the original photo.

In the above example, *ListView* shows properties of a specific result directly derived from the domain ontology visualized as *label-value* pairs. For multi-value properties such as *Type* in Fig. 2, a column with all values is shown. We either show all result properties to maximize information value or apply personalization to select only the most relevant properties (detailed personalization description was described in [13]).

In *MatrixView*, the generation principles remain the same except for the editing pane, which is generated separately. For each specific result type, we identify all applicable properties from the domain ontology metadata, construct editing widgets based on property types (e.g., text boxes with language selection or auto-complete combo boxes with single/multi-value support). Properties with existing values are shown first, while properties without values are shown at the bottom (see Fig. 3).

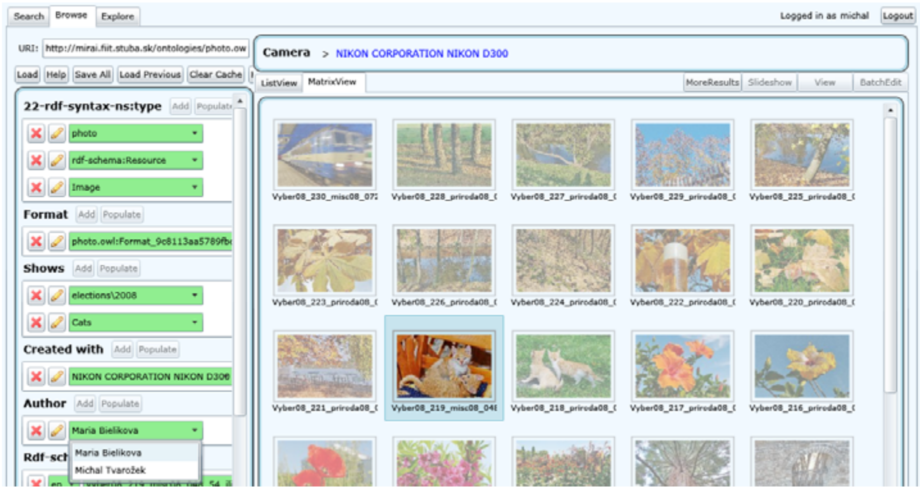


Fig. 3. Example of a generated matrix result overview showing image thumbnails (right), and the correspondingly generated annotation pane for collaborative content creation (left)

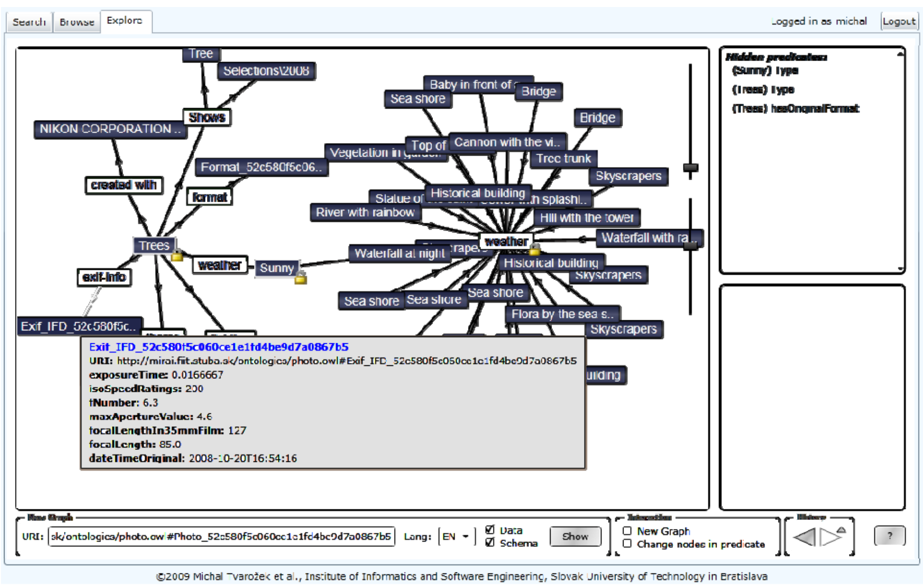


Fig. 4. Example of our generated graph-view exploration interface. Dark nodes represent individual resources, white nodes correspond to relations (top). Hovering over nodes shows the attributes of a node (center); additional tools include zooming, node hiding and history (right), with additional filtering options for languages and data/schema only visualization (bottom). Note that the color scheme has been modified for printing purposes.

The generated graph exploration view consists of the graph visualization window, predicate filtering windows and an options toolbar (see Fig. 4). Users can access the view either directly by typing in the URI of the node they wish to explore or by exploring a result found in the faceted browser (see Fig. 3).

The graph itself is generated directly from the domain ontology showing only individuals (objects) and their relations. Dark nodes correspond to individual resources, white nodes correspond to relations between them; arrows denote relation directions, node attributes (i.e., values of literal properties) are normally hidden and only shown as tooltips after hovering over a node.

Relations are intentionally visualized as separate nodes connecting resources to reduce information overload as one relation can have multiple values and to improve graph layout. E.g. in Fig. 4, the relation *weather* shown on the right would otherwise have to be displayed on all edges making the graph less readable.

Each exploration sessions starts by showing the first dark node (individual) and its direct neighbors, which corresponds to a window or a view of the graph. The user can next move the visible window by selecting another central node, or incrementally expand the view by expanding one or more of the visible nodes. The view in Fig. 4 was initiated by showing the node *Trees* (left) and expanding the node *Sunny*.

To make the graph understandable, we employ a force-based layout algorithm, but also allow the user to fix and manually reposition nodes in the resulting graph. Apart from traditional view panning, users can use two zoom options – regular zoom enlarges or shrinks the view, advanced zoom spatially expands dense node clusters to make them less crowed. Lastly, as ontologies often contain much data, we also enable users to hide irrelevant nodes manually thus reducing information overload.

4 Facet Generation

During facet generation, we examine metadata describing the information space, identify patterns corresponding to facets, construct facet restrictions based on the identified metadata and map the resulting facet onto the graphical user interface and the semantic back-end, which provides querying services. As such, facet generation must define these facet properties:

- A *facet template*, which corresponds to a pattern found in domain metadata and specifies the overall type and behavior of the facet.
- A *restriction template*, which defines how the individual restrictions in the facet are constructed and mapped onto the domain ontology.
- A *query template*, which defines how the back-end query engine creates database queries and maps them onto facet restrictions.
- A *visualization and interaction template* (i.e., the corresponding widget type), which binds the facet to the graphical user interface and handles user input.

The purpose of the facet generation process is to identify specific predefined patterns in the metadata and map them onto a set of predefined templates in three successive steps: facet identification, construction and mapping as described below.

4.1 Facet Identification

During the facet identification stage, we identify the facet template, restriction template and query template. We first search for eligible candidate properties by examining properties of individual instance types and their transitively associated properties, which can be used for automated facet construction from the domain ontology based on low-level metadata facet templates. We distinguish

- *object facet templates* that correspond to properties having complex object values (e.g., a class such as *di:Author*), and
- *literal facet templates* that correspond to properties having simple values (e.g., numbers, dates or strings).

In the above example (see Fig. 1), the *di:createdBy* is a suitable candidate property matching the *class-property-class* pattern between *di:Photo* and *di:Author*, resulting in a object facet template.

Similarly, we define several *restriction templates*:

- *enumeration*, which corresponds a flat list of restrictions (e.g., days of the week),
- *hierarchical taxonomy*, which corresponds to a hierarchical tree of values connected via a transitive property in the domain model (e.g., a hierarchy of geographical locations such as country-state-city-street).

We distinguish the following *query templates* based on the instance-property relation:

- *Direct query template*, which corresponds to the direct property pattern: *{instance} property {value}*
- *Indirect query template*, which corresponds to the indirect property pattern: *{instance}* $property_1$ *{} ... {}* $property_N$ *{value}*

Our facet identification algorithm tries to match these predefined templates and their variations onto the domain ontology metadata, evaluates possible matches and forwards successful matches to the successive facet construction stage.

4.2 Facet Construction

After a facet has been identified, its internal representation must be constructed before it can be used in the browser. The facet construction stage applies the templates identified in the previous stage, constructs facet restrictions based on the restriction template, and persistently stores facet metadata for future use.

The crucial step of facet construction is the initialization of facet restrictions using the restriction template and the definition of the interaction mode. Normally a facet can work as a list of restrictions from which users can select one or more values, or as a search box where users can search for and select a specific restriction. We determine the interaction mode based on the overall number of potential restrictions; *list mode* is used for a small number of predefined values (e.g., days of the week), *search mode* is used for large numbers of values (e.g., all cities on Earth). If an ordering of values is defined in the ontology for object values, we can also create restriction intervals to cover continuous values (e.g., real numbers or dates).

4.3 Facet Mapping

The last facet mapping stage selects a suitable user interface widget to render the generated facet in the faceted browser, and maps the constructed facet and restriction values onto the widget. The widget provides facet visualization (see Fig. 2) and handles user interaction forwarding events and facet metadata to the back-end search services, which use the *query template* and the user selection in the facet to construct SPARQL queries in order to retrieve results corresponding to the generated facet.

Although a broad range of potential interface widgets could be developed, such as lists, histograms, maps, timelines, etc., their detailed description is beyond the scope of this paper as automated discovery of what specific visualization/interaction to use would likely prove difficult. Thus we only employ list widgets at this time and leave the use of more advanced widget types as one possible direction of future work.

5 Validation

We reworked our original personalized faceted browser prototype [14] with the interface generation principles described above as a client-side Silverlight application working inside a web browser to minimize deployment effort. This allowed us to move user specific functionality onto the client and also provide interactive features not supported by HTML (e.g., the interactive graph view). We performed several experiments to validate individual parts of our approach in the digital image domain.

Our goal was to validate two primary aspects of our solution – our facet and result overview generation approach, and the novel graph-view Semantic Web exploration approach. Consequently, we performed two groups of experiments:

- A proof of concept experiment with the facet generation approach, where the goal was to verify that our approach generates meaningful and usable facets for our personalized faceted browser. Note that the goal was not to generate the best possible set of facets, but rather a good enough set to use for personalization.
- A user study with our graph exploration approach, the goal being to gather user feedback on the generated GUI and its usefulness for Semantic Web exploration.

Data. Our domain ontology of images is based on the popular Kanzaki EXIF ontology (http://www.kanzaki.com/ns/exif) and contains about 8 000 manually and semi-automatically annotated images. The entire ontology consists of 35 classes, 50 properties (including relations and attributes), more than 32 000 individuals and in excess of 150 000 facts. For individual photos, the ontology describes EXIF metadata as supplied by the camera, information about formats in which the photos are available (e.g., resolution, aspect ratios), and optional additional annotations such as the author, the object and background of the photo, the place, overall theme and expression, lighting conditions, weather and the event to which the photo belongs.

Methodology. In the proof of concept experiment, we generated facets from the available data and examined how the original browser behaved in practice and whether the interface was still usable for its intended purpose in terms of usability and performance. We performed several experiments with and without personalization, and also after some changes in the information space have been made.

In the user study with our graph exploration interface, we made our browser available to a target group of 10 end-users aged between 20 and 25 years with an IT background. As none of the users had previous knowledge of Semantic Web principles nor had used similar graph-based tools before, each user was given a brief introduction about the functionality of the browser. Next, the users were asked to complete a set of 5 tasks using the browser which also counted the time and number of clicks made (e.g., finding a specific image, discovering image properties or getting a better understanding of the domain). Lastly, each user was asked to fill out a questionnaire with the results of the tasks and his experience with the browser.

Results and Lessons Learned. The experiments with facet generation proved the approach was viable for interface generation with minimal performance impact. We successfully managed to distinguish facet and restriction templates, direct query templates, and construct and map facets to interface widgets and use them in our exploration interface without any significant negative impact over manually created facets due to facet generation.

Based on our experiments, we want to point out these lessons learned:

- Identification of direct query templates resulted in many facets being generated, which we expected to handle at the personalization stage later in the browser. However, this had negative impact on performance and we had to employ selection metrics (e.g., based on significance) already during the facet identification stage.
- The identification of indirect query templates was limited due to the complexity of selecting viable options. Consequently, either the identification algorithm must be further refined or a workaround via indirect nested facets (i.e., facets in facets) needs to be used complicating facet generation and mapping.
- During facet generation and result overview generation, blank nodes and helper objects in the domain ontology caused problems as, e.g., empty, meaningless or unnamed interface items were generated and had to be accounted for.
- Some generated facets such as location eventually had too many restrictions (e.g., hundreds) making them unusable and significantly decreasing performance. This required the change of the interaction mode from *list mode* to *search mode*, where users could type in their desired restriction instead of selecting from a list of hundreds of items. This problem could also be alleviated by prior hierarchical structuring of the information space before facet generation.
- The users preferred *alphabetical restriction ordering* in facets; other orderings such as relevance based or potency based had negative impact on user experience as users were unable to seek in the restrictions which were in an unexpected order.
- Using type/information specific facet widgets instead of list widgets would likely improve usability in specific cases, such as date selection via calendars, location selection via maps or timeline selection via histograms as was done in [4], but effectively generating mappings for advanced widgets would be more complex.

The user study with the graph exploration interface showed that 9 out of 10 users managed to find the specified image, although the time required varied widely – 141 seconds and 8 clicks were required on average, although the fastest user needed less than 50 seconds while the slowest one required almost 5 minutes. Overall, the users managed to answer 75% of the questions correctly leaving 25% false answers (this also includes answers that were close to the correct ones, but not exactly right).

Based on these results, we conclude that graph-based exploration is viable for Semantic Web browsing as most users were able to accomplish the given tasks despite having no prior experience with a similar interface. Still, improvements to layouting and node selection are necessary to improve understandability and task times, which was also confirmed by user feedback which indicates that non-expanded graphs are easy to understand (rating 4.5 on a 5 level Likert scale), while expanded graphs are less readable (rating 3.4). Further feedback indicates that although response times were generally acceptable, some operations took too long to complete (e.g., loading the new graph after expanding a node took sometimes too long).

6 Conclusions

We described our novel approach to exploratory user interface generation for the Semantic Web with specific focus on facet generation, result overview generation and graph view generation. Based on our experiments in the digital image domain, we argue that the approach works and is viable for its intended purpose. Consequently, we see our primary contribution in:

- Devising an *all-around method for (semi)automated generation of Semantic Web exploratory search interfaces* addressing interfaces for query formulation, result overview browsing and individual result exploration.
- *Enabling Adaptive Social Semantic Web exploration* and thus facilitating the adoption of the Semantic Web by end-users.

Despite the positive feedback we got, we encountered scalability issues with exploratory search approaches over remote repositories, as network delays increase the response time of both interface generation (if performed online) and of actual end-user exploration of the respective information spaces.

Although the described method was primarily intended for Semantic Web repositories and possibly Linked data exploration, most of the described principles could be extended to a Deep Web relational database, provided that metadata describing the structure of the information space were available.

Even more interesting is the extension of the proposed approach for legacy web content for specific pages (e.g., personal browsing history) or for whole web sites (e.g., generating a faceted browsing interface for a typical corporate web site). This could be accomplished by taking advantage of contextual/navigational links between pages and entity extraction approaches *ultimately providing a seamless search and browsing experience for both legacy web and semantic web content.*

Acknowledgments. This work was partially supported by the grants VEGA 1/0508/09, KEGA 028-025STU-4/2010 and it is the partial result of the Research & Development Operational Programme for the project Support of Center of Excellence for Smart Technologies, Systems and Services, ITMS 26240120029, co-funded by the ERDF.

References

1. Basu Roy, S., Wang, H., Nambiar, U., Das, G., Mohania, M.: DynaCet: Building Dynamic Faceted Search Systems over Databases. In: Proceeding of International Conference on Data Engineering, pp. 1463–1466. IEEE CS, Los Alamitos (2009)
2. Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. LNCS, vol. 4321. Springer, Berlin (2007)
3. Dakka, W., Ipeirotis, P.G., Wood, K.R.: Automatic Construction of Multifaceted Browsing Interfaces. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pp. 768–775. ACM Press, New York (2005)
4. Dörk, M., Carpendale, S., Collins, C., Williamson, C.: VisGets: Coordinated visualizations for web-based information exploration and discovery. IEEE Transactions on Visualization and Computer Graphics 14(6), 1205–1212 (2008)
5. Gulli, A., Signorini, A.: The indexable web is more than 11.5 billion pages. In: Special Interest Tracks and Posters of the 14th Int. Conf. on WWW, pp. 902–903. ACM, New York (2005)
6. Kules, B., Capra, R., Banta, M., Sierra, T.: What do exploratory searchers look at in a faceted search interface? In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2009, pp. 313–322. ACM, New York (2009)
7. Marchionini, G.: Exploratory search: from finding to understanding. Communications of the ACM 49(4), 41–46 (2006)
8. Oren, E., Delbru, R., Decker, S.: Extending Faceted Navigation for RDF Data. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 559–572. Springer, Heidelberg (2006)
9. Shadbolt, N., Berners-Lee, T., Hall, W.: The Semantic Web Revisited. IEEE Intelligent Systems 21(3), 96–101 (2006)
10. Schulz, H.-J., Schumann, H.: Visualizing Graphs - A Generalized View. In: IV 2006: Tenth International Conference on Information Visualization, pp. 166–173. IEEE CS, Los Alamitos (2006)
11. Staab, S., Domingos, P., Mika, P., Golbeck, J., Ding, L., Finin, T., et al.: Social Networks Applied. IEEE Intelligent Systems 20(1), 80–93 (2005)
12. Tvarožek, M., Bieliková, M.: Collaborative Multi-Paradigm Exploratory Search. In: Proceedings of the Hypertext 2008 Workshop on Collaboration and Collective Intelligence, pp. 29–33. ACM Press, New York (2008)
13. Tvarožek, M., Bieliková, M.: Visualization of Personalized Faceted Browsing. In: Forbrig, P., Paternò, F., Pejtersen, A.M. (eds.) IFIP: Human-Computer Interaction Symposium. IFIP, vol. 272, pp. 213–218. Springer, Heidelberg (2008)
14. Tvarožek, M., Bieliková, M.: Reinventing the web browser for the semantic web. In: WI-IAT 2009: Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, pp. 113–116. IEEE CS, Los Alamitos (2009)
15. Wilson, M.L., Schraefel, M.C., White, R.W.: Evaluating advanced search interfaces using established information-seeking models. Journal of the American Society for Information Science and Technology 60(7), 1407–1422 (2009)
16. Zhang, J., Marchionini, G.: Evaluation and evolution of a browse and search interface: relation browser. In: Proceedings of the 2005 National Conference on Digital Government Research, vol. 89, pp. 179–188. Digital Government Society of North America (2005)

Can Adaptive Interfaces Improve the Usability of Mobile Applications?

Janet L. Wesson, Akash Singh, and Bradley van Tonder

Department of Computing Sciences, Nelson Mandela Metropolitan University,
PO Box 77000+2741 504 2323
{Janet.Wesson,Akash.Singh,Bradley.vanTonder}@nmmu.ac.za

Abstract. Mobile applications are becoming increasingly widespread and complex. Many of these applications suffer from usability issues, including information overload, screen clutter, lack of task support and limited interaction mechanisms. Adaptive user interfaces (AUIs) have been proposed to address some of these usability issues. The aim of this paper is to investigate how AUIs can improve the usability of mobile applications. This paper discusses several simple types of adaptation that have been shown to yield significant usability benefits for mobile applications. Two case studies are presented to illustrate how an AUI can be incorporated into different types of mobile applications. This paper also discusses the lessons learned from these case studies and presents some implications for designing adaptive systems in the future.

Keywords: Adaptive user interfaces, adaptive systems, mobile applications, usability evaluation.

1 Introduction

Research has shown that modern information systems suffer from several usability problems (issues). These issues are mostly attributed to the complexity and lack of flexibility of the user interface (UI). Information systems developed for mobile phones also suffer from similar usability issues. The physical constraints of these devices have an impact on screen size and available interaction mechanisms. These constraints contribute to the existence of several usability issues for mobile applications.

Adaptive User Interfaces (AUIs) can provide potential benefits for addressing these usability issues [1]. Adaptation of the UI has been identified as an important aspect to be considered in the design of modern information systems [2]. Adaptation techniques include adapting what information to present *(information* adaptation), how to present this information (*presentation* adaptation) and how to interact with this information (*interface* adaptation).

The aim of this paper is to investigate how AUIs can improve the usability of mobile applications. Several simple types of adaptation are discussed that have been shown to yield significant usability benefits for desktop systems and mobile applications. Two case studies are presented to illustrate how an AUI can be incorporated into a typical mobile application and the benefits obtained.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 187–198, 2010.

This paper is organized as follows. Section 2 discusses some usability studies highlighting the usability issues associated with mobile applications. Section 3 discusses the benefits of AUIs and how AUI techniques can be used to address these problems. Sections 4 and 5 describe the design and evaluation of two case studies as well as the different forms of adaptation which were incorporated into these systems. Section 6 presents some lessons learned from these case studies and some implications for the design of adaptive systems in the future.

2 Usability of Mobile Applications

The importance of mobile phones as a communication and computational device is increasing daily. The number of mobile phone users is increasing exponentially and the usability of mobile applications is becoming a critical factor [3].

Looije et al. [4] maintain that mobile devices currently suffer from a number of usability issues. These usability issues can be grouped into three main categories:

- Technical - which refers to the battery life of the phone, network connectivity and the limited screen size;
- Environmental - which refers to temperature, light conditions, noise, distraction, mobility of the user, cognitive and psychological constraints for the user, competition for attention from other tasks and the need to manipulate objects other than the mobile device; and
- Social - which refers to usability issues relating to privacy, acceptance, adoption, comfort and personalization.

A study on the usability of mobile systems conducted by the Nielsen Norman Group in 2009 in the United Kingdom and the United States showed that only 59% of all the required tasks were completed successfully. The usability issues identified in this study included [5]:

- Speed - the biggest factor contributing to users performing poorly.
- Screen - the screen size is too small. There is only a limited amount of information that can fit on the screen.
- Typing – typing on mobile devices is hard. An example is logging in because passwords and usernames often contain a combination of digits and letters as well as special characters.

The next section discusses how AUIs can be used to address some of the usability issues identified in this section, especially the limited screen size and interaction mechanisms.

3 Adaptive User Interfaces

3.1 Types of AUIs

An adaptive user interfaces (AUIs) can be defined as: "… *a software artifact that improves its ability to interact with a user by constructing a user model based on* *partial experience with that user."*[6] The basic premise behind AUIs is that users are different and therefore have different needs from an interactive system. The system should adapt to the user, rather than forcing the user to adapt to the system. Each user's characteristics and/or past behaviour are modelled in an attempt to adapt to his/her needs and desires.

AUIs can be classified according the input variables which influence adaptation and the types of adaptation effects. The following four variables commonly influence adaptation [4]:

- *User:* User-based adaptation is commonly employed in AUIs. AUIs can adapt to the user's preferences, knowledge and skills.
- *Task:* Adaptation according to the user's current task can ensure that the adaptation is relevant and helps users in their current activity.
- *System:* Adaptation can take place to adjust to differing device capabilities and variables such as network connectivity (in the case of mobile devices).
- *Context:* Adaptation according to the user's current context can play an important role, particularly in mobile applications.

AUIs can adapt in many different ways in response to the above variables. Several AUIs have been implemented which adapt the UI to suit the individual user [7], provide assistance with routine tasks [8], recommend and filter information based on user preferences and personalize the presentation of information [9].

Three broad classes of adaptation have been identified as suitable for mobile map-based applications [24]:

- *Information:* The information selected and the organization of the information can be adapted.
- *Visualization:* The presentation of information can be adapted.
- *User Interface:* The user interface can be adapted in a variety of ways, including adaptation of the UI controls and the interaction techniques.

3.2 Benefits and Shortcomings

Browne *et al.* [10] identified a number of high-level potential benefits of AUIs, including improving accuracy and efficiency and aiding in user learning. AUIs have been touted as potential solutions for problems such as information overload and filtering, learning to use complex systems and automated task completion [11]. AUIs also provide the potential to deliver the benefits of customized software at a lower development cost per user [12].

AUIs can provide many potential benefits, but are not without problems. One of the fundamental usability principles identified by Nielsen [13] is user control and freedom. AUIs can be seen as taking control away from the user and putting them at the mercy of the system. Other possible problems and pitfalls include privacy issues, confusion, learnability and obtrusiveness [7, 14, 15]. Some adaptive systems try to make the adaptation process more transparent to users to aid their understanding and to help them develop a more accurate mental model of the system [16].

The following sections describe two case studies investigating the use of AUIs in the domain of mobile map-based applications. Mobile map-based applications were chosen as the application domain as these types of systems are widely used and suffer especially from limited screen size and static interfaces. These two case studies were chosen to investigate how an AUI could be incorporated into a mobile map-based application and the benefits obtained. In the first case study, an adaptive system was developed from scratch. In the second case study, an existing system was re-engineered to incorporate an AUI based on adaptation requirements identified in a field study.

4 Case Study: MediaMaps

MediaMaps was developed as an example of an adaptive mobile map-based visualization (MMV) system. MediaMaps was only developed as an adaptive system and no non-adaptive system was developed, due to the nature of the functionality provided. MediaMaps supports three types of adaptation, namely information, visualization (presentation) and interface adaptation [17]. The design and evaluation of MediaMaps are discussed in the following sections.

4.1 Functional Requirements

The functional requirements for MediaMaps were determined by looking at existing MMV systems and interviewing potential users. This process produced the following functional requirements: Capturing and location-tagging of photos, videos and sound recordings; organization of multimedia into collections, based on time and location; lap-based visualization of media collections and individual items; list-based browsing of media collections; searching of collections based on time and location; and saving and loading of views. Figure 1 shows a screenshot of MediaMaps, showing several media collections and items being visualized in a map-based view.

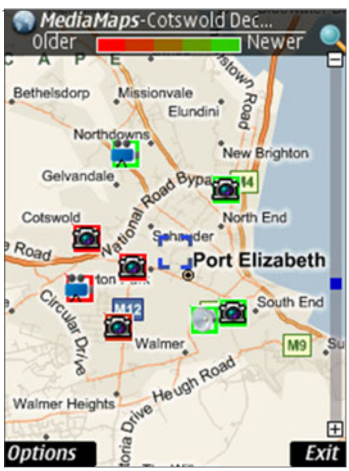


Fig. 1. Screenshot of MediaMaps showing different media collections

4.2 Design

MediaMaps was designed to adapt to the users, their tasks and their context. In order to learn user preferences, a logging mechanism was built into MediaMaps. User interaction is recorded, and various algorithms are used to perform different types of adaptation. The different types of adaptation supported in MediaMaps are briefly discussed below.

Information adaptation is implemented in MediaMaps by sorting media items into media collections based on time and location [18]. The Real-time Event Detection (RED) algorithm was extended to include different media items. The sorting of media items into collections helps to minimize the clutter problem that would otherwise result if a user's entire media collection was visualized in a single view. The grouping of media items is based on time and location. Previous user behavior is used to model the user's media capturing behavior. Media items captured at a similar time and at a similar location are grouped together to form a collection.

Visualization adaptation is implemented in MediaMaps in terms of adapting the visualization to the user's previous behaviour. One the most significant problems of standard zooming and panning techniques is that the user often needs to perform many operations to find information. To address this problem, the visualizations in MediaMaps are adapted in terms of the zoom-level, latitude, longitude, the time period and the current map style (road, satellite photograph or hybrid) [19]. An example of visualization adaptation is shown in Figure 2. The same media collection is shown before adaptation (left) and after adaptation (right). In this example, the zoom-level, map style and location have been adapted based on previous user behaviour.

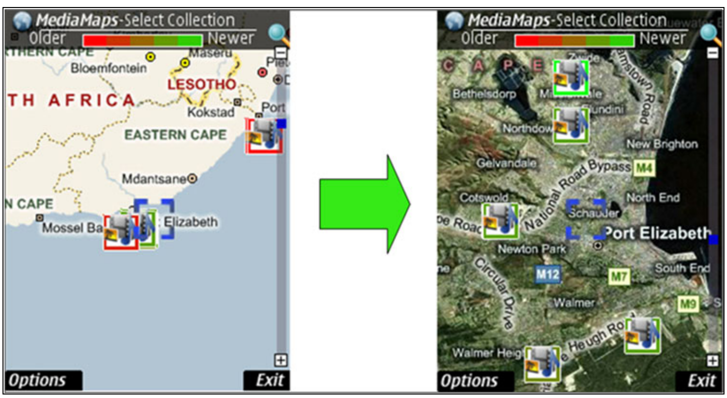


Fig. 2. Visualization adaptation in MediaMaps, before adaptation (left) and after adaptation (right)

Interface adaptation is implemented in the form of list ordering in MediaMaps. These lists include the list of media collections, list of cities and the list of saved views. An example of interface adaptation in MediaMaps is shown in Figure 3. The top section contains the adaptive section of the list, including the most recently used (MRU) option and the two most frequently used (MFU) options [20].

Fig. 3. Example of interface adaptation in MediaMaps, showing the adaptive section at the top of the list

4.3 Evaluation

An international field study of MediaMaps was conducted to evaluate the benefits of the adaptations over a three week period. The following metrics were used to evaluate the three forms of adaptation:

- *Accuracy:* Information and interface adaptation were evaluated in terms of the accuracy of the adaptations performed. Information adaptation was considered to be accurate when the user accepted a recommendation regarding the sorting of a media item into a media collection. Interface adaptation was regarded as accurate when the user selected an item from the adaptive section of the list.
- *User Satisfaction:* Several different aspects of user satisfaction were measured, including the general system usability and the usefulness of the three types of adaptation implemented.

User interaction was logged in order to evaluate the accuracy of the information and interface adaptations. A post-test satisfaction questionnaire was used to elicit user satisfaction feedback. International test participants were recruited using a bidding process [21]. Twenty participants (14 male and 6 female) completed the field study. Participants were recruited from twelve countries on four different continents. The participants all had at least three years of mobile phone experience. The results of the evaluation were analyzed in terms of performance and satisfaction metrics and are discussed below.

4.4 Results

High levels of accuracy were achieved for the sorting of media items into collections (mean = 87.44%). A significant correlation co-efficient of 0.79 (p-value < 0.05) was calculated between the accuracy achieved and the number of media items captured. This implied that the accuracy improved as the number of media items captured increased.

Accuracy information was also calculated for the interface adaptation algorithm. List option selections were split into three categories, namely pre-adaptive selections (when lists were too short to warrant adaptation), adaptive selections and non-adaptive selections. On average, 76.78% of the selections were selected from the adaptive section of the lists.

Five-point Likert scales were used in the post-test user satisfaction questionnaire, with antonyms at the extremes of the scales. The questionnaire was divided into five sections: Sections B and C related to the general usability of capturing media and searching for media items. Sections D-F evaluated the three forms of adaptation.

User satisfaction results regarding the information, interface and visualization adaptation were very positive. High ratings were obtained for the usefulness of sorting media items into collections (mean=4.20 and median=4.50). The usefulness of the ordering of list options was also highly rated (mean=3.80 and median=4.00). Visualization adaptation was also rated as useful by most of the participants.

The results therefore suggest that the three types of adaptation implemented in MediaMaps were accurate and useful. The next section describes the design of an adaptive mobile tourist guide, called A-POInter, as another example of an adaptive MMV system.

5 Case Study: A-POInter

The aim of this case study was to determine the benefits of incorporating an AUI into an existing mobile tourist guide, called POInter [22].

5.1 Requirements Analysis

A list of adaptation suggestions for A-POInter was compiled based on typical requirements for a mobile tourist guide [23] and suggested AUI techniques for MMV systems (Section 3.1). A field study was conducted in order to measure the extent to which participants agreed with these suggestions.

The results of the field study clearly showed that the participants would like to use an adaptive mobile tourist guide. Participants generally preferred most frequently used (MFU) selections for most adaptation parameters, including search criteria, filtering search results, reordering menu items, zooming, panning and the selection of map style. This was to be expected since these tasks involve repetitive user interactions with the system.

Most recently used (MRU) was preferred for the selection of a starting map. This was to be expected as this would be useful for tourists needing to quickly access the same area of interest (AOI) that they used the last time. A combination of MFU and MRU was preferred for placing items at the top of the search criteria list.

5.2 Design

The design and implementation of the adaptive features in A-POInter were separated into the same three types of adaptation, namely information, visualization and interface adaptation. For the purposes of this paper, only some examples of the adaptive features will be discussed.

Upon load completion, a user is taken directly to his/her MRU map (Figure 4). If no MRU map exists (e.g. the system is being used for the first time), then the Home Screen will be displayed instead. In POInter tapping "View Map" from the Home Screen would take the user to a screen for selecting a map, or to detect the starting map using GPS. In A-POInter, tapping "View Map" from the Home Screen will immediately visualise the MRU map. If no MRU map exists, tapping "View Map" will take the user to the map selection screen.

Fig. 4. MRU map with main menu in A-POInter

After a user initiates a search for a category of POIs in A-POInter, the system automatically selects the MFU criteria (Figure 5). Figure 5 also shows the adaptation of list reordering, using the Base Adaptive Partitioning algorithm. The two MFU criteria are selected, but the MRU (3rd item) is not automatically selected.

Fig. 5. A-POInter automatically places MFU criteria at the top of the list

After calculating a set of search results, A-POInter will automatically set the filter percentage to visualise the user's preferred number of search results. After calculating this filter level, the optimal zoom level is calculated and the map is zoomed in or out accordingly (Figure 6).

Fig. 6. Search results displayed in A-POInter

5.3 Evaluation

An international field study was conducted to determine the usability benefits of A-POInter [25]. The majority of participants were recruited from Mob4Hire [21]. Participants were provided with a test plan, which contained a list of tasks and guidelines regarding the functionality of A-POInter. Tasks included selecting, navigating and manipulating a map, searching for points-of-interest (POIs), viewing and filtering search results, interacting with the search results and using the GPS functionality. The tasks performed were the same tasks used during the requirements field study.

A total of 15 participants successfully completed the study. Twelve participants (80%) were below the age of 39; 13 (87%) were male; six (40%) were South African participants and nine (64%) were international participants from four different continents. The majority of participants (93%) had over five years of general computer experience and 12 participants (80%) had experience with a smartphone or PDA (at least one year).

Participants were also asked to complete a post-test satisfaction questionnaire, comprising three sections: Information adaptation, visualization adaptation and interface adaptation. The three sections of the questionnaire were composed of statements scored using a 5-point Likert scale (e.g. 1=Frustrating, 5=Satisfying).

5.4 Results

The results for information adaptation were very positive. Participants found it highly noticeable and highly useful that A-POInter automatically selected the MRU map (median=5). Participants were generally satisfied with this adaptation feature (median=4). Participants found selecting and browsing maps that had been downloaded to be relatively easy (median=4). Several positive aspects were identified from the users' comments, including ease of use, usefulness and enjoyment (satisfaction).

Similar results were also obtained for visualization and interface adaptation. One participant stated the following regarding reordering of search criteria: "*After repeated use, I actually enjoyed A-POInter selecting and moving my preferred search criteria to the top. This prevented me from having to scroll and reselect those criteria each time I searched*". High ratings for ease-of-use (median>=4) were obtained for all the key tasks, namely selecting and browsing maps, searching for POIs, and zooming and panning. Participants indicated that they were highly satisfied overall (median=5) with the adaptations provided by A-POInter.

6 Discussion

This paper has shown that AUIs can be used to effectively address some of the usability issues affecting mobile applications. The results of the two case studies show that including adaptive features in a mobile map-based application can result in improved usability, usefulness and user satisfaction.

The results of the case studies also show that all three types of adaptation can provide significant usability benefits for mobile applications. Information adaptation can assist in sorting data into collections, displaying relevant data and remembering search preferences. Visualization adaptation can assist in remembering preferred visualization techniques, areas of interest and levels of detail. Interface adaptation can assist in remembering preferred interaction techniques, reordering long lists and menu items.

These adaptation techniques can be included in the design of a new mobile application, such as MediaMaps, or incorporated into an existing mobile application, such as A-POInter. More case studies are needed, however to explore the benefits of AUIs for different types of mobile applications.

7 Conclusions

Modern mobile applications suffer from several usability issues, including complex interfaces, poor task support, limited interaction techniques and static navigation. These problems are aggravated on mobile devices due to the limited screen size, speed and network connectivity. This paper has shown that AUIs can be used to address some of these usability issues for mobile map-based applications. The results of the two case studies provide evidence that AUIs can improve the perceived usability, usefulness and satisfaction of mobile applications.

Information, visualization and interface adaptation can also assist in providing personalization and supporting flexibility. User modelling is needed, however, to store user preferences and record user interaction. Existing algorithms can then be used to perform the different types of adaptation.

Acknowledgements. Acknowledgements are due to Ryan Hill who developed the A-POInter system as part of his Masters research at NMMU in 2009.

References

1. Al-bar, A., Wakeman, I.: A Survey of Adaptive Applications in Mobile Computing. In: Proc. International Conference on Distributed Computing Systems (ICDCSW 2001), pp. 246–251. IEEE Computer Society, Los Alamitos (2001)
2. Alvarez-Cortes, V., Zayas-Perez, B.E., Zarate-Silva, V.H., Ramirez Uresti, J.A.: Current Trends in Adaptive User Interfaces: Challenges and Applications. In: Proc. the Electronics, Robotics and Automotive Mechanics Conference, pp. 312–317. IEEE Computer Society, Los Alamitos (2007)
3. Hussain, A., Ferneley, E.: Usability metric for mobile application: a goal question metric (GQM) approach. In: Proc. iiWAS 2008, pp. 567–570. ACM, New York (2008)
4. Looije, R., te Brake, G., Neerincx, M.: Usability Engineering for Mobile Maps. In: Proc. International Conference on Mobile Technology, Applications, and Systems (Mobility 2007), pp. 532–539 (2007)
5. Budiu, R.: Overcoming Mobile Web Usability Challenges, `http://www.forum.nokia.com/Technology_Topics/Design_and_User_Experience/User_Experience/Overcoming_Mobile_Web_Usability_Challenges.xhtml`
6. Langley, P.: User Modeling in Adaptive Interfaces. In: Proc. Seventh International Conference on User Modeling, pp. 357–370. Springer, Heidelberg (1999)
7. Jameson, A.: Adaptive Interfaces and Agents. In: Jacko, A.J., Sears, A. (eds.) Human-Computer Interaction Handbook, pp. 305–330. Erlbaum, Mahwah (2003)
8. Zukerman, I., Albrecht, D.: Predictive Statistical Models for User Modeling. User Modeling and User-Adapted Interaction (UMUAI) 11(1-2), 5–18 (2001)
9. Koelle, D.: Intelligent User Interfaces, `http://www.davekoelle.com/intint.jsp`
10. Browne, D., Norman, M., Riches, D.: Why Build Adaptive Systems? In: Browne, D., Totterdell, P., Norman, M. (eds.) Adaptive User Interfaces, pp. 16–57. Academic Press, London (1990)
11. Höök, K.: Designing and Evaluating Intelligent User Interfaces. In: Proc. the International Conference on Intelligent User Interfaces (IUI 1999), pp. 5–6. ACM Press, New York (1999)
12. Ross, E.: Intelligent User Interfaces: Survey and Research Directions. Department of Computer Science, University of Bristol (2000)
13. Nielsen, J.: Heuristics for User Interface Design, `http://www.useit.com/papers/heuristic/heuristic_list.html`
14. Opperman, R.: Adaptive User Support: Ergonomic Design of Manually and Automatically Adaptable Software. Lawrence Erlbaum Associates, Hillsdale (1994)
15. Paymans, T.F., Lindenberg, J., Neerincx, M.: Usability Trade-offs for Adaptive User Interfaces: Ease of Use and Learnability. In: Proc. the International Conference on Intelligent User Interfaces (IUI 2004), pp. 301–303. ACM, New York (2004)
16. Kühme, T.: A User-Centered Approach to Adaptive Interfaces. In: Proc. the International Conference on Intelligent User Interfaces (IUI 1993), pp. 243–245. ACM Press, New York (1993)

17. Van Tonder, B.P., Wesson, J.L.: Design and Evaluation of an Adaptive Mobile Map-Based Visualisation System. In: Proc. IFIP INTERACT 2009, pp. 839–852. Springer, Heidelberg (2009)
18. Chen, W., Chen, M.: Event detection in real time on a personal camera phone image collection. Multimedia Systems (12), 479–492 (2007)
19. Elkan, C.: Naive Bayesian Learning, Department of Computer Science and Engineering. University of California, San Diego (1997)
20. Findlater, L., McGrenere, J.: Impact of Screen Size on Performance, Awareness and User Satisfaction With Adaptive Graphical User Interfaces. In: Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI 2008), pp. 1247–1256. ACM, New York (2008)
21. Mob4Hire. Mob4Hire - Crowdsourced Mobile Application Testing, http://www.mob4hire.com
22. Hill, R.T., Wesson, J.L.: Using Mobile Preference-Based Searching to Improve Tourism Decision Support. In: Proc. SAICSIT 2008: Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, pp. 104–113. ACM, New York (2008)
23. Baus, J., Chreverst, K., Kray, C.: A Survey of Mobile Tourist Guides. In: Map-based Mobile Services: Theories, Methods and Implementations, pp. 197–216 (2005)
24. Cena, F., Console, L., Gena, C., Goy, A., Levi, G., Modeo, S., Torre, I.: Integrating Heterogeneous Adaptation Techniques to Build a Flexible and Usable Mobile Tourist Guide. AI Communications 19(4), 369–384 (2006)
25. Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., Stenild, S.: It's worth the hassle! the added value of evaluating the usability of mobile systems in the field. In: Proc. NordiCHI 2006, pp. 272–280. ACM, New York (2006)

Video Game Design for Mobile Phones

Jaime Sánchez and Matías Espinoza

Department of Computer Science, Center for Advanced Research in Education (CARE)
University of Chile
Blanco Encalada 2120, Santiago, Chile
jsanchez@dcc.uchile.cl, mespinoza@c5.cl

Abstract. This article explores the use of mobile phones for educational purposes through the design, development and evaluation of Role Playing Games (RPG). To accomplish this, we differentiated between the functionality and design of the videogame, developing a videogame engine and videogames prototypes. Thus the engine was responsible for controlling the videogame's functionalities regarding the prototypes designed. Once designed and developed, a usability evaluation of the videogames with end-users was administered. Results show that the videogames implemented were usable, easy and pleasant to use, and that they fully motivated learners as mobile learning tools.

Keywords: cell phone, mobile, RPG, videogame, problem solving.

1 Introduction

Videogames have a significant educational potential, as in addition to motivating students they allow people to learn and develop skills, abilities and strategies [3]. There are wide-ranging possibilities for applying videogames to learning. Some aspects in which videogames can contribute to learning are: knowledge of digital literacy, videogame skills through problem solving, comprehension skills and academic abilities, among others [3][2].

In an RPG (Role-Playing Game) videogame, the player takes on a specific role in a storyline through a character who lives in a virtual world. The storyline develops based on the decisions and actions taken by the player during the game. The character can gain experience, specific items or other characteristics as the game develops. Through the interaction with an RPG videogame, one can take advantage of the educational potential regarding the development of skills and the learning of content by carrying out the tasks that are assigned within the videogame.

In this way, the research objective of our work is the development of an educational, RPG videogame engine for mobile cellular phones that defines the characteristics and functionalities for the various videogames that can be designed and implemented using such a platform. This is in order to obtain a tool that supports school learning through the stimulation of the user's problem solving abilities. In order to achieve this objective, a usability evaluation of two videogame prototypes created using the educational RPG videogame engine for mobile phones was carried out.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 199–210, 2010.

In the following we present a review of the pertinent literature that was carried out based on the issues that make up the primary influences for our work. Afterwards, the engine and the videogames developed are described, and the evaluation of the prototypes is presented. Finally, we finish by establishing some conclusions based on our results.

2 Literature Review

Computers, PDAs, mobile phones and other technological artifacts all make up part of the way in which work, social relations, the creation of knowledge, learning and education have been transformed in recent decades [2]. In particular, mobile phones have been massified intensely in recent years, and the number of applications developed for these devices has increased. As a result, the general use of these applications has also become more widespread. The use of mobile phones in education can be considered as part of the evolution of the use of these devices. In addition, due to the massification of mobile phones, there is currently a student/phone ratio of almost 1:1, which represents a great opportunity to develop applications oriented towards education that are based on this kind of technology. Mobile phones are characterized by being light and portable, and are adequate for working while walking or even in uncomfortable places [4], and they can be used both inside and outside of the classroom. Despite their low processing speed, reduced screen size and unsatisfactory capacity for storage and volatile memory [4], these devices have what is needed to be able to manage information, generate collaboration and foment the construction of knowledge in both formal and informal situations, while being subject to the previously mentioned restrictions.

Through the use of an RPG videogame, learning activities such as imitation, feedback, training, practice, case reviews, incremental challenges and immersion can be carried out, among others [10]. Such activities involve the development of problem solving skills, specific content, and communication and concentration abilities [6][1][7][21][5].

It is interesting to explore the educational potential of videogames, in which the student learns from the experiences gained through the tasks that he/she carries out within the videogame, thus breaking with the traditional teaching scheme [8][15]. This points clearly to the sense of the expectations held by children today, in that they are immersed in a world that is surrounded by technology and multimedia elements, made stronger by a high degree of interactivity. These expectations are a very important factor, which makes the traditional educational system look irrelevant to the lives of today's students. In this search for new educational media, it has been sought to attract children to learning through an RPG videogame for mobile phones, contributing to an improvement in the traditional ways of learning [10][11][12][14][13].

Conceptually speaking, doing an exercise and solving a problem are not the same. In exercises, an algorithm or routine process is applied in a mechanical manner. Solving a problem means providing a coherent explanation for a set of related information in a specific context. To do this, one must pause, reflect and even execute original actions that have not previously been practiced in order to reach a solution to the problem. Polya points out that problem solving is a process that implies the consideration of 4 steps [9]: 1. Comprehension and Identification of the problem; 2. Creation of a plan to solve the problem; 3. Carrying out a plan to solve the problem; 4. Evaluation of the solution to the problem.

In the design of an RPG videogame it is convenient to follow good practices in order to construct the storyline. With regards to this aspect, our work considers some guidelines for the creation of hyperstories [16][17], as well as a contribution to the design of the software to be developed.

Finally, some previously developed videogames were taken into account as a reference for some characteristics to consider and incorporate into the proposed engine and the videogame prototypes developed. Link [18] and The Natomy's Journey [20] are RPG videogames that present a simple level of play in which the player exchanges items during the development of the videogame in order to fulfill his/her mission in the context of an educational objective. On the other hand, BuinZoo and Museo [18] are software programs for Pocket PC and Classmate computers that consist of a quiz-type videogame that are developed while at a visit to the zoo or a museum; questions on the different issues are asked, and the information necessary to answer the questions are located within the physical environment that is being visited, which is an interesting support characteristic for a mobile phone-based videogame.

3 The Engine and the Videogames

The educational, RPG videogame engine developed in this study allows for the development of videogames with different themes, but identical functionalities and characteristics. In this way, the videogames were designed for individual use, and the user can navigate freely through the virtual world that the videogames offer in the available scenarios.

The engine integrates different functionalities that characterize the videogames that were developed. Of these, the most significant are:

Elements: Visually, the videogames present the user with a virtual world made up of Scenarios, in which a Main Character (the person controlled by the user) can move about freely, and in which we can find several different elements such as Characters (not controlled by the user), Items (can be used for different purposes) and Decorations (delimit and decorate the scenarios). Other elements that are visible to the user are the dialogues between the Characters, which appear as information and/or questions referring to a problem that the player needs to solve.

Interactions: When the Main Character collides with other elements in the Scenario, the following kinds of interactions can be presented: Inform, Describe, Ask, Pick up and Trade (see table 1). An interaction only occurs when there is a collision with some element in the game, and when certain conditions are met the user can change from one interaction to another, if it is available.

Playability: The engine allows the videogames to present the users with challenges, providing them with problem-solving issues, and obligating them to make decisions.

The quality of the videogame, in this way, depends a lot on the design of the hyper-story [16]. Playability is determined by the characteristics of the Interactions together with the incorporation of problem solving aspects. Structurally, the videogame is made up of a Mission or Central Objective, which presents the player with a global problem to solve. The problem is solved when the player completes a minimum number of sequential Tasks, which are specific objectives within the videogame and which correspond to shorter-term problem-solving issues. There is a "minimum number" of Tasks because the user can simultaneously carry out other tasks in the videogame that are complementary and that are not strictly necessary in order to complete the Mission. A Task is made up of a sequence of Interactions, each of which corresponds to a node of the hyper-story.

Table 1. Possible interactions when a main character collides with the different elements

Interaction	Description	Character	Item	Decoration
Inform	The element provides information that is relevant to the videogame	Yes	Yes	Yes
Describe	The element describes itself	Yes	Yes	Yes
Trade	The element presents a problem that requires a certain item as a response, and awards another item as a prize	Yes	No	No
Pick up	The element presents a problem that requires a certain item as a response, and awards the element itself as a prize in the form of an additional item	Yes	Yes	Yes
Ask	The element asks a multiple-choice question, and awards an item as a prize for providing the correct answer	Yes	No	No

3.1 Interfaces

3.1.1 Screen

The size (in pixels with dimensions x and y) of the graphic objects on the screen is fixed, and no proportional adjustments are made when the size of the screen is smaller or larger.

The screen shows a portion of the entire Scenario, which we will denominate the clip. The size of the clip is assigned with dimensions that are the size of the display on the device, which is obtained when opening the videogame. In figure 1A we can appreciate the visible areas for different screen resolutions: Area I is for 128x128 pixel resolution, Areas I and II are for 128x160 pixel resolution, and Areas I, II and III are for 240x320 pixel resolution.

3.1.2 Entry Buttons

In order to keep the control of the actions within the videogame as simple as possible, we opted to use only joystick buttons, which are identified as: up button, down button, left button, right button and fire button (see figure 1B):

- Movement of the *Main Character*: up, down, left and right buttons.
- Continue after the *Inform and Describe Interactions*: fire button.
- Select one of the alternatives available in the *Pick up, Trade and Respond Interactions*: left and right buttons.
- Confirm a correct alternative in the *Pick up, Trade and Respond Interactions*: fire button.
- Perform a *Scroll* of the text shown in the interactions: up and down buttons.

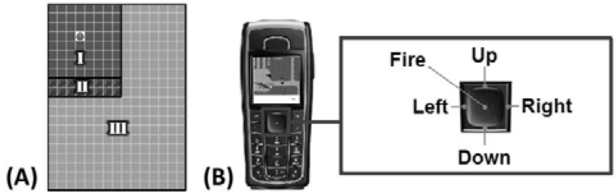


Fig. 1. (A) Clip of a scenario for different resolutions (B) Joystick buttons on a mobile phone

3.2 Prototypes

Two videogame prototypes were developed using the videogame engine, which we will call prototype 1 and prototype 2.

In prototype 1 a scenario was created with a simple story involving two characters and one item. The mission consists of capturing a Cockroach, and to do so the Main Character must previously obtain a Torch, which can be obtained without any prerequisites. There is a Tortoise that asks a question once the Cockroach has been captured.

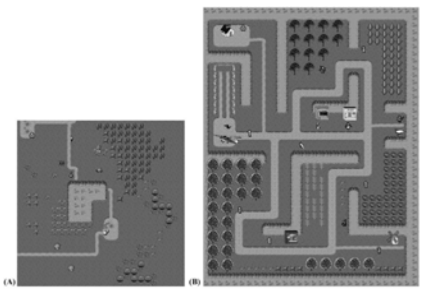


Fig. 2. Images from the maps of (A) Prototype 1 and (B) Prototype 2

In prototype 2, a visibly more extensive videogame was developed (see figure 2) with 13 characters, which involved three interactions for each one. This game developed the issue of nutrition as an educational objective. The mission for this prototype consists of getting off of an island, and to do this the player must get a pilot into an airplane. The pilot cannot get on the plane immediately because there is a dog that does not let him get near to it. Thus the pre-condition for the pilot's Interaction is to have captured the dog, and having completed this, the videogame is over. While the Main Character has still not captured the dog, the pilot presents a dialogue expressing that he cannot get close to his plane. On the other hand, the dog wants to be fed with a protein-rich food. In the videogame, three kinds of protein-rich food can be obtained: eggs, fish and meat. However, the meat is what satisfies the dog, and when he is fed a portion of meat, he is captured.

In the same way that the dog asks for his food (a protein-rich food), the other Characters ask for Items by presenting a problem consistent with each Character's need to obtain a specific food with certain nutritional characteristics. As such, the Characters provide one item in exchange for another Item, cloaking the exchange in a need presented as a problem, and providing clues for the answer but never asking for the Item by its specific name (see table 2).

Table 2. Characters and problems presented

Character	Presents a problem asking for
Baker	Base ingredient to make bread
Butcher	Cereal-based food that is rich in carbohydrates
Pilot	Free the way to the plane that is blocked by the dog
Fisherman	Vegetarian lunch rich in vitamins
Bee Farmer	-
Corn Girl	-
Lettuce Girl	Sweet concentrate food
Business Girl	Her lost hen
Apple Boy	-
Wind Mill Boy	Raw material needed to produce flour
Farmer	Sweet, protein-rich food
Dog	Protein-rich food
Hen	Cereal-type food

4 Usability Evaluation

This evaluation was focused on the characteristics of the videogame engine, using the two previously described test prototypes.

First, prototype 1 was evaluated regarding baseline characteristics of the videogame engine such as the behavior of the interactive elements and the main character's movement, as well as the use of the most complicated controls such as the text scroll and the multiple choice selection. This was done because it was necessary to detect problems in order to redesign and improve the various aspects of the application interfaces.

Second, prototype 2 was tested by evaluating all of the characteristics of the videogame. The high number of elements and interactions in this version put the playability of the videogame to the test. For this reason, it was convenient to evaluate the level of acceptance of the videogame, and to detect design problems.

4.1 Objectives

The objective consisted of detecting, correcting problems with and validating the characteristics of the videogames developed using the videogame engine designed, through the evaluation of the design of the main elements, the level of playability, the use of interactive material in general, the use of controls in the main character's movements, the text scroll in the dialogues and in the selection of the multiple choice alternatives.

4.2 Participants

There were two groups of students from different grade-levels of general, elementary school education, with ages that fluctuated between 9 and 11 years old. Group one was made up of nine students (three girls and six boys), all of who had experience with the use of mobile phones. Group two was made up of ten students (nine girls and one boy), all of which also had previous experience using mobile phones.

An engineer who had participated in the development of the engine and the videogames acted as an evaluator during the evaluation of the prototypes.

4.3 Instruments

Observation questionnaire 1: seeks to obtain a record of the actions associated with the interactions with the main elements, with which it is sought to detect problems regarding the playability and the use of the controls of the videogame. The actions recorded were: if the student asked what to do with each element, if in interacting with each element the student used the text scroll, if he/she pressed the *Fire* button to close the dialogue with the element, if the student pressed the *Fire* button for another action that had not been implemented, and if he/she performed the expected action with each element.

User feedback questionnaire: this questionnaire was inspired by some of the open-ended questions from the end user evaluation questionnaire for blind students [15]. In this case, the questions were centered on the elements of the prototype, asking the students if they liked each one, with possible answers of: yes, so-so, and no. In addition, some open-ended questions were incorporated, which can be seen in the results section. With these questions it was sought to evaluate the users' opinions on the design, playability, the use of the controls and the interactive material of the videogame.

End user evaluation questionnaire: consists of an adaptation of a summarized evaluation questionnaire regarding software for blind students [21], in which the specific questions designed to measure the interpretation of the sounds were replaced with questions that seek to measure the interpretation of the images. The original questions, "I like the software's sounds", "The software's sounds are clearly identifiable" and "The software's sounds provide me with information", were replaced with the questions "I like the videogame's images", "The videogame's images are clearly identifiable", and "The videogame's images provide me with information", respectively. In addition, within this questionnaire the word "software" was replaced with "videogame" wherever it appeared. This questionnaire also includes a section with open-ended questions. Through the use of this questionnaire, it is sought to evaluate the quality of the videogame based on the user's satisfaction regarding issues of design, playability, the use of controls, and the interactive material of the videogame.

Observation questionnaire 2: with the use of this questionnaire it is sought to measure the use of the functionalities, in which the evaluator answers the following questions with Yes or No regarding the user's performance: The user correctly uses the movement controls, The user navigates the map, The user interacts with the characters, The user interacts with the decorations, The user at some point uses the text scroll, The user selects the multiple choice alternatives, The user obtains the items he/she is looking for, The user understands the relation between the item sought after and the character that needs it. These questions seek to evaluate and validate the controls, the playability of the application and the interaction with the elements of the videogame.

4.4 Procedure

The usability tests were separated into two different moments, first to test prototype 1 with the application being developed and then to test prototype 2 with the final application including the changes redesigned and incorporated as a result of the first evaluation. In order to perform the usability tests, the necessary copies of the previously described instruments were printed out on paper. In addition, for both of the evaluations a Nokia 6230 mobile phone was used, with both prototype 1 and prototype 2 installed.

First the usability evaluation was applied individually in group one in the following steps: 1. Verbal instructions on what was to be done in the videogame and on the controls for the videogame were provided to each student; 2. The interactions performed in the videogame were recorded in the observation questionnaire 1; 3. When a student had doubts on what he/she had to do, instructions were provided that would allow him or her to continue playing; 4. Finally, the user feedback questionnaire was applied verbally. The evaluator read the questions out loud, and the answers were recorded on the same sheet.

Afterwards, the usability evaluation was applied individually in group two using the following steps: 1. Verbal instructions were provided to each student regarding what had to be done in the videogame, and on the controls for the videogame; 2. Each student played for about 7 minutes; 3. The student's performance in the different functionalities were recorded in the observation questionnaire 2; 4. When a student had doubts on what he/she had to do, instructions were provided that would allow him or her to continue playing; and 5. After the 7 minutes of playing time had passed, the students were invited to fill in the end user evaluation questionnaire.

4.5 Results

The results for the initial tests showed that the interaction with the videogame environment was highly intuitive, as very little was asked about what they had to do, and in general they were able to discover things in the videogame through exploration alone. Because of this, it was not necessary to add special guides within the videogame regarding this aspect.

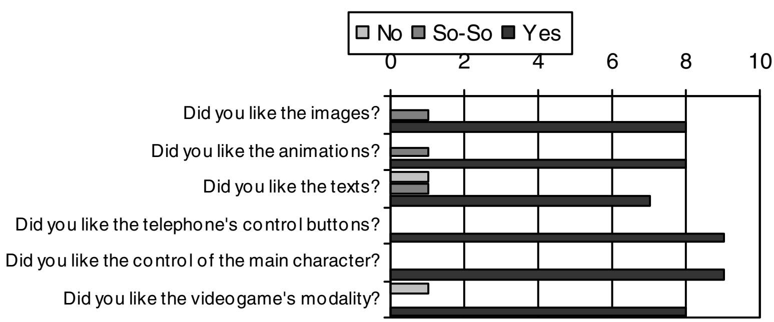


Fig. 3. Summary of the results of the user feedback questionnaire

The entry buttons for the videogame's controls and for the main character's movements were unanimously accepted (see figure 3), which provides support for the choice adopted in the videogame design. However, the use of the controls for the text scroll and the selection of the multiple-choice alternatives (interaction with the Tortoise) were not intuitive, as no users performed the text scroll and only two users were able to execute the action for the selection of the multiple-choice answer. For this reason, in the prototype 2 we avoided using extensive dialogues that would make it necessary to use the text scroll, and guides were added that indicated that the left and right buttons could be used to perform the selection of the multiple-choice alternatives.

The interactive elements (images, animations and texts) were well received by the students (see figure 3), for which reason we maintained the same line with regards to this aspect.

From the answers to the open-ended questions of the user feedback questionnaire, we were able to conclude that the students understood that the objective of the videogame consists of capturing the cockroach with the torch. This is good because it implies that they were able to identify a basic aspect of the game's playability, namely obtaining and using things in order to obtain other things. When asked what the videogame could be useful for, the students did not perceive any educational utility; they only expressed that it would be useful for entertainment, which presents a challenge for the second prototype in order to improve the sensation that the game leaves the users with. Also, in answering these questions the students suggested adding more trees, plants, bushes, vegetation and other elements in order to confuse the user, to be able to change characters, and increasing the difficulty of the game. This means that they wanted more interactive elements in the videogame. Faced with this scenario, for prototype 2 a more wide-ranging hyperstory was proposed, with more elements and a higher degree of complexity regarding the solution to the game, incorporating a baseline educational framework.

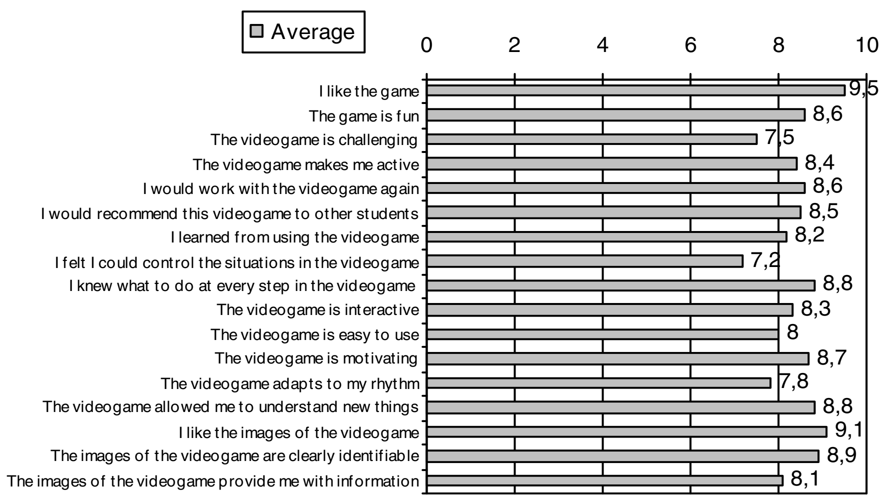


Fig. 4. Results of the end user evaluation questionnaire, showing the average percentage of each indicator on a scale of 1 to 10 points

From the second usability tests, the results obtained from the application of the observation questionnaire 2 show that, in general, the videogames' functionalities are appropriately used by the students, according to the criteria of what was expected by the videogame designer. However, the text scroll continues to be a low point, and it is necessary to seek new ways in order to improve the use of this function.

The results collected from the end user questionnaires (see figure 4) are very good for each of the indicators of acceptance, which validates the choices regarding design, playability, and control of the videogame, in addition to the materials used.

In the answers to the questions included in the end user acceptance questionnaire, the students were shown to like various aspects of the videogame, such as "looking for the things that people ask for and need", "giving food to the dog", "providing other people with things", "the adventure format of the videogame", "the characters", "the playing fields", "helping people", and "it is fun". They also expressed their opinions on the elements of the videogame that they did not like, such as "when it was hard to find the things that people asked for", "there were no instructions", "there was no friend for the main character", and "people asked for too many things". The students suggested adding various aspects to the videogame, such as: "the characters should talk more", "the game should be bigger", "more instructions", "more characters", "more problems", "a character that helps the main character", "a pretty girl", "more things to eat" and "more elements that can do other things". They also expressed their general opinions on the videogame, with statements such as, "I liked the videogame", "it is very good", "the videogame was incredible", "it was very good", and "it is recommendable". These results show that the students had very positive feelings for the videogame, its playability and the elements that were available to them. However, the students demand a more wide-ranging and complete videogame, with more characters, more dialogues, more items and new functions, in addition to improving the videogame's instructions. When asked what the videogame could be useful for, the students answered, "to know about different harvests", "to know more about animals", "to learn", "to learn to look for things", "to learn to navigate the different parts of the videogame", "to teach", "to help", and "to discover new things". This represents a very interesting result, in that the students clearly perceive that the videogame is related to learning activities and different kinds of objectives, which represents notable progress when compared to prototype 1.

5 Conclusions

In our work we successfully and satisfactorily fulfilled our objectives to design and implement an educational, RPG videogame engine. This engine offers a set of functionalities that allow for an educational, RPG videogame to be developed within it. The engine was developed and used for the creation of prototypes 1 and 2, which were tested with end users.

Another one of the proposed objectives was to support the development of the application through the usability evaluation of the educational, RPG videogames during the different stages of development. With this we were able to improve the design of the application as it was being developed, and at the same time justify design and implementation decisions that were supported by the results of the evaluations.

The results of the usability evaluations show that the students adopted the interaction with the different elements of the videogame naturally. They were able to explore the virtual world and the possibilities offered by the videogame. This is supported by the simplicity of the choice for the videogame's controls, which represented a strong point in the evaluations regarding the facility of its use, with the exception of the text scroll, for which an acceptable improvement is still pending. Together with this aspect, the playability was also well received by the students, who understood that the dynamic of the videogame consists of providing things to the different Characters and receiving other things in exchange, being able to effectively stimulate their problem-solving abilities. Finally, we add that in general the students associated the videogame with being useful for learning-related activities. Thus, from these results we can infer that this educational, RPG videogame engine can be exploited for the development of videogames that support school learning through the stimulation of the user's problem-solving abilities.

The students' opinions from the application of the end user evaluation questionnaires showed a very high level of acceptance of the videogame, validating the choices regarding design, playability, the control of the videogame, as well as the materials used. This translates into a high level of motivation to use this tool for learning.

It is important to point out that the sample of users utilized to apply the end user evaluation questionnaire is not representative in terms of quantity. In this sense, we propose future work to evaluate such games with more users in order to assure the acceptance of the applications developed.

In addition, the usability evaluations allowed us to propose substantial improvements to the functionalities of the videogame engine. This allows the videogames developed to be more appropriate regarding the interests and ways of interacting of the end users.

Although the evaluation of the videogame engine was not the initial focus of our work, this could be an interesting aspect to evaluate in the future with a small sample of videogame designers and developers. However, the design, implementation and evaluation of the characteristics of the videogame engine used in the two prototypes involved in this project, represent guidelines that can be used by the designers and developers of educational RPG videogames for mobile phones.

We also propose future work that includes a cognitive test of the application, in order to be able to measure the effect that the use of this application has on learning. The videogame engine provides for certain aspects of learning through the available functionalities, such as the construction of questions and the presentation of information with specific content. To deal with these aspects, we intend to perform a future evaluation of the level of significance regarding the impact that the videogame engine has on users' learning. With this we would be able to completely sustain our main objective regarding the development of a tool that supports school learning through the stimulation of the user's problem-solving skills.

Acknowledgments. This report was funded by the Chilean National Fund of Science and Technology, Fondecyt #1090352 and Project CIE-05 Program Center Education PBCT-Conicyt.

References

1. Eriksson, Y., Gärdenfors, D.: Computer games for children with visual impairments. In: Proceedings of the 5th International Conference on Disability, Virtual Reality and Associated Technologies, New College, Oxford, UK, September 20-22, pp. 79–86 (2004)
2. Gros, B., Bernat, A.: Videojuegos y aprendizaje. Graó, Barcelona (2008)
3. Gros, B., Aguayos, J.: Pantallas, juegos y educación: La alfabetización digital en la escuela. Desclée de Brouwer, Bilbao (2004)
4. Guerrero, L., Ochoa, S., Pino, J.: Selecting Computing Devices to Support Mobile Collaboration. Group Decision and Negotiation 15(3), 243–271 (2006)
5. Kelly, H., Howell, K., Glinert, E., Holding, L., Swain, C., Burrowbridge, A., Roper, M.: How to Build Serious Games. Communication of the ACM 20(7), 45–49 (2007)
6. Klopfer, E., Yoon, S.: Developing Games and Simulations for Today and Tomorrow's tech Savvy Youth TechTrends. Linking Research & Practice to Improve Learning 49(3), 33–41 (2005)
7. McDonald, K., Hannafin, R.: Using web-based computer games to meet the demands of today's high stakes testing: A mixed method inquiry. Journal of Research on Technology in Education 55(4), 459–472 (2003)
8. McMichael, A.: PC Games and the Teaching of History. The History Teacher 40(2), 203–218 (2007)
9. Polya, G.: How to solve it. Doubleday and Co., Inc., Garden City (1957)
10. Prensky, M.: Digital Game-Based Learning. McGraw-Hill, New York (2001)
11. Prensky, M.: Digital Natives. Digital Immigrants Part 1: On The Horizon - The Strategic Planning Resource for Education Professionals 9(5), 1–6 (2001)
12. Prensky, M.: Digital Natives. Digital Immigrants Part 2: Do They Really Think Differently? On the Horizon 9(6), 1–6 (2001)
13. Prensky, M.: Engage Me or Enrage Me: What Today's Learners Demand. Educause Review 40(5), 60, 62,64 (2005)
14. Prensky, M.: Listen to the Natives. Educational Leadership 63(4), 8–13 (2006)
15. Proserpio, L., Viola, D.: Teaching the virtual generation. Academy of Management Learning & Education 6(1), 69–80 (2007)
16. Lumbreras, M., Sánchez, J.: Hyperstories: A model to specify and design interactive educational stories. In: En Baeza, R. (ed.) Computer Science, USA, pp. 135–146. Editorial IEEE Computer Society, Los Alamitos (1997)
17. Sánchez, J., Lumbreras, M., Bibbo, L.: Hyperstories for learning. In: En Nanard, M. (ed.) Workshop in Computing, Hypermedia Design. Springer, Heidelberg (1996), ISBN 3-540-199-85-3
18. Sánchez, J., Elías, M.: Science Learning in Blind Children through Audio-Based Games. In: Redonde, M., Bravo, C., Ortega, M. (eds.) Engineering the User Interaction: From Research to Practice, pp. 87–102. Springer, London (2008)
19. Sánchez, J., Mendoza, C., Salinas, A.: Mobile serious games for collaborative problem solving. In: Brenda, K. (ed.) Wiederhold and Giuseppe Riva (Editors) the Annual Review of Cybertherapy and Telemedicine 2009. Studies in Health Technology and Informatics (SHTI) Series, vol. 144, pp. 193–197. IOS Press, Amsterdam (2009)
20. Sánchez, J., Sáenz, M.: Video Gaming for Blind Learners School Integration in Science Classes. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009, Part I. LNCS, vol. 5726, pp. 36–49. Springer, Heidelberg (2009)
21. Sánchez, J., Salinas, A., Sáenz, M.: Mobile Game-Based Methodology for Science Learning. In: Jacko, J.A. (ed.) HCI 2007, Part IV. LNCS, vol. 4553, pp. 322–331. Springer, Heidelberg (2007)
22. Sanchez, J.: End-user and facilitator questionnaire for Software Usability. In: Usability Evaluation Test. University of Chile, Santiago (2003)

EPISOSE: An Epistemology-Based Social Search Framework for Exploratory Information Seeking

Yuqing Mao[1], Haifeng Shen[2], and Chengzheng Sun[1]

[1] School of Computer Engineering
Nanyang Technological University
Block N4, Nanyang Avenue, Singapore, 639798
{maoy0002,czsun}@ntu.edu.sg
[2] School of Computer Science,
Engineering and Mathematics
Flinders University, Adelaide, Australia
hfshen@csem.flinders.edu.au

Abstract. Search engines are indispensable for locating information in WWW, but encounter great difficulties in handling exploratory information seeking, where precise keywords are hard to be formulated. A viable solution is to improve efficiency and quality of exploratory search by utilizing the wisdom of crowds (i.e., taking advantage of collective knowledge and efforts from a mass of searchers who share common or relevant search interests/goals). In this paper, we present an epistemology-based social search framework for supporting exploratory information seeking, which makes the best of both search engines' immense power of information collection and pre-processing and human users' knowledge of information filtering and post-processing. To validate the feasibility and effectiveness of the framework, we have designed and implemented a prototype system with the guidance of the framework. Our experimental results show that an epistemology-based social search system outperforms a conventional search engine for most exploratory information seeking tasks.

Keywords: Exploratory Search, Information Seeking, Social Search, Search Epistemology, Collaborative Search.

1 Introduction

Web search has become the best source of information for many people. The following factors are relevant to the success of a web search: 1) to provide information that can be easily accessed (by information providers), 2) to formulate precise keywords that can express search goals (by information consumers), and 3) to rank search results according to their relevance to the search goals (by search engines).

To information providers, some research on search engine optimization (SEO) [10] has been done to achieve easy accessibility of their information. However, the abuse of SEO can cause information overload by search engines.

To search engines, the ultimate goal is to understand the goals of information consumers and to return what they want. Unfortunately, the processing of natural languages and the deployment of semantic web still has a long way to go. Researchers

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 211–222, 2010.

are currently focusing on improving the intelligence of search engines in order to rank the results pertinent to a search query. For example, personalized web search systems collect users' profiles together with users' queries to measure similarities with techniques such as collaborative filtering for web search [2]. On the other hand, studies have shown that nearly 40% of queries were repeated [11]; a lot of work has been done to re-use previous successful queries. For example, one will see the keyword suggestions from a search engine like Google when she/he types an incomplete keyword in the search box. Queries can also be enhanced and modified for query expansion by the analysis of similar query logs in a search engine [1].

To information consumers, field knowledge and search skills are essential to a successful web search task. As individual users' knowledge and skills are limited, they can complement each other by turning to social search, which utilizes "the wisdom of crowds" [4]. As opposed to algorithm-based searches, human-labor-based searches are collectively called "social search". Representative examples include ChaCha[1], which offers live expert assistance to guide ordinary users in performing search tasks and SearchTogether [8], which provides a platform for a group of people to communicate and share search results in real time. As collaborative tagging has grown in popularity on the web, social media and social annotations have also attracted a lot of attention. Successful examples include question and answer systems like Yahoo! Answers[2], social annotations like Mahalo[3], and websites for discovering expert in social networks like Google Adarvark[4]. Social search has changed a search process from an individual activity to a social one. It has been proved effective in information filtering (e.g., understanding users' search goals) and helpful in information post-processing (e.g., annotating search results).

Although these social search systems are good complements to conventional search engines, they are generally inadequate for supporting effective "exploratory search" [7], where a search process starts with a vague goal rather than precise keywords, and the information consumer is not clear about what search results she/he is expecting. She/he can only evaluate every search result by measuring its relevance to her/his search goal. This is mainly due to the fact that information providers are not actively involved in search processes to help information consumers easily access the information they have provided. According to Heymann et al. [3], although some information providers provide information not currently available from other sources (e.g., answers to a question), they need to make extra efforts to make the information easily searchable by information consumers.

Our solution to supporting effective exploratory search is based on the effective sharing of "search epistemology" – the knowledge acquired by users in information collection, pre-processing, filtering and post-processing – among a social search community. Search epistemologies are aggregated and well-structured information packages derived from successful search processes, such as queries, results, rankings, annotations, comments, and evaluations. In this solution, information consumers are

[1] http://www.chacha.com
[2] http://answers.yahoo.com
[3] http://www.mahalo.com
[4] http://www. vark.com

also information providers, who are actively involved in a social search process to help get their contributed information easily accessed by others.

In this paper, we present an Epistemology-based Social Search (EPISOSE) framework for designing exploratory information seeking systems. The framework leverages existing algorithmic search engines to collect and pre-process information and human users to construct search epistemologies by filtering and post-processing information, and allows a social search community to effectively share their intimate search epistemologies. The EPISOSE framework can be applied to the design and implementation of a range of social search systems with different strategies and algorithms. To validate the feasibility and effectiveness of the framework, we have designed and implemented a prototype system *Baijia* with the guidance of the framework and conducted a set of experiments to measure the system's performances in supporting exploratory information seeking.

The rest of the paper is organized as follows. Section 2 describes the architecture and components of the EPISOSE framework. Section 3 presents an implementation of the framework – the *Baijia* prototype system, and Section 4 discusses the setting and results of experiments. Finally, Section 5 concludes the paper with a summary of major contributions and future work.

2 The EPISOSE Framework

Contrary to the linear relationship between a user and a search engine in current web search systems, one of EPISOSE's distinctive characteristics is a positive feedback cycle: information consumers are also information providers so that their search epistemologies can be contributed to subsequent search processes for the benefit of other consumers. As depicted in Fig. 1, the EPISOSE framework consists of the following major components.

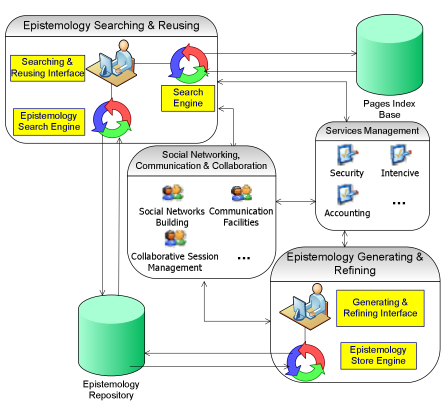


Fig. 1. The EPISOSE Framework

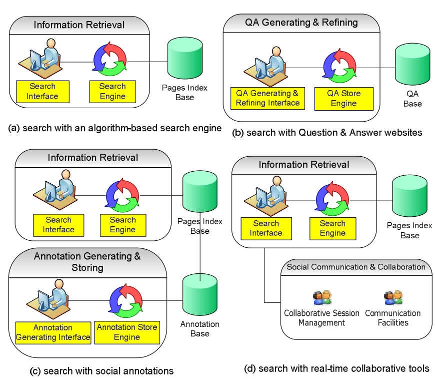


Fig. 2. The A comparison of EPISOSE with other Web search frameworks

Epistemology Searching and Reusing: this component is for a social search community to share and reuse search epistemologies. While a user types a query through the *Searching & Reusing Interface*, the *Epistemology Search Engine* will first search the epistemology repository and return the relevant search epistemologies. These epistemologies were contributed by other users with the same or relevant search interests or goals through the *Generating & Refining Interface* powered by the *Epistemology Store Engine*. If no relevant epistemology is found, the *Search Engine* (e.g., Google) will search the *Pages Index Base* and return relevant pages according to the keywords. Users can generate their own epistemologies from the result pages returned by the search engine or refine others' relevant epistemologies retrieved from the epistemology repository. In this component, knowledge discovery techniques (such as classification or clustering) can be applied to analyze the relevance of the input queries to the stored search epistemologies in the epistemology repository. Because each epistemology is tagged with plenty of additional information by processing raw information returned by a search engine with the user's intimate knowledge and understanding, the *Epistemology Search Engine* has a very high probability to return relevant epistemologies pertinent to a query. Consequently, other users can save their time in repeating the course of collecting and processing raw information through a search engine.

Epistemology Generating & Refining: this component is for users to easily generate new search epistemologies or refine existing ones, and store them into the *Epistemology Repository* through the *Epistemology Generating & Refining Interface* powered by the *Epistemology Store Engine*. A search epistemology includes the subsequence of search keywords, the approbatory results selected by the user, the commentaries on these results added by the user, other useful information about the search topic provided by the user, and evaluation of the search provided by other users.

This component is significantly different from answering questions or adding annotations because stored search epistemologies can be easily retrieved by *Epistemology Searching & Reusing* without relying on search engines, which have difficulties in retrieving most relevant information if precise keywords are hard to be formulated. Since epistemologies are generated during exploratory information seeking processes, the retrieval of epistemologies would be a heuristic search. Users can be quickly led to the final search goal by the epistemologies generated from others' successful searches. Furthermore, as existing epistemologies can be refined by subsequent users, they tend to be more relevant and accurate as they are refined by more users.

Epistemology Repository: this component stores search epistemologies contributed by the users in a social search community. Search epistemologies are information packages regarding specific search goals. They are indexed by the combination of the search goal and relevant keywords for the sake of easy retrieval by the *Epistemology Search Engine*. The *Epistemology Repository* also stores extra information about epistemologies for building social search communities. For example, building a social network needs the information about who contribute relevant epistemologies.

Social Networking, Communication & Collaboration: this component has the following functions: *Social Networks Building* helps users with the same or similar search goals build up social networks to complete search tasks together. While users are doing exploratory information seeking, they would be likely to look for help from others. Users can find people with same hobbies or similar information requirements from search epistemologies and thence build a social network with them.

Furthermore, the EPISOSE framework can adopt effective strategies to search for expertise in social networks, so that exploratory searches can benefit from the building of social networks. When one is not sure about what she/he is looking for, seeking advices from experts is always a good option.

Communication Facilities allow users in a social search community to communicate via tools such as messenger or email. *Collaborative Session Management* allows a group of users to share their search epistemologies in an ongoing search process synchronously or asynchronously. While users are doing exploratory information seeking, they may invite others to work with them together. Users can discuss with the contributors of certain epistemologies in order to better refine them, or invite buddies in their social networks to join their ongoing search processes. *Services Management* provides some common services for making exploratory social search viable, reliable, and sustainable. For example, *Incentive* encourages users to share their epistemologies, *Security* handles issues related to privacy and security in a social search community, and *Accounting* can estimate users' contribution to establish a profit model for social search.

EPISOSE is uniquely distinguished from current web search frameworks listed in Fig. 2(a)-(d). Fig. 2(a) shows the conventional method of web search with a search engine. In these systems, the *Search Engine* is an algorithm-based search engine such as Google. Search knowledge cannot be shared since there is no knowledge generation and storage mechanism. Peer to peer search engines can support each peer to publish its local documents (or local index), but the purpose is to build efficient topic-specific search engines rather than supporting general exploratory search through knowledge sharing. Fig. 2 (b) shows the method of web search with Question & Answer websites, e.g. asking for help in forums or Yahoo! Answers, which is also known as the man-powered search engines. Users can share knowledge through those social web sites, but knowledge sharing is not effective as shared knowledge still has to be retrieved by conventional search engines. Fig. 2(c) shows the method of web search with social annotations, such as Yoople[5], which allows people to publish their attitudes toward certain search results through voting or editing. But annotations are not as rich in content as search epistemologies (annotations are only an integral part of epistemologies) and sharing of annotations is not effective as they have also to be retrieved by search engines. Fig. 2(d) shows the method of web search with real-time collaborative tools such as SearchTogether [8], which supports search and sharing of Web pages with others through communication facilities. Searched Web pages cannot be shared to people who are not invited or online when the search process is ongoing. Moreover, it shares only Web pages rather than a package of knowledge about a specific topic.

The primary goal of a social search system for exploratory information seeking is to utilize existing successful searches. Many people have searched for the same topic, for example, wedding planning, but they couldn't re-use previous successful searches because they were not shared or fell into oblivion. Yet it's not true that people are unwilling to share the knowledge gained from their searches. The problem is that they couldn't find convenient and effective ways to share it. As a matter of fact, many people write their knowledge of wedding planning in blogs or discussion forums after effortful web searches. However, only a small group of people in a specific period of

[5] http://www.yoople.net

time may benefit from such kind of knowledge sharing, while others still need to spend tremendous time repeating the process of searching for the same topic.

The EPISOSE framework addresses this problem as follows. First, it provides a social search platform and interfaces for a wide range of users to share their search epistemologies directly. Second, search epistemology is based not only on a rich set of objective results returned by a search engine, but also on a user's subjective judgment and intimate knowledge. It is far beyond just a set of search keywords and a list of linear results. Third, searching and sharing are seamlessly integrated in the framework so that users can enrich their knowledge about a search topic and improve their search skills by learning from their peers. It is worth pointing out that an epistemology-based social search system is self-reinforcing in the sense that epistemologies tend to be more relevant and accurate as they are refined by more users.

3 Implementation

To validate the EPISOSE framework, we have designed and implemented a prototype system *Baijia* (see Fig. 3 & 4), which allows a social search community to share their search epistemologies.

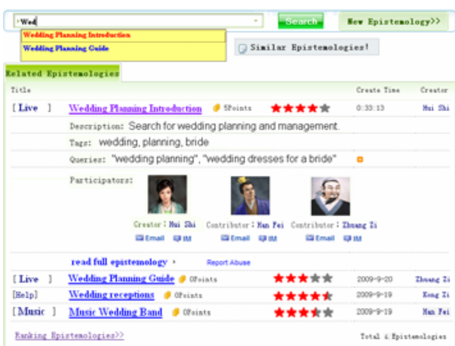


Fig. 3. The *Baijia* system interface

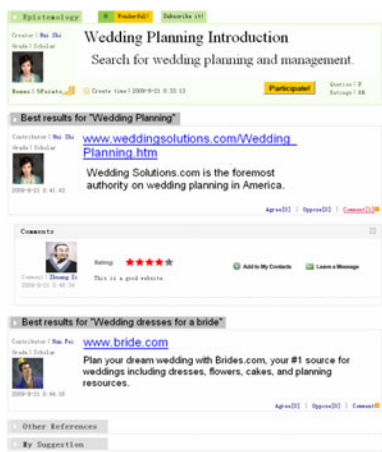


Fig. 4. Epistemology generation and refinement

3.1 Data Collection

We selected the AOL query logs [9] as the initial dataset for the *Baijia* system. The AOL query logs consist of about 20 million search queries from about 650,000 users. Although the dataset doesn't contain explicit user's feedback on the search results, the URL clicking can be regarded as the implicit positive feedback because relative feedback signals generated from the search engine users' clicking behaviors have been proved to correspond well with explicit judgments [6].

The dataset includes {AnonID, Query, QueryTime, ItemRank, ClickURL}, where AnonID presents an anonymous user ID number, ClickURL is the URL user clicked and ItemRank is the rank of the clicked item on the listed results.The initial epistemology repository of *Baijia* is built by importing the dataset and the search epistemologies are automatically generated for each user. For some selected topics, search epistemologies are clustered by keywords of the queries, and formed by specified rules, e.g., different weights are given to URLs according to the number of click times for the same queries. With the initial epistemology repository, we can also setup experiments to evaluate the effectiveness of *Baijia* system by contrasting the new search processes with those in the query logs.

3.2 Searching and Sharing

Take a search of wedding planning as an exploratory information seeking example. A user (AnonID=19913) begins with the keyword “wedding planning”. After the user inputs “wedding planning” in the concise search interface (Fig. 3), the system will return a linear list of results through the common API of a search engine (currently the Google AJAX Search API). If previous search epistemologies exist, a re-ranked results list will be returned. The details of these results show who have contributed to the search epistemologies and how do other users evaluate these epistemologies, which will help the user judge their relevance.

After browsing a few pages of the results, the user regards the page from “www.weddingsolutions.com” as the best page about the solutions of wedding planning and drags the item from “search results” to “best results”. The user can also add some personal comments on this result, e.g., “This is a good website about wedding planning”. The interfaces for search and sharing are presented in one page so that users can easily drag and drop items and read and write comments without any popup windows.

Having gained the basic knowledge about wedding planning, the user may continue the search for “wedding receptions” if she/he considers having a wedding reception (AnonID=884092), or “wedding dresses” if she/he wants to get some good suggestions about the dressing in the wedding ceremony (AnonID=5761104). She/he can choose “share epistemology” to publish the epistemology after completing the whole search process.

She/he may entitle the epistemology as “Wedding Planning Introduction” and classify it into a “successful search” or “partially successful search” category. She/he can even supplement the search with her/his own epistemology: “All those preparations are not enough and I think there should be…” (Fig. 4).

When another user also starts exploratory information seeking about wedding planning with the keyword “wedding planning”, she/he will get a list of related epistemologies from others (Fig. 3). The relevant search epistemologies can be collected by filtering according to the similarity to the query. For a user who knows little about wedding planning, what has been searched by others is a good starting point for her/him. Hence, the user can browse the list and dig out the details of those in which she/he is interested, e.g. “Wedding dresses for a bride”, if she/he (e.g. AnonID=16852248) wants to know more about wedding dresses. The user can also get help from others’ epistemologies if she/he is not sure about what she/he is searching for. For example, the user (AnonID=12199341) has only a vague idea about “how can

I write the wedding invitations”. After browsing the list, she/he finds that many others are looking for “wedding invitation wording”, which effectively suggests her/him how to start the search.

Furthermore, a user can learn search skills from others’ epistemologies. For example, when a user (AnonID=2200929) wants to search information about a band playing music in the wedding ceremony and she/he uses the keyword “wedding band”, she/he would be puzzled since all results returned by a search engine in the first page are all about the wedding rings. If a previous successful search epistemology named “Music Wedding Band” has been generated in the epistemology repository, the user can easily find relevant information from the related epistemologies list. Meanwhile, she/he knows how to formulate keywords and to filter out unwanted ones. Such knowledge will probably help accelerate her/his search process.

The user can continue her/his search on the basis of these searches, add new good results and commentaries or remove expired or faulty information. If a page is dynamic, she/he may add annotations or modify the link with parameters to ensure it will not expire. The user can also evaluate others’ epistemologies, e.g., give positive or negative feedback. All search epistemologies will be accumulated in the epistemology repository for further utilization.

3.3 Epistemology Search Engine and Epistemology Repository

The epistemology search engine derives the relevance between search epistemology and a search query. For example, if the “www.weddingsolutions.com” page of wedding has appeared as the best result in both the search epistemologies of “Wedding Planning Guide” and “Traditional Wedding Planning”, we are likely to draw the conclusion that it is relevant to the query about “wedding planning”. Besides, some methods such as feature extraction and clustering will be employed to accurately discover more relevant search epistemologies to certain search goals.

For conflicting search epistemologies, the reputation and the expertise of users can be used to help make the judgment. One’s reputation will increase/decrease when one’s published search epistemologies receive positive/negative evaluation by others. This will discourage users from publishing misleading or irresponsible search epistemologies for malicious purposes or undeserved reward points. The epistemologies published by users with better reputation will be assigned heavier weights

Since all search epistemologies are stored in the epistemology repository, its volume grows fast with time. Therefore, it is important to build suitable indices for search epistemologies in order to improve the retrieval efficiency and accuracy. Moreover, with the epistemology repository, we can also support semantic epistemology retrieval by building various ontologies. For example, the “band” would be modeled with the meaning of “ring” for an ontology about the domain of wedding, but with the meaning of “instrumentalists” for another ontology about the domain of music. As such, our system could process users’ requests more accurately based on the context of their queries.

3.4 Communication and Personal Information

Baijia allows users to communicate with the system and with each other via email or a web-based instant messenger (Fig. 3). It can support synchronous sharing of search epistemologies among users and allow users to communicate with other instant messaging users through gateways of the jabber server so that users can get instant help from their online buddies or experts during a search process. For example, the bride who is searching for the wedding planning may have difficulties in choosing the dress to wear in the wedding, she can immediately ask her online buddies or parents to generate their search epistemologies for her or modify her rudimentary epistemologies.

Users can manage personal information through a control panel. Security settings can also be done here. For privacy-sensitive information, a user can choose to share within a small group, e.g., family members, or not to share the search epistemologies at all. It is helpful in some cases, e.g., colleagues searching for new techniques in order to get a project done. All members can take part in the search task and share epistemologies with each other. Users are motivated to share or evaluate search epistemologies by an incentive mechanism based on accumulated points. Besides, a user can subscribe to an interested search epistemology in order to receive notifications when it is updated (Fig. 4).

4 Experimental Evaluations

We have done some experiments to evaluate to what extent the *Baijia* system can improve exploratory information seeking as compared to a conventional search engine (e.g., AOL). We plan to conduct user evaluations after the system has accumulated adequate epistemologies.

4.1 Experiments Setup

For the sake of simplicity, clicked URLs are regarded as relevant search results pertaining to the corresponding queries in our experiments. It is a reasonable hypothesis since clicked URLs must have attracted users' attentions although they might not necessarily be good matches. More importantly, it is consistently applied to both the *Baijia* system and the AOL search engine; therefore, the comparison results should be fair and conclusive.

In our experiments, search epistemologies are contributed and shared through the following steps:

Step 1: "Users" completed their searches through iterative interaction with the system and contributed their search epistemologies. To simulate the contribution from users, we extracted every user's exploratory searches from their queries. Each exploratory search contains several queries that are contextually related. Cosine distance function is used to measure the contextual similarity between every two queries. We have totally extracted 1,201,497 exploratory searches.

Step 2: The system returned other users' search epistemologies that are relevant to the current user's queries from its epistemology repository. To simulate the sharing of epistemologies, we retrieved the epistemology repository for relevant search epistemologies. An epistemology is relevant to an exploratory search if its queries are similar to the search queries, and the selected pages of the epistemology completely/partially match the clicked URLs of the search.

Step 3: If no relevant epistemologies are found at step 2, the exploratory search itself will be formulated as a search epistemology; otherwise, it will be integrated into existing relevant epistemologies. “Users” participated in the search activity by re-ranking the re-ranked results from other users or the ranked results from the search engine. To simulate the refinement of epistemologies, a random score is assigned to every clicked URL to represent the judgment from the current user. Actually it is common that users may have different opinions on the same search result. As the motivation of social search is to believe the wisdom of crowds, the result that is ranked highest by the majority of searchers is regarded as the best. For a social search system, a ranking mechanism that is based on the average scores of all participators‘ follows this rationale. This ranking mechanism is adopted by the *Bajia* system, which is adaptive to accumulative users’ rankings, no matter whether they are computer-generated random scores or real human evaluation scores. Therefore, the random score assignment serves our purpose for the experiments and we could envisage an even better performance gain if real human scores were used in the future experiments. In addition, URLs that are repeatedly clicked are given higher scores. The selected URLs of every epistemology are re-ranked according to the scores. We have finally built 480,254 records in the epistemology repository.

Following the above steps, we have build up the initial epistemology repository for our system by importing all exploratory searches derived from the AOL query logs. It is worth pointing out that the initial epistemology repository can immediately benefit new exploratory information seekers, but the system can actually work without it. The system relies on the search engine to build up the epistemology repository at its initial stage and gradually relies more on the epistemology repository itself.

Several metrics have been adopted to evaluate *Baijia*’s performance, such as Mean Average Precision (MAP) [13], Precision at K (e.g., Precision@10) [13], and Normalized Discounted Cumulative Gain (NDCG) [5]. In this article, we present our experimental results based on MAP, which shows the overall performance of *Baijia* is superior to that of AOL search engine in exploratory information seeking.

Average precision of a query is defined as follows:

$$AP = \frac{1}{\mathrm{Re}\,l}\sum_{r=1}^{\mathrm{Re}\,l}\frac{Pos_r}{r} \tag{1}$$

where Rel is the total number of documents relevant to the query, and Pos_r is the position of the r^{th} relevant document in the list of all resultant documents. MAP is defined as the mean AP over all queries. It can stably reflect the overall performance of a search system [12].

4.2 Results and Discussion

We traced the generation of the epistemology repository. Our major concern is whether an exploratory search can benefit from the system’s epistemology repository. Therefore, we introduce the Epistemology-repository Acquisition Rate (EAR) metric – the ratio of exploratory searches that successfully retrieve relevant epistemologies.

We computed the EAR and MAP scores of the *Baijia* system at different stages. In our experiments, the MAP is computed as the mean of every exploratory search’s AP, which is the weighted mean of precisions of all queries it consists of.

$$MAP = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j=1}^{Q_i} w_j AP_j\right) \tag{2}$$

where N is the total number of exploratory searches, Q_i is the total number of queries in the i^{th} exploratory search, AP_j is the AP of the j^{th} query in the i^{th} exploratory search and w_j is the weight of it according to its importance to the exploratory search (e.g. a query with more clicked URLs will be assigned a heavier weight than a query without any clicked URL).

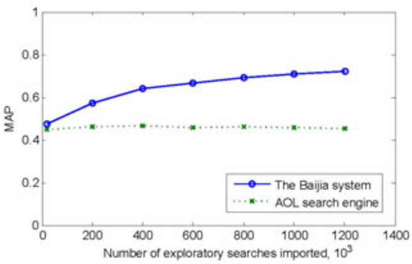


Fig. 5. MAP scores of the *Baijia* system and the AOL search engine

Table 1. Epistemology repository size and EAR at different stages

Number of exploratory searches imported	Epistemology repository size	EAR
20,000	7,612	18.35%
200,000	74,634	29.67%
400,000	151,392	34.20%
600,000	230,273	37.37%
800,000	311,167	39.93%
1,000,000	394,266	41.71%
1,201,497	480,254	42.52%

Fig. 5 shows the MAP scores of the *Baijia* system as compared to those of the AOL search engine (the original data). The results show that increase of the number of exploratory searches imported leads to improvement of MAP scores in the *Baijia* system while the MAP scores of AOL search engine remain steady. This can be explained with Table 1: the EAR increases as more exploratory searches are imported, which implies a user who submits an exploratory search will have a higher probability to get relevant search epistemology from the epistemology repository. Since search epistemology is extracted from clicked URLs and unclicked URLs have been filtered out, the MAP scores of the *Baijia* system are clearly higher than those of the AOL search engine, which can never benefit from the search epistemologies at all. Furthermore, when the search epistemologies in the epistemology repository are re-ranked according to users' feedback rather than random weights, the MAP scores will even be significantly improved.

Unlike previous studies which mainly focus on augmenting search results with relevant data aggregated from the Semantic Web by pre-defined ontologies [29], our work aims at automatically constructing ontologies based on the sequence of queries in every exploratory information seeking process.

5 Conclusions and Future Work

We have proposed a novel epistemology-based social search framework EPISOSE for supporting exploratory information seeking, where search epistemologies – aggregated and well-structured information packages derived from successful search processes contributed by a mass of searchers – are effectively shared, reused, and refined by others with same or relevant search interests or goals. We have also implemented a prototype system *Baijia* based on the framework and conducted a set of experiments to prove that the proposed solution can outperform a conventional search engine in supporting exploratory information seeking. Preliminary usage study indicates that utilizing search engines' immense power and human users' intelligence is an effective and pragmatic solution to exploratory web search.

We have introduced *Baijia* on our intranet to get some initial usage feedback. Most feedback confirms the improvement of the search efficiency in various situations. A main dissatisfaction is that no enough well-refined up-to-date search epistemologies were available to benefit from at the elementary stage. This situation will be improved as the user number increases and the epistemology repository grows. As EPISOSE is an epistemology-based social search framework, *Baijia* relies more on search engines at its initial stage and is self-adaptive to the growing human factors of the system.

Currently, we are conducting a user study to validate the proposed approach. As a social search framework based on users' contribution and reusing, a user satisfaction study is helpful. For example, we could tune the configuration of the prototype system after testing the protocol and algorithm in the real world. In the future, data mining and other artificial intelligence technologies will be adopted to discover the definite objective of an exploratory web search task from the accumulated search epistemologies. We also plan to research and develop components for users to conveniently build social networks and share their intimate search epistemologies for exploratory information seeking.

References

1. Baeza-Yates, R., Tiberi, A.: Extracting semantic relations from query logs. In: Proceedings of KDD 2007, pp. 76–85 (2007)
2. Herlocker, J., Konstan, J., Borchers, A., Riedl, J.: An algorithmic framework for performing collaborative filtering. In: Proceedings of SIGIR 1999, pp. 230–237 (1999)
3. Heymann, P., Koutrika, G., Garcia-Molina, H.: Can social bookmarking improve web search? In: Proceedings of WSDM 2008, pp. 195–205 (2008)
4. Surowiecki, J.: The Wisdom of Crowds: Why the many are smarter than the few and How collective wisdom shapes Business, Economies, Societies and Nations (2004)
5. Jarvelin, K., Kekalainen, J.: IR evaluation methods for retrieving highly relevant documents. In: Proceedings of SIGIR 2000, pp. 41–48 (2000)
6. Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of KDD 2002, pp. 133–142 (2002)
7. Marchionini, G.: Exploratory search: from finding to understanding. Communications of the ACM 49(4), 41–46 (2006)
8. Morris, M.R., Horvitz, E.: SearchTogether: an interface for collaborative web search. In: Proceedings of UIST 2007, pp. 3–12 (2007)
9. Pass, G., Chowdhury, A., Torgeson, C.: A Picture of Search. In: Proceedings of the 1st International Conference on Scalable Information Systems, Infoscale 2006 (2006)
10. Ravi, S.: Optimal search engine marketing strategy. International Journal of Electronic Commerce 10(1), 9–25 (2005)
11. Teevan, J., Adar, E., Jones, R., Potts, M.: Information Re-retrieval: Repeat queries in Yahoo's Logs. In: Proceedings of SIGIR 2007, pp. 151–158 (2007)
12. Voorhees, E.M., Harman, D.: Overview of the 7th Text REtrieVal Conference (TREC-7). In: Proceedings of the 7th Text REtrieval Conference (TREC-7), pp. 1–23 (1999)

Artificial Emotion Generation Based on Personality, Mood, and Emotion for Life-Like Facial Expressions of Robots

Jeong Woo Park, Woo Hyun Kim, Won Hyong Lee, and Myung Jin Chung

Department of Electrical Engineering, KAIST, Daejeon, Korea
{pjw,ishsrain,leestation}@rrlab.kaist.ac.kr,
mjchung@ee.kaist.ac.kr

Abstract. We can't overemphasize the importance of robot's emotional expressions as robots step into human's daily lives. So, the believable and socially acceptable emotional expressions of robots are essential. For such human-like emotional expression, we have proposed an emotion generation model considering personality, mood and history of robot's emotion. The personality module is based on the Big Five Model (OCEAN Model, Five Factor Model); the mood module has one dimension such as good or bad, and the emotion module uses the six basic emotions as defined by Ekman. Unlike most of the previous studies, the proposed emotion generation model was integrated with the Linear Dynamic Affect Expression Model (LDAEM), which is an emotional expression model that can make facial expressions similar to those of humans. So, both the emotional state and expression of robots can be changed dynamically.

Keywords: Artificial emotion, Human-like emotional expression, Integration of emotion generation model and expression model.

1 Introduction

As robots step into the human's daily lives, we can't stress enough the importance of natural human-robot interaction. Especially, emotional expression is an indispensable part of that natural interaction.

Many previous studies about generation and expression of artificial emotion have been conducted so far. In Kismet's emotional architecture, emotion is a medium to construct a kind of social relationship through emotional expression [1]. Kshirsagar et al. used the 'Multilayer Personality Model' to implement a virtual agent with artificial emotion [2]. A. Egges proposed the PME model to generate artificial emotion of conversational agents [3]. J. C. Park proposed a method to generate emotions using a two-layered architecture that consists of a reactive layer and a deliberative layer [4]. Sajal Chandra Banik suggested a Markovian emotion model using a conditional probability at each emotional state [5]. H. Miwa mentioned emotion dynamics using second order differential equations [6].

However, there have been few studies about the dynamics of facial expressions. In some of the previous studies, we can observe that facial expressions are changed with

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 223–233, 2010.

dynamics, but it is simply caused by motor dynamics. Furthermore, there were also very few studies about unified emotion generation and the expression system. Most of the previous studies have handled only the emotion generation model. So, in this paper, we will propose an integrated architecture of emotion generation and an expression system that has emotional dynamics.

2 Overall Architecture of Emotion Generation and Expression Model

Our model consists of two parts. The first is an emotion generation model and the second is an emotion expression model. The former is based on the PME (Personality Mood Emotion) model [3] and the latter is based on the LDAEM (Linear Dynamic Affect Expression Model) [10]. New emotion ${}^{d}\mathbf{e}$ is generated using user's input and personality as an external input. In addition to the external inputs, previous emotion and current mood are considered internally. The generated emotion is transferred to the emotion expression module, LDAEM, which can obtain a trajectory of all action units from current expression to the next expression.

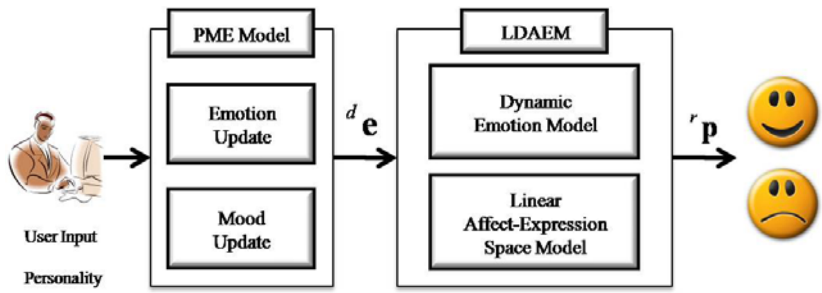


Fig. 1. Overall architecture of emotion generation and expression model

3 Emotion Generation Model

3.1 PME Model

Our proposed emotion generation model is based on the PME (Personality Mood Emotion) model [3]. The PME model is a quite excellent basic framework for generating artificial emotions. This basic framework consists of a mood updating step and an emotion updating step, as shown in (1) and (2). In the mood updating step, the previous mood, $\mathbf{m}_t$, an amount of mood change by input $\Psi_m(\cdot)$, and the internal updating by itself $\Omega_m(\cdot)$ are considered to generate the new mood $\mathbf{m}_{t+1}$. In the emotion updating step, the previous emotion e_t, an amount of emotion change by input $\Psi_e(\cdot)$, and the internal updating by constant term $\Omega_e(\cdot)$ are considered to generate the new emotion $\mathbf{e}_{t+1}$.

$$\mathbf{m}_{t+1} = \mathbf{m}_t + \Psi_m(\cdot) + \Omega_m(\cdot) \tag{1}$$

$$\mathbf{e}_{t+1} = \mathbf{e}_t + \Psi_e(\cdot) + \Omega_e(\cdot) \tag{2}$$

So, we can make various emotion generation models by using this framework with other kinds of mood and emotion updating functions.

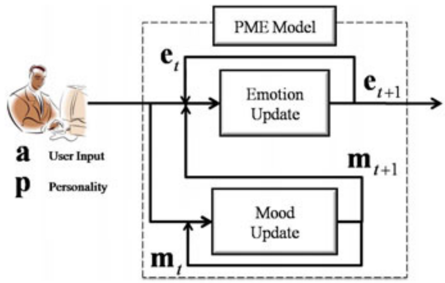


Fig. 2. Block diagram of PME Model

3.2 Proposed Emotion Generation Model

In this paper, we used a five dimensional personality based on the Big Five Model [7], a one dimensional mood (good-bad), and a six dimensional emotion based on Ekman's six basic emotions [8].

According to A. Egges and S. Kshirsagar, the characteristics of personality, mood, and emotion are as listed in Fig. 3.

Fig. 3. Characteristics of personality, mood, and emotion [3]

However, mood state in their results is as dynamic as emotion state, because they considered only current input stimuli in the function $\Psi_m(\cdot)$. Namely, the mood state is varied immediately by input stimuli.

Thus, we used some functions different from those used in the previous PME model to update mood and emotion. First, we considered not only current input stimuli but also the history of input stimuli and emotions to update mood, as shown in (3).

$$\Psi_m(\cdot) = R \cdot Q \cdot \mathbf{a} \;\rightarrow\; \Psi_m(\cdot) = R \cdot Q \cdot (\boldsymbol{\gamma}_t + \boldsymbol{\omega}_t) \tag{3}$$

Here, Q is an emotion-mood influence matrix that defines the relation between the emotions and each mood dimension, $\mathbf{a}$ is a vector of user inputs, $\boldsymbol{\gamma}_t$ is a history of user inputs, and $\boldsymbol{\omega}_t$ is a history of emotions. Both $\boldsymbol{\gamma}_t$ and $\boldsymbol{\omega}_t$ can be obtained as follows:

$$\begin{aligned}
&\boldsymbol{\gamma}_t = \frac{1}{s}\sum_{i=0}^{s}\left(\mathbf{a}_i \frac{\exp(i)-1}{\exp(s)-1}\right)_i, \quad \boldsymbol{\omega}_t = \frac{1}{p}\sum_{i=0}^{p}\left(\mathbf{e}_i \frac{\exp(i)-1}{\exp(p)-1}\right) \\
&\mathbf{a}_i = \begin{bmatrix} a_1 & \cdots & a_l \end{bmatrix}^T, \ l: number\ of\ input\ stimuli \\
&\mathbf{e}_i = \begin{bmatrix} e_1 & \cdots & e_m \end{bmatrix}^T, \quad m: dimension\ of\ emotion, \quad m = l \\
&s: number\ of\ used\ input\ stimuli\ in\, a\ history \\
&p: number\ of\ used\ emotions\ in\ a\ history
\end{aligned} \tag{4}$$

We can make a diagonal matrix R using vector $\mathbf{v}$. This vector $\mathbf{v}$ indicates the stability of each mood. In this paper, vector $\mathbf{v}$ is a scalar because we used only one dimensional mood. So, if this value is too big, the mood state will be unstable.

$$\begin{aligned}
&R = diag(\kappa_1, \cdots, \kappa_k) \\
&\mathbf{v} = R_0 \cdot \mathbf{p} = \begin{bmatrix} \kappa_1 & \cdots & \kappa_k \end{bmatrix}^T, \ k: dimension\ of\ mood \\
&\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}^T, \ n: dimension\ of\ personality
\end{aligned} \tag{5}$$

Here, R_0 is a personality-mood influence matrix that indicates how each personality factor influences each mood dimension.

In the internal updating part of mood, we simply used the same self-decay process that was used in the previous PME model.

$$\Omega_m(\cdot) = -C_m \cdot \mathbf{m}_t \quad , \quad C_m: constant \tag{6}$$

As can be seen in the above equation, we used a history of inputs and emotions to update mood. This was to implement a general property of mood, which is the fact that mood is less dynamic than emotion.

In the part of the emotion updated by input stimuli, the same function in the previous PME model was used.

$$\Psi_e(\cdot) = P \cdot \mathbf{a} + (V \cdot \mathbf{m}_{t+1})^T \cdot \mathbf{a} \tag{7}$$

Here, V is a mood-emotion influence matrix that defines the relation between the mood and each emotion.

We can construct a diagonal matrix P using vector $\mathbf{u}$. This vector $\mathbf{u}$ means the importance of each emotion depending on the personality.

$$\begin{aligned}
&P = diag(\kappa_1, \cdots, \kappa_k) \\
&\mathbf{u} = P_0 \cdot \mathbf{p} = \begin{bmatrix} \varepsilon_1 & \cdots & \varepsilon_m \end{bmatrix}^T, \ m: dimension\ of\ emotion \\
&\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}^T, \ n: dimension\ of\ personality
\end{aligned} \tag{8}$$

Here, P_0 is a personality-emotion influence matrix that indicates how each personality factor influences each emotion.

Even though the same $\Psi_e(\cdot)$ in the previous PME model was used, we employed a different internal updating function $\Omega_e(\cdot)$. We used not a constant term but some functions for the internal updating to differentiate the amount of emotion that is internally updated according to emotion states. Vector **d** makes it possible.

$$\Omega_e(\cdot) = \begin{bmatrix} -C_e \\ \vdots \\ -C_e \end{bmatrix}_{m\times 1} \rightarrow \Omega_e(\cdot) = -C_e \cdot \mathbf{d} \cdot t^2$$
$$\mathbf{d} = \begin{bmatrix} d_1 & \cdots & d_m \end{bmatrix}^T,\ m: \mathit{dimension\ of\ emotion} \tag{9}$$

Here, C_e is a constant value, **d** is an emotion-decay vector, and t is sampling time. So if (9) is used to internally update emotions, we are able to reflect different characteristics of each emotion's decay.

4 Emotion Expression Model

In chapter 3, we proposed an emotion generation model based on the PME model. Now, we will introduce an emotion expression model, LDAEM, and an integration of the expression model with the proposed emotion generation model.

4.1 LDAEM

Using LDAEM, we can easily get dynamic facial expressions according to the change of generated emotions. This model consists of the Linear Affect-Expression Space Model (LAESM) and the Dynamic Emotion Model (DEM), as in (10) and (11).

$${}^{r}\mathbf{p} = T \cdot \mathbf{e} \tag{10}$$

$$M\ddot{\mathbf{e}} + C\dot{\mathbf{e}} + K\mathbf{e} = \mathbf{s} \tag{11}$$

In LAESM [9], affect space and expression space of a developed robot are defined first, and a relationship between the two spaces is determined. So, if an artificial emotion **e** is determined in the affect space, corresponding emotional expression ${}^{r}\mathbf{p}$ can be immediately manifested by (10). Here, T is a transition matrix between affect space and expression space of the developed robot.

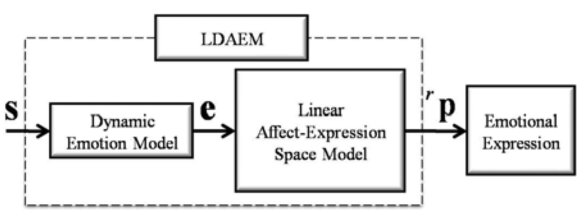


Fig. 4. Block diagram of LDAEM

In DEM [10], trajectories of emotion **e** for input **s** are generated automatically considering characteristics of each emotion, as in (11). Here, *M*, *C*, and *K* are the coefficient matrices that determine the characteristics of bases in the affect space.

4.2 Integration between Emotion Generation Model and Emotion Expression Model

To integrate the emotion expression model with the emotion generation model, all we have to do is use the generated emotions as inputs of DEM, as in (12).

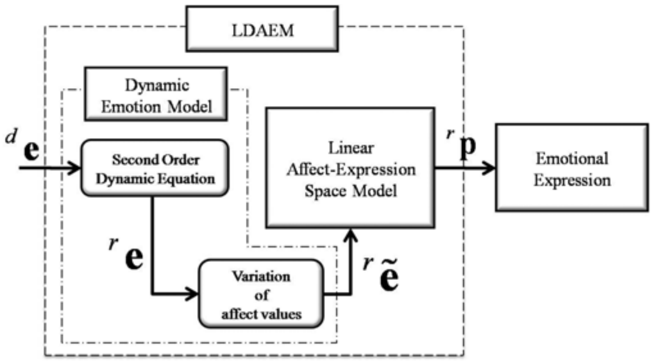


Fig. 5. Modified block diagram of LDAEM

Before applying the second order dynamic equation, we have to rectify the desired emotion $^{d}\mathbf{e}$ that is generated from the proposed emotion generation model. First, we multiplied the desired emotion $^{d}\mathbf{e}$ by matrix *E,* which represents a position of emotions in the affect space because the dimension of the desired emotion $^{d}\mathbf{e}$ and that of the current emotion $^{r}\mathbf{e}$ are different. Next, matrix *K* has to be multiplied by the above result to minimize the steady-state error. Then, a trajectory between current emotion $^{r}\mathbf{e}$ and desired emotion $^{d}\mathbf{e}$ is determined through the second order differential equation.

$$\begin{aligned} &\mathbf{s}_{D_e\times 1} = K_{D_e\times D_e}\cdot E_{D_e\times m}\cdot\ {}^{d}\mathbf{e}_{m\times 1} \\ &M\cdot\frac{d^2}{dt^2}\ {}^{r}\mathbf{e} + C\cdot\frac{d}{dt}\ {}^{r}\mathbf{e} + K\cdot\ {}^{r}\mathbf{e} = \mathbf{s} \\ &D_e : dimension\ of\ \ affect\ space \\ &m : number\ of\ \ generated\ emotions \end{aligned} \tag{12}$$

To make more life-like facial expressions, we added a random value to the current emotion $^{r}\mathbf{e}$ in the affect space within an emotional boundary that was determined based on the recognition rate of facial expressions [11].

$$
\begin{aligned}
&{}^{r}\tilde{\mathbf{e}} = \begin{bmatrix} \tilde{e}_1 \\ \vdots \\ \tilde{e}_{D_e} \end{bmatrix} = \begin{cases} {}^{r}\mathbf{e} + rand(\cdot)\ , & \left\| {}^{r}\mathbf{e} - {}^{t}\mathbf{e} \right\| \le {}^{t}b \\ {}^{r}\mathbf{e} \qquad\qquad , & \left\| {}^{r}\mathbf{e} - {}^{t}\mathbf{e} \right\| > {}^{t}b \end{cases} \\
&{}^{t}\mathbf{e} = E \cdot {}^{d}\mathbf{e}\ , \quad -1 < rand(\cdot) < 1, \\
&{}^{r}\mathbf{e} : current\ emotion\ vector,\ {}^{t}\mathbf{e} : t\arg et\ emotion\ vector \\
&{}^{r}\tilde{\mathbf{e}} : newly\ generated\ emotion\ vector \\
&{}^{t}b\ : radius\ of\ emotional\ boundary \\
&\quad for\ target\ emotion\ in\ the\ affect\ space
\end{aligned}
\tag{13}
$$

5 Experimental Results

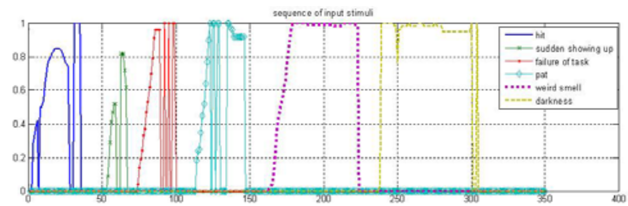
In this chapter, the results of generated emotions and their corresponding facial expressions under various conditions will be shown. We used six kinds of external stimuli such as hit, sudden showing up, failure of tasks, pat, weird smell, and darkness as inputs of the emotion generation model; generated emotions are determined based on Ekman's six basic emotions: anger, surprise, sadness, happiness, disgust and fear. One dimensional mood (positive-negative) and five dimensional personality based on the Big Five Model were used.

Determination of each matrix used in the emotion generation model is very important. We might get absolutely different results according to the values of the four matrices. P_0 is suggested based on Julie's results [12] except for the emotion surprise. Surprise is generally considered as a reactive response, so it can be generated identically regardless of personality. Thus, we decided that the entries of P_0 related with the surprise emotion would be one. Q was determined intuitively and the others were calculated as in (14).

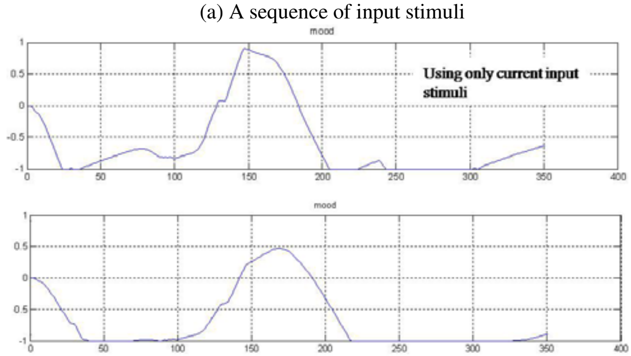
$$
\begin{aligned}
&P_{0_{m\times n}} = \begin{bmatrix} 0.04 & -0.01 & -0.01 & 0.01 & 0.12 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ -0.01 & 0.00 & -0.16 & 0.06 & 0.06 \\ 0.18 & 0.23 & 0.29 & 0.23 & -0.16 \\ -0.17 & -0.12 & -0.24 & -0.04 & 0.24 \\ -0.24 & -0.32 & -0.22 & 0.06 & 0.38 \end{bmatrix} \\
&Q_{k\times m} = \begin{bmatrix} -0.3 & 0.0 & -0.1 & 0.3 & -0.2 & -0.15 \end{bmatrix} \\
&R_{0_{k\times n}} = Q_{k\times m} \cdot P_{o_{m\times n}} = \begin{bmatrix} 0.113 & 0.144 & 0.187 & 0.059 & -0.195 \end{bmatrix} \\
&T_{m\times k} = Q_{k\times m}{}^{T}
\end{aligned}
\tag{14}
$$

The first experiment looks at how differently mood is generated according to whether history of input and emotion are used or not. Here, we considered 50 previous input stimuli and emotion states. As shown in Fig. 6 (b), mood is changed more dynamically when only the current input stimuli are used than when history of input stimuli and emotions are used. For instance, a slop of mood's change in bottom graph of fig. 6 (b) is less gentle than that of the upper graph of fig. 6 (b), and mood in the bottom graph of fig. 6 (b) is not changed immediately according to user inputs.

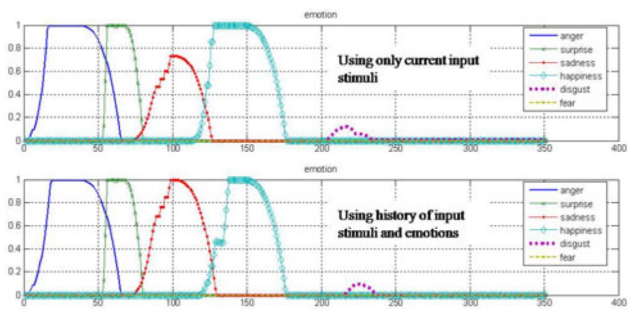
In addition to the change of mood, patterns of generated emotions such as intensity and lasting time of emotion are changed a little because current mood participates in updating emotions.



(a) A sequence of input stimuli



(b) A change of mood



(c) A change of emotions

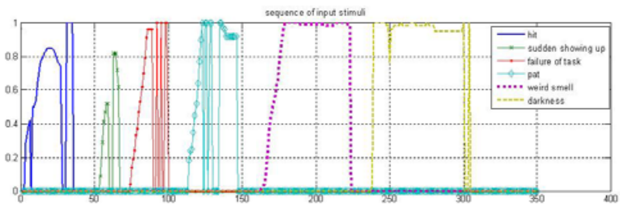
Fig. 6. A change of mood and emotions according to the using history of input stimuli and emotions

Second, we performed the experiment for two kinds of personality, such as strong openness and strong agreeableness to observe differences of the generated emotions.

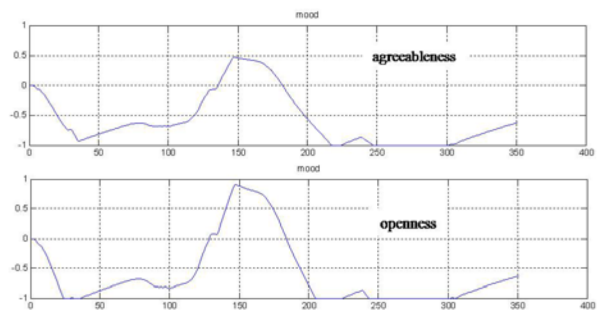
Elements of **p** in (15) are openness, conscientiousness, extraversion, agreeableness, and neurotic orderliness.

$$\mathbf{p} = \begin{bmatrix} 1.0 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix} \;:\textit{strong openess}$$
$$\mathbf{p} = \begin{bmatrix} 0.1 & 0.1 & 0.1 & 1.0 & 0.1 \end{bmatrix} \;:\textit{strong agreeableness} \tag{15}$$

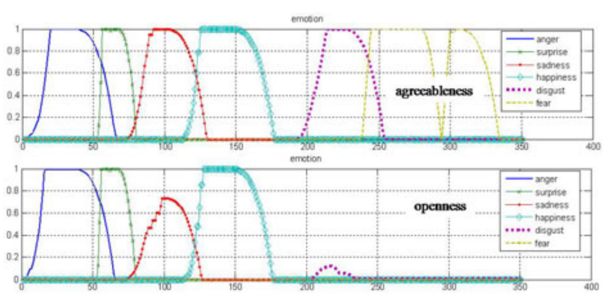
The sequence of input stimuli is the same as that shown in Fig. 6 (a). As shown in Fig. 7 (c), we found that generated emotions are completely different from each other according to each personality. In the case of agreeableness, all kinds of emotions were generated easily according to input stimuli. In the case of openness, however, sadness, disgust, and fear emotions were not generated as easily as they would be with an agreeable personality. This is due to the personality-emotion influence matrix $P_{0.}$ In



(a) A sequence of input stimuli



(b) A change of mood



(c) A change of emotions

Fig. 7. A Change of mood and emotions according to personalities

(14), the first column of P_0 is for the relationship between openness of personality and the six basic emotions, which are anger, surprise, sadness, happiness, disgust, and neurotic orderliness. The fourth column is for agreeableness of personality. Comparing the two columns, values of sadness, disgust, and fear for agreeableness are greater than those for openness. Therefore, the above three emotions for agreeableness were generated more easily than they were for the personality of openness. Furthermore, intensity of mood is quite dissimilar according to personalities because history of emotions participates in updating mood.

The final experiment is to check the change of the internal update of emotion according to each emotion. We can find that degree of internal update for surprise and happiness is relatively bigger than that value would be for other emotions.

Fig. 9 is a sequential change of facial expression using the results of Fig.8 as an input of proposed emotion expression model.

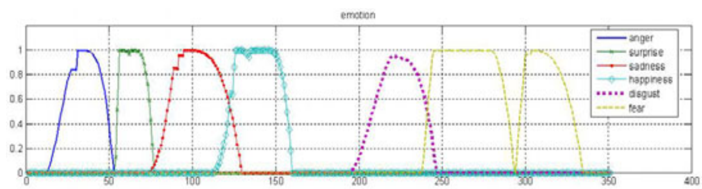


Fig. 8. Different internal update according to the kind of emotion

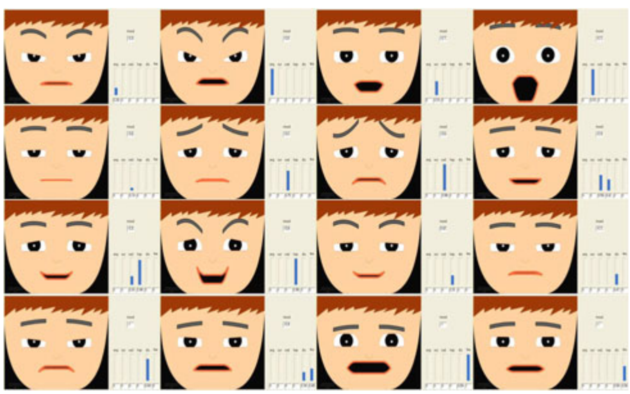


Fig. 9. A sequence of facial expressions using generated emotion

6 Conclusion

In this paper, we proposed an emotion generation model based on the PME model and a method to integrate the emotion generation model with the emotion expression model LDAEM. To implement the general characteristics of mood and emotion, we used a history of input and emotions. Through the experiments, we confirmed the differences between the previous methods and the proposed one. Furthermore, we suggested a guideline to determine the four matrices that were used in the PME model. Entries of those matrices are quite significant factors for generating emotions. We also enhanced the internal updating function of emotion. Finally, we integrated the proposed emotion generation model with LDAEM using matrix *E,* which represents a position of each emotion in the affect space.

Here we only updated emotions and mood, but the personality of robots also can be changed when a long term interaction between humans and robots is occurring. In this paper, we generated only six basic emotions, so this report might be considered slightly simple. So, we need to do more research about personality updating and need to expand the kinds of generated emotions.

Acknowledgement

This work was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the IITA(Institute for Information Technology Advancement).

References

1. Breazeal, C.: Designing Social Robots. The MIT Press, Cambridge (2002)
2. Kshirsagar, S., Magnenat-Thalmann, N.: A Multilayer Personality Model. In: Proc. of 2nd International Symposium on Smart Graphics, pp. 107–115 (2002)
3. Egges, A., Kshirsagar, S., Magnenat-Thalmann, N.: Generic personality and emotion simulation for conversational agents. Computer Animation and Virtual Worlds 15(1), 1–13 (2004)
4. Park, J.C., Kim, H.R., Kim, Y.M., Kwon, D.S.: Robot's Individual Emotion Generation Model and Action Coloring According to the Robot's Personality. In: The 18th IEEE International Symposium on Robot and Human Interactive Communication Toyama, Japan, September 27-October 2 (2009)
5. Banik, S.C., Watanabe, K., Izumi, K.: Intelligent Behavior Generation of Benevolent Agents with a Stochastic Model of Emotion. In: Ishikawa, M., Doya, K., Miyamoto, H., Yamakawa, T. (eds.) ICONIP 2007, Part II. LNCS, vol. 4985, pp. 147–156. Springer, Heidelberg (2008)
6. Miwa, H., Itoh, K., Takanobu, H., Takanishi, A.: Development of Mental Model for Humanoid Robots. In: ROMANSY, 2004 Montreal, Canada, June 14-18 (2004)
7. Digman, J.M.: Personality structure: Emergence of the five-factor model. Annual Review of Psychology 41, 417–440 (1990)
8. Ekman, P., Friesen, W.V.: Unmasking the Face. Malor Books, Cambridge (2003)
9. Lee, H.S., Park, J.W., Chung, M.J.: A Linear Affect–Expression Space Model and Control Points for Mascot-Type Facial Robots. IEEE Transactions on Robotics 23(5) (October 2007)
10. Lee, H.S., Park, J.W., Jo, S.H., Chung, M.J.: A Linear Dynamic Affect-Expression Model: Facial Expressions According to Perceived Emotions in Mascot-Type Facial Robots. In: Proceedings 16th IEEE International Conference on Robot & Human Interactive Communication, August 26 (2007)
11. Park, J.W., Kim, W.H., Lee, W.H., Kim, W.H., Chung, M.J.: Lifelike Facial Expression of Mascot-Type Robot based on Emotional Boundaries. In: 2009 IEEE International Conference on Robotics and Biomimetics, Guilin, Guangxi, China, December 19-23 (2009)
12. Penley, J.A., Tomaka, J.: Associations among the Big Five, emotional responses, and coping with acute stress. Personality and Individual Differences 32, 1215–1228 (2002)

Human Error Categorization: An Extension to Classical Proposals Applied to Electrical Systems Operations

Daniel Scherer[1,2], Maria de Fátima Q. Vieira[1,3], José Alves do N. Neto[1]

[1] LIHM, Federal University of Campina Grande, Electric Engineering Department
Av. Aprígio Veloso, 882 - Bodocongó CEP 58.109-970 Campina Grande, PB – Brazil
[2] Center for Science and Technology, State University of Paraiba
Bodocongó - Campina Grande, PB – Brazil
[3] Centre for Excellence in Signal & Image Processing, Dept of Electronic & Electrical Engineering, University of Strathclyde - Scotland, UK
{daniel.scherer,jose.alves}@ee.ufcg.edu.br,
fatima@dee.ufcg.edu.br

Abstract. Accident and incident analysis is essential to the study of human error and the development of error prevention measures. Human Error research deals essentially with the classification of error and the identification of the causal relation between the error detected and the level of human performance at which it occurred. As a result the literature proposes many error categorization methods and taxonomies. These are not, in themselves, sufficient, however, to analyze (and understand) the circumstances surrounding the error occurrence. For a more complete understanding of human error, it is necessary to associate each error with the sequence of steps taken by the human operator during the task that led to it. This paper proposes an extension of the existing error categorization found in the literature and applies it to the analysis of human error reports originating in the electricity industry.

Keywords: Human Error Categorization, Error Analysis, Error taxonomy.

1 Introduction

The study of human error has an important application in supporting the reporting and analysis of accidents and incidents in industrial automated systems. Based on the existing literature, studies may grouped as follows: error identification ([5], [2], [3], [4], [10] and [6]), the human performance level at which the error occurs ([5] and [7]), and the context in which the error occurs ([5] and [10]). Analyzing the error and its background, however, may not be sufficient for a complete understanding of the situation which preceded its occurrence. A crucial requirement for the effective analysis of an error report is the need to identify the precise decision process, made by the operator[1] of an automated system, which might have caused the accident or incident.

[1] Along this paper, the term operator will imply human operator in industrial automated systems.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 234–245, 2010.

Hollnagel [10], Reason ([6] and [8]) and Norman [4] address the human error issue from a different viewpoint; that of describing the mechanisms of human malfunction (such as unintentional action, data-driven activation, spoonerisms, problems with causality, etc.). In contrast Rasmussen et al. [5], and particularly the extensions of this work proposed by Rouse et al. (apud Cellier [1], [2], [3]), identify the relationship between the human error and the sequence of decisions that were made prior to it. This is similar to the extension proposed in this paper. Our experience, obtained from analyzing a *corpus* of human error reports using this method of error categorization, suggests that, in addition to considering the error occurrence, it is necessary to consider correct actions taken by the operator. This additional consideration is central to understanding the context and decisions that lead up to an error. This motivates our proposed extension of the classic error classification. We apply the proposed extension to electric power systems operation in the particular context of automated substations.

The rest of the paper is organized as follows. Section 2 presents a review of the model proposed by Rasmussen et al. [5] and its extension, proposed by Rouse et al. ([1], [2], [3]), and draws comparisons between them and the taxonomy adopted by the company which serves as a case study. Section 3 describes our proposed error categorization and applies it to the analysis of a corpus of error reports that comprises the case study. Finally, Section 4 discusses the results and proposes new directions for this work.

2 Human Error Classification

Rasmussen proposed a model to describe the human decision sequence. This can be represented as a graph with the typical sequence of steps a human takes when carrying out a task.

2.1 Rasmussen Error Classification

In [5], Rasmussen et al. represents the model as an adequate taxonomy for reporting industrial incidents and events, involving *human malfunctions*. One of the proposed error categories is the *internal human malfunction*, described in **Fig.1**. The sequence of steps depicted provides the basis for human error categorization as corroborated by other authors (e.g. [2], [3] and [1]). Rouse & Rouse [1] take this sequence and expand it, associating to each step of the decision sequence, errors that might occur during task execution. Rasmussen et al. use this model to explain information processing activities (represented by rectangles in **Fig. 1**) and the resulting knowledge (represented by ellipses in **Fig. 1**) associated with the possible sequence of decision leading to the action performed. Another error category is the *external mode of malfunction*, which describes the immediate and directly observable human malfunction (observable errors) during the execution of a given task (**Table 1**).

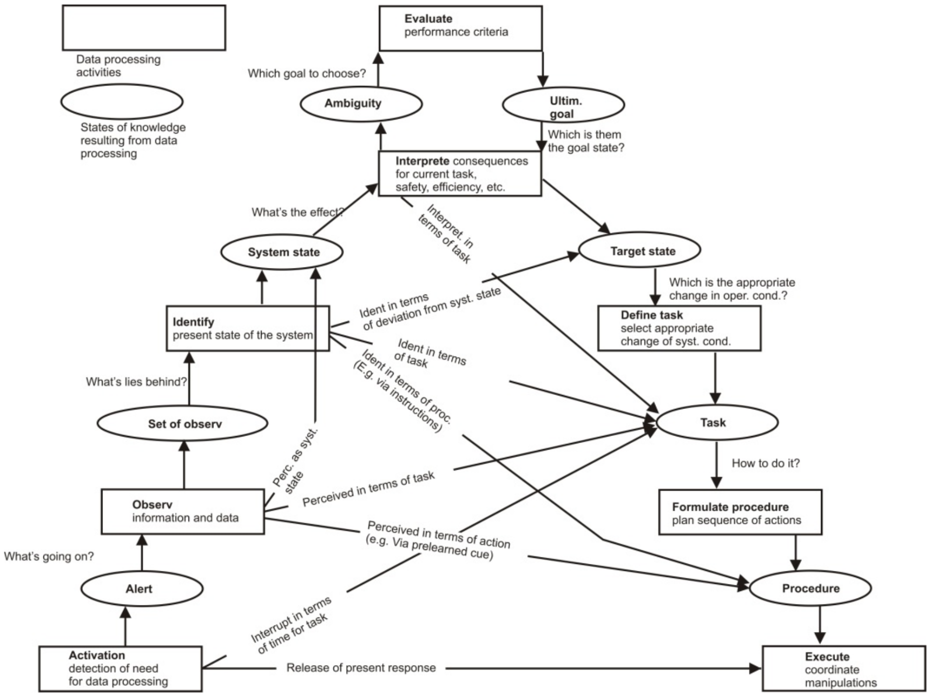


Fig. 1. Rasmussen's et al. model of human decision sequence [5]

Table 1. Set of observable errors proposed by Rasmussen et al.

specified task was not performed:
- omission of task;
- omission of act;
- inappropriate, inaccurate performance;
- inappropriate/wrong timing;
- actions in wrong sequence.
the effect is due to specific, erroneous act:
- wrong act executed on correct component, equipment;
- wrong component, equipment;
- wrong timing.
the effect is due to extraneous act;
the effect is due to coincidence, several events or faults;
not stated, not applicable.

2.2 Eekhout, Johnson, Rouse and Rouse Error Classification

Based on the work of Rasmussen et al., Eekhout & Rouse [2] proposed a simplified model of the human decision sequence [5] resulting in a sequential list of general and specific categories; each error type being a member of one specific category. This sequential list of general and specific categories was modified by Johnson & Rouse [3]. Later Rouse & Rouse (apud Cellier [1]) proposed a larger set of specific categories for each general category, **Table 3**.

2.3 Confronting Error Classifications

The two categories proposed by Rasmussen et al. (*internal human malfunction* and *external mode of malfunction*) greatly influenced the work of Eekhout & Rouse, Johnson & Rouse and Rouse & Rouse, resulting in only slight differences between the general error categories proposed by these authors (evident from **Table 2**). Our proposed categorization has three additional general categories (**Table 2**). These are recovery, consequences and causes, and are included to expose, respectively, (i) the time that an operator takes to recover from an error, (ii) the consequences resulting from the error, and (iii) the causes (state level of the operator and contextual situations) of the error. There is a significant difference between the categorizations concerning observable errors as shown in **Table 3**.

Table 2. General Error Categories compared

Rasmussen	Eckhout & Rouse	Johnson & Rouse	Rouse & Rouse	Proposed categorization
ACTIVATE	Observation of system state	Observation of system state	Observation of system state	Observation of system state
OBSERVE				
IDENTIFY present state of the system	Identification of fault	-		
INTERPRET consequences for current task, safety, efficiency, etc.	-	Choice of hypotheses	Choice of hypotheses	Choice of hy-potheses
EVALUATE performance criteria	-	-	Hypotheses evaluation	Hypotheses evaluation
DEFINE TASK	Choice of goal	-	Choice of goal	Choice of goal
FORMULATE PROCEDURE	Choice of procedure	Choice of procedure	Choice of procedure	Choice of procedure
EXECUTE	Execution of procedure	Execution of procedure	Execution	Execution
-	-	-	-	Recovery
-	-	-	-	Consequences
-	-	-	-	Causes

Table 3. Comparing categories of observable errors

General Category	Rasmussen	Eckhout & Rouse	Johnson & Rouse	Rouse & Rouse	Proposed categorization
Observation of system state		- incomplete - inappropriate - lack	- incomplete - misinterpreted - repeated	- excessive - misinterpretation - incorrect - incomplete - inappropriate - absent	- excessive - misinterpretation - incorrect - incomplete - inappropriate - absent - unnecessary - correct
Choice of hypotheses		- incomplete - inappropriate - lack	- inconsistent with symptoms - consistent but unlikely - consistent but costly - functionally irrelevant	- inconsistent with symptoms - consistent but unlikely - consistent but costly - functionally irrelevant	- inconsistent in relation to observation - consistent, but less probable - consistent, but extremely costly -non-pertinent from the viewpoint of functionality - absent - insufficiently consistent - unnecessary - correct
Hypotheses evaluation				- incomplete - acceptance of an incorrect hypothesis - rejection of a correct hypothesis - absent	- incomplete - acceptance of an incorrect hypothesis - rejection of a correct hypothesis - absent - unnecessary - correct
Choice of goal		- incomplete - inappropriate - lack		- incomplete - incorrect - superfluous - absent	- incomplete - incorrect - superfluous - absent - unnecessary - correct
Choice of procedure		- incomplete - inappropriate - lack	- incomplete - inappropriate - lack	- incomplete - incorrect - superfluous - absent	- incomplete - incorrect - superfluous - absent - unnecessary - correct
Execution	specified task not performed: - omission of task; - omission of act; - inappropriate, inaccurate performance; - inappropriate/ wrong timing; - actions in wrong sequence. the effect is due to specific, erroneous act: - wrong act executed on correct component, equipment;	- incomplete - inappropriate timing - inadvertent action	- omission of steps - other - inadvertent action	- omitted operation - repeated operation - addition of an operation - operation out of sequence - intervention at some non-appropriate time - position of inappropriate operation - incomplete execution - non-related, inappropriate	- omission - replication - inclusion - sequence - intervention at some inappropriate time - incorrect operator position - incomplete action - unrelated or inappropriate action - right action on wrong object - unintended action

Table 3. (*Continued*)

	- wrong component, equipment; - wrong time. the effect is due to extraneous act; the effect is due to coincidence, several events or faults; not stated, not applicable.			action	
Recovery					- too late - late - immediate
Consequences					- non-interrupted load - interrupted load - equipment overload - loss and damaged of equipment - personal damages
Causes					- lack of concentration occasioned by haste - lack of concentration occasioned by excessive self-confidence - lack of concentration - hurry - stress - confusion - pressure - anxiety - improvisation - inexperience - excessive self-confidence - personal problems - lack of technical capacity - tiredness - excessive concentration

3 Proposed Categorization of the Error

The extension of Rouse & Rouse's categorization was motivated by an analysis of errors in the operation of an electrical power system extracted from 10 years of reports provided by the Brazilian electricity company, CHESF.

Due to the general focus of the Rouse & Rouse categorization the number of errors classed as *Execution* errors resulted in mapping the terms defined by Rouse & Rouse onto those found in the reports. Although the original structure [9] was maintained, the three added general categories (*Recovery*, *Causes* and *Consequences*) and their specific subcategories were proposed to accommodate the corpus analyzed. When scrutinizing the error reports, however, we found situations that were not addressed by the specific categories proposed by Rouse & Rouse, e.g. when an operator correctly observes the system state, or the situation described makes it unnecessary to evaluate the hypothesis. We therefore added the specific categories *unnecessary* and *correct* to every general category (except execution). We also observed that situations occurred where the choice of hypothesis was partly consistent, but did not justify a subsequent action. The specific category *insufficiently consistent* was therefore added to the general category *Choice of hypotheses*.

3.1 Report Analyses

A Pareto chart, aiming to identify the incidence distribution of specific categories, was generated using the set of analyzed data for each general category at each step of the Rasmussen model. This information was then used to propose strategies to prevent the causes of failures that accounted for more than 50% of the errors. **Table 4** summarizes the results of the reports analyzed.

Table 4. Categorization of failure reports

General Category	Specific Category	Incidence
Observation of system state	Excessive	1
	Misinterpretation	8
	Incorrect	2
	Incomplete	11
	Inappropriate	0
	Absent	14
	Unnecessary	1
	Correct	3
Choice of hypotheses	Inconsistent in relation to observation	6
	Consistent, but less probable	1
	Consistent, but extremely costly	0
	Non-pertinent from the viewpoint of functionality	8
	Absent	11
	Insufficiently consistent	10
	Unnecessary	1
	Correct	0
Hypotheses evaluation	Incomplete	3
	Acceptance of wrong hypothesis	21
	Rejection of right hypothesis	0
	Absent	11
	Unnecessary	2
	Correct	0
Choice of goal	Incomplete	3
	Incorrect	7
	Superfluous	0
	Absent	0
	Unnecessary	0
	Correct	26

Table 4. (*Continued*)

Choice of procedure	Incomplete	9
	Incorrect	12
	Superfluous	1
	Absent	0
	Unnecessary	0
	Correct	17
Execution	Omission	5
	Replication	0
	Inclusion	0
	Sequence	3
	Intervention at some inappropriate time	3
	Incorrect operation position	0
	Incomplete action	5
	Unrelated or inappropriate action	4
	Right action on wrong object	15
	Unintended action	2
Recovery	Too late	7
	Late	6
	Immediate	15
Consequences	Non-interrupted load	15
	Interrupted load	19
	Equipment overload	0
	Loss and damaged of equipment	0
	Personal damages	0
Causes	Lack of concentration occasioned by haste	3
	Lack of concentration occasioned by excessive self-confidence	5
	Lack of concentration	14
	Hurry	4
	Stress	5
	Confusion	7
	Pressure	2
	Anxiety	3
	Improvisation	5
	Inexperience	2
	Excessive self-confidence	11
	Personal problems	2
	Lack of technical capacity	5
	Tiredness	5
	Excessive concentration	1

From the analysis it was found that 31 out of 35 failures occurred at the stage of *Observation of system state* (**Fig. 2**). In three out of 35 reports the *Observation of system state* (**Fig. 2**) was correct, failure occurring at some other stage of the decision sequence. 10 out of 35 errors occurred during the *choice of goal* (**Fig. 3**). The *choice of procedure* was correct in 17 out of 35 reports (**Fig. 4**).

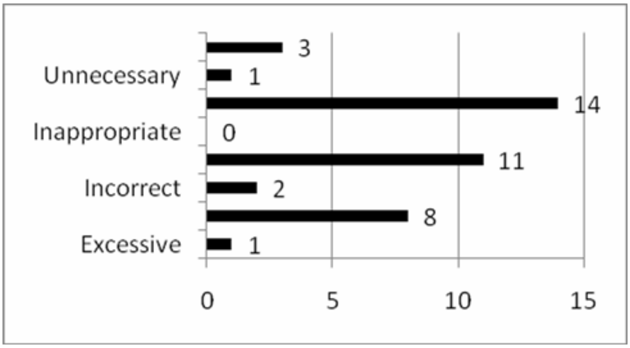


Fig. 2. Observation of system state stage

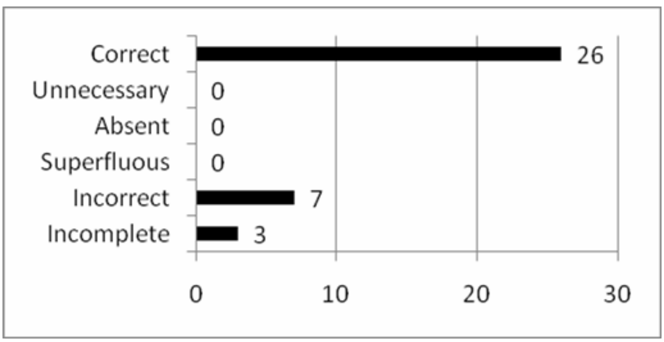


Fig. 3. Choice of goal stage

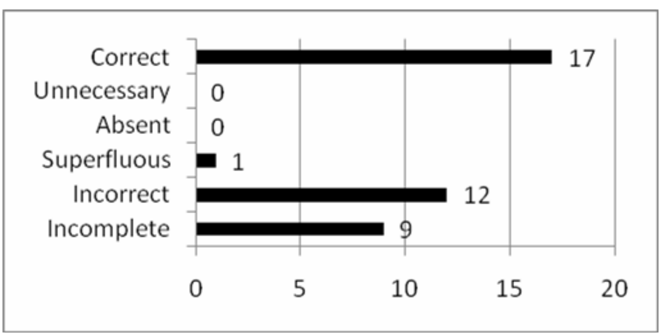


Fig. 4. Choice of procedure stage

4 Final Considerations and Future Directions

This paper proposes an extension of Rouse & Rouse's error categorization that resulted from the analysis of a *corpus* of 35 reports. It seeks to identify not only the final error that occurred (e.g. omission, right action on wrong object), but also the actions taken throughout the sequential decision process.

The extension was based on the taxonomy proposed by Rasmussen et al., on the error classification by Rouse et al. and on the analysis of a corpus of human error reports from the operation of an electrical power system.

The extended classification was reapplied in the analysis of the same set of reports which made it possible to identify situations unobserved using the original (Rouse & Rouse) error categorization. The analysis results were then projected into Pareto diagrams to identify the faults accountable for more than 50% of all errors.

For a complete understanding of human error, it is necessary to associate the error with the sequence of steps taken by the human operator during task performance. We have three reports in which the observation of the system was correct. Scrutinizing the decision-making sequence in these reports **(Fig. 5**), it is clear that there were problems in the *choice of hypothesis* and *evaluation of the hypothesis*. In two of the reports the right goal was chosen with the correct procedure being followed. In contrast, in the third report (R34), the procedure followed was incomplete, possibly leading to incomplete execution.

Finally we have the recovery, the causes and consequences reported. In R20 an *absence of hypothesis* culminated in an unintended action. A hypothesis inconsistent in relation to an observation in R25 led to right action on wrong object. In R34 we have a hypothesis *insufficiently consistent* leading to an incomplete action.

Table 5 illustrates the incidence of hits and misses along the decision making sequence. The relation between the events that culminated in an execution error and its causes are illustrated in the graph depicted in **Fig. 5**. These relations were extracted from the analysis of the error reports, excerpts of which are given in **Fig. 5**.

Table 5. Excerpts from error reports: R20, R25 and R34

R.20: "...decided to go to the equipment in order to inspect it. After opening its comand box directed the attention to reading the label fixed on its door. At this point stepd on false and in the attempt to hold on and avoid falling caused the equipment trip."
R25: "... the operator knows the manouvers in detail and had the manouver instructions at hand... when selecting the switch X got confused and selected switch Y, since both were placed on the same chassis", "self-confidence- simple, standard and routine manouver", "momentary loss of concentration due to self-confidence", "incorrect selection of switch to be manouvered".
R34: "Operator's lack of concentration during manouver execution due to procedure disregard (task considered simple) and manouver execution during working shift exchange". "Second operator missing resukting in no double checking or follow up". "Lack of operator concentration... causing incomplete execution of one item in the manouver, agravated by this item's characteristics which demanded two actions performed in separate places".

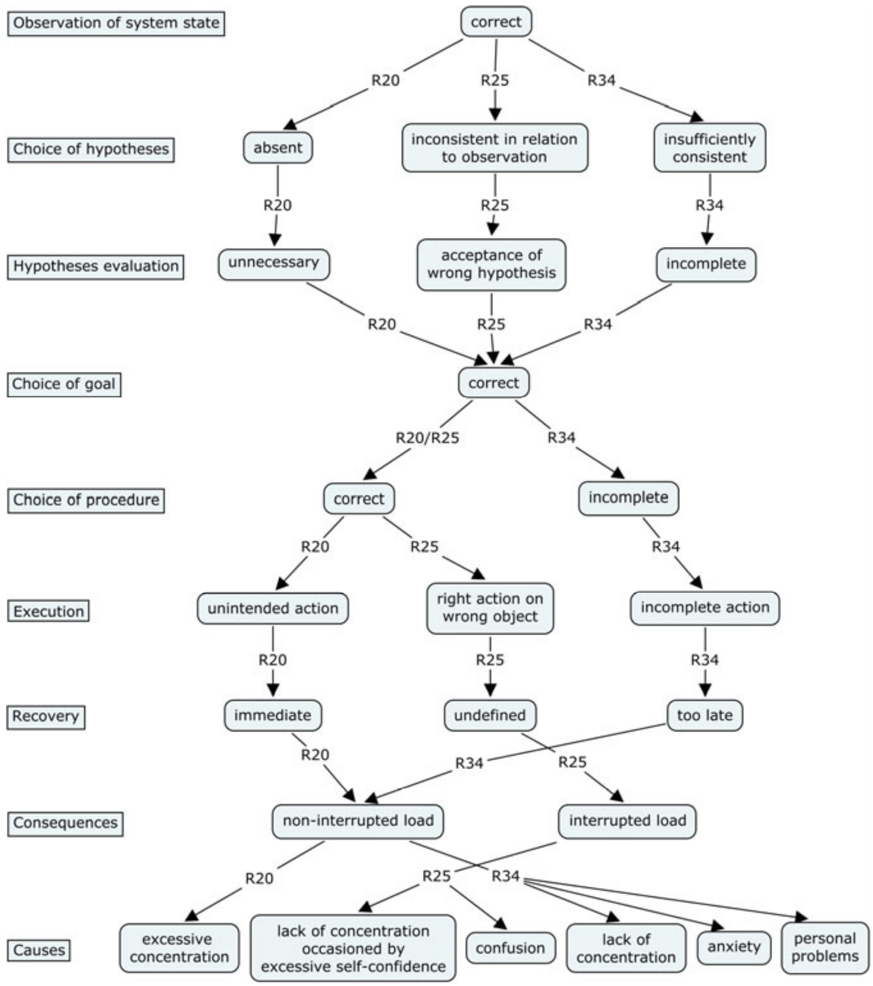


Fig. 5. Human decision sequence of the reports 20, 25 and 34

Although the error categorization extension proposed here focuses on error situations reported during the operation of an electrical power system, the authors expect to address other contexts by adapting the specific categories: recovery, consequences and causes.

The more refined categorization presented here will not prevent the occurrence of errors but will enable to explicit the relations between causes and consequences leading to more effective strategies to prevent the causes that lead into errors.

It may appear to be a long way to relate causes and consequences and from there come to strategies to prevent the errors, but although the gap still remains, this paper tries to narrow it. The authors plan to invest more time into researching the accidents and incidents based upon this taxonomy, uncovering relations between causes and consequences in order to propose more effective error prevention strategies. Given the scope of the problem this could not be dealt within the scope of this paper.

The following future work is proposed:

- To verify the completeness of the extended categorization, analyzing a new set of reports;
- To develop software to facilitate both the analysis of reports and the visualization of the analysis results;
- To associate the proposed categorization to a test usability protocol in order to support the observation and the analysis of the user's behavior during system operation.

Acknowledgments. The authors wish to thank the Brazilian government agencies CNPq and CAPES for the financial support to this project as well as the engineers of the electricity company CHESF for their support and collaboration with this work.

References

1. Cellier, J.M.: L'erreur human dans le travail. In: Leplat, J., De Terssac, G. (eds.) Les facteurs humains de la fiabilite dans les systemes complexes, Marseille (1990)
2. van Eekhout, J.M., Rouse, W.B.: Human Erros in Detection, Diagnosis, and Compensations for Failures in the Engine Control Room of a Supertanker. IEEE Transactions on System, Man, and Cybernetics SMC-11(12) (Dezembro 1981)
3. Johnson, W.B., Rouse, W.B.: Analysis and Classification of Human Errors in Troubleshooting Live Aircraft Power Plants. IEEE Transactions on System, Man, and Cybernetics SMC-12(3) (Maio 1982)
4. Norman, D.A.: Design Rules Based on Analyses of Human Error. In: Denning, P.J. (ed.) Communications of the ACM, vol. 26(4), pp. 254–258 (April 1983)
5. Rasmussen, J., Pedersen, O.M., Mancini, G., Carnino, A., Griffon, M., Gagnolet, P.: Classification system for reporting events involving human malfunctions. In: Relatório RISO-M-2240. RISO National Laboratory, Dinamarca (March 1981)
6. Reason, J.: A framework for classifying errors. In: Rasmussen, J., Duncan, K., Leplat, J. (eds.) New Technology and Human Error, pp. 5–14. John Wiley & Sons Ltd., Chichester (1987)
7. Reason, J.: A preliminary classification of mistakes. In: Rasmussen, J., Duncan, K., Leplat, J. (eds.) New Technology and Human Error, pp. 15–22. John Wiley & Sons Ltd., Chichester (1987)
8. Reason, J.: Human Error. Cambridge University Press, Cambridge (1990)
9. Guerrero, C.V.S.: Modelo conceitual de cenários de acidentes causados pelo erro humano em sistemas industriais críticos com foco na concepção de interfaces ergonômicas. Doctoral Thesis. UFCG (2006)
10. Hollnagel, E., Woods, D.D.: Cognitive systems engineering. New wine in new bottles. International Journal of Man-Machine Studies (1983)

Exploring the Influence of Cultural Diversity in Collaborative Design Teams: Preliminary Findings

Vivian Hsueh-Hua Chen[1] and Henry Been-Lirn Duh[2]

[1] Wee Kim Wee School of Communication and Information,
Nanyang Technological University
chenhh@ntu.edu.sg
[2] Department of Electrical and Computer Engineering, NUS-KEIO CUTE Center/Mixed Reality Lab, National University of Singapore
eledbl@nus.edu.sg

Abstract. Past studies loosely define culturally heterogeneous group composition as any combination of mixed nationality and/or ethnicity. A case study with three project groups was conducted to investigate if culturally diverse groups composed differently in terms of nationality/ ethnicity mix will experience different types of communication problems. Communication problems of discussion dominance, reduced communication frequency and discussion exclusion were coded from taped discussions. Findings indicate that ethnicity and nationality give rise to different communication problems. Future studies should be specific when defining heterogeneous groups in terms of whether they are manipulating differences in nationality, ethnicity or both.

Keywords: Cultural diversity, group composition, collaborative design, group process.

1 Introduction

Past research investigating the effects of cultural diversity on group communication for collaboration compared the communication processes between culturally homogeneous vs. heterogeneous groups [see examples: 6, 17]. Group composition is defined by nationality and ethnicity. Majority of the studies compared European/ American homogenous groups with any set of heterogeneous individuals. It is assumed that all culturally diverse groups are alike [19].

However, there are indications in literature that this may not be the case. Take studies that investigate the effects of group heterogeneity on cooperative behavior in group setting for example. Cox, Lobel, and McLeod [6] conducted an experiment to compare cooperation and competition between homogeneous Anglo-American groups against heterogeneous groups composed of Asian, Hispanic, Black and Anglo-Americans. The Anglo-American participants represented communication behavior in individualistic culture. The Asian, Hispanic and Black participants were all taken to represent communication behavior in collectivistic culture. No differentiation was made between the Asian, Hispanic and Black participants. The study concluded that ethnically-diverse groups acted more cooperatively than homogeneous Anglo-American groups due to the influence of collectivistic group members.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 246–254, 2010.

Past studies have used the term “homogeneous” to refer to groups composed of members from the same nationality and ethnic group, typically Anglo-White. The term “heterogeneous” is used to refer to groups composed of members from different nationality and/or ethnicity. However, not all heterogeneous groups are alike. Heterogeneous group members could differ in terms of ethnic cultural background; they could differ in terms of national cultural background. Hence, this paper studies how groups composed of members from different cultural backgrounds will experience different communication problems in the collaborative design process.

2 Communication Problems in Collaborative Design Teams

Collaborative design is largely a series of decision choices [5]. When team members come from different cultural backgrounds, they are able to contribute a wide range of perspectives to the decision-making process. However, communication problems can hinder the collaborative design process. Three main problems were discussed based on past research.

2.1 Discussion Dominance

In the context of group setting, dominance refers to a state of power, a reflection of team members’ hierarchy in the group [15]. It is a function of power differences between group members in discussion. Watson et al [25] showed that groups composed of culturally different individuals experience more occurrences of discussion dominance than homogeneous groups. Their study compared the interaction process and performance of heterogeneous and homogeneous groups over time. When the workgroups were newly-formed, homogeneous groups reported fewer power struggles, equal participation and higher levels of cohesion heterogeneous groups on the other hand had more frequent occurrences where group members dominated the discussion and hindered member contribution.

The context of group discussion affects the exhibition of dominance by team members. Hence, to understand communication issues in collaborative design groups, it is important to examine the relationship between cultural diversity and actual exhibition of dominating behavior during discussion.

2.2 Reduced Communication Frequency

Studies have consistently concluded that frequent communication between group members to their colleagues, both inside and outside of their project group, is vital to high project performance [10]. The frequency of communication between team mates is dependent on the familiarity and proficiency of each in the shared language of communication [14].

Culturally diverse teams experience reduced frequency of communication if team members are not proficient with using the common language of discussion. Loosemore and Lee’s [13] survey of the construction industry in Singapore and Australia found that language barrier was the biggest difficulty that construction site supervisors faced when communicating with workers from other cultures. As a result, foreign workers were less likely to communicate potential problems to their supervisors.

Foreseeable problems were not brought to light. Foreign workers either ignored the problems or tackled them without consulting with the supervisor. Similarly, a reduction of communication between team members has dire consequences to the success of collaborative design.

2.3 Discussion Exclusion

Individuals in a team define their social identities by the process of social categorization using salient characteristics like ethnicity [23]. This in turn leads to social comparison with others so as to maintain a high level of self-esteem. Such social categorization results in the formation of in-group/ out-groups in teams [22]. People like to interact with those who are more like themselves, i.e., the in-group. Conversely, people exclude those who are less like themselves, i.e., the out-group, in their interactions [3]. Hence, in-group/ out-group differentiation results in the exclusion of ethnically different team members from the communication process [11]. The exclusion of out-group members from the communication loop reduces team cohesion. It also disadvantages the out-group from opportunities in mentoring and development [16].

Further, the exclusion of out-group members in the discussion can develop into prejudice and discrimination. Bochner and Hesketh [4] surveyed 263 employees of an Australian bank to understand inter-ethnic work-related friction. Cultural diversity was defined by nationality. The analysis compared between Anglo-Celtic vs. non Anglo-Celtic employees. Anglo-Celtic employees referred to those who identified themselves as Australian, British, New Zealanders and “other western countries”. Non Anglo-Celtic employees were those who identified themselves as from Asian, Indian or Middle Eastern nations. Following the Australian societal context, Anglo-Celtic employees were defined as the in-group. Non Anglo-Celtic employees were defined as the out-group. Comparing between in-group vs. out-group employees, the authors found that the out-group reported perceiving more discrimination and inequality.
Hence, for group members who are willing and able to contribute to the discussion, exhibition of prejudice and discrimination in their teams will stymie their contribution in the team.

2.4 Research Issues

As discussed above, cultural diversity increases the project team’s potential to generate more creative and high quality decisions. However, cultural diversity also brings communication issues that impede on the realization of process gain. Discussion domination reduces air-time for important contributions to be heard. A low proficiency in the common language used during discussion reduces team members’ ability to articulate their ideas and air their views on problems they foresee with decisions made. Team members excluded from the discussion will not be able to contribute to the discussion. Their ability to excel in the role they play in the team is also impeded as they are not kept in the information loop. If exclusion develops into prejudice and discrimination against ethnically different team members, their willingness and ability to contribute to the discussion will be stymied.

This study investigates whether groups composed of different nationality and ethnicity mix will experience different types of the most common communication problems, and/ or different degrees of the most common communication problems highlighted above.

3 Study

Three teams were formed through purposive sampling and given a collaborative design task to complete in laboratory conditions. The entire process was captured on video-tape and their behaviors were coded for analysis.

3.1 Participants in Experimental Conditions

Nine engineering undergraduates were recruited via snowballing to participate in the study. They were rewarded with a small cash reward for completing the assignment. All had attended the relevant engineering design courses that provided the knowledge for collaborative design. In addition to the homogenous discipline, participants selected for this research study were in their second year of undergraduate study. They aged between 20 to 24 years of age. As the cohort was big, care was taken to select participants who did not know each other prior to the study.

Nationality and ethnicity were manipulated to generate three groups. Group One comprised of three Singaporean Chinese participants. Group Two comprised of three Singaporean participants with varying ethnicity – Chinese, Indian and Malay. Group Three comprised of three participants from Singapore, Malaysia and China but with Chinese ethnic culture. Table 1 gives a summary of the group composition:

Table 1. Group Composition

Group 1	Group 2	Group 3
Same Cultural Background	Different Cultural Background (Ethnic)	Different Cultural Background (National)
- Same nationality - Same ethnicity	- Same nationality - Different ethnicity	- Different nationality - Same ethnicity

3.2 Procedure

Teams were tasked to construct a model airplane within four sessions over a period of two weeks. Each session was one hour in duration and two sessions were conducted each week. Participants were given an instruction guide at the start of the experiment detailing the deliverables for each session.

Session 1 was allocated for participants to familiarize with each to other and undergo first half of the planning stage. Participants were expected to complete the functional analysis and brainstorm about the airplane modeling. In the second session, participants continued with the brainstorming session and were required to complete their discussion and finalized the model of the airplane by the end of the session. In the third session, participants constructed the model airplane. The last session was the testing phase. The model airplanes were tested for their flight capability. Participants were allowed to make modifications if their planes that did not meet the required flight distance of three meters.

3.3 Coding and Measures

The discussion sessions were recorded on videotape. The videotapes were then coded by trained coders. Coders worked simultaneously but independently from one another. Each of the measures was coded by two coders. In addition, one of the coders timed the speaking duration for each participant in the experiment using a stop-watch. Inter-coder reliability was reasonable (Spearman's rho .293, .307 and .406) and there was no significant difference between the coders' analysis of each of the measures.

Discussion dominance. Discussion dominance was assessed via total number of successful interruptions during discussion. Groups where members spent significantly more time interrupting each other are considered to have a greater problem with discussion domination than other groups [27]. The greater number of successful interruptions, the greater the problem the group faces.

Coders measured the number of successful interruptions in each group. Groups with a higher occurrence of successful interruptions were plagued by discussion dominance compared to groups where members interruption each other less often. Coders first identified areas of overlapping speech between team members. Interruptions were coded as successful when the interrupter completed an utterance and prevented the interruptee from completing an utterance. When one or both of these conditions were not met, an unsuccessful interruption was coded [21]. Discussion domination hence consists of a within-group comparison as well as a between-group comparison.

Communication frequency. Most organizational studies examine communication frequency amongst group via self-report. Team members are asked how often they communicate with each other, e.g., daily, weekly, etc [1, 9, 32].

However, for this study, team members are only allowed to communicate with each other within the laboratory session and not outside. Hence, this study operationalizes communication frequency differently. Communication frequency is assessed via total time spent on discussion and total number of turn taken during discussion. Both are between-group measures. The first compares the total amount of time that the group spoke, between groups; the second compares the total number of turns that the group took, between groups.

Discussion exclusion. Discussion exclusion is defined as when team members ignore someone who is present from the ongoing conversation [11]. Discussion exclusion was assessed by counting the number of times a communication act was ignored or not reciprocated. This was a between-group assessment. In other words, discussion exclusion was deemed to be a problem in the group when there were instances of non-reciprocal communication. The analysis here was to assess if discussion exclusion was a problem of a greater degree in certain groups more than others.

4 Results

One-way Anova was computed to analyze for communication problems that each group experienced. All communication problems were observed in all three groups. However, communication problems occurred more frequently in different groups. Groups with different cultural composition experienced different degrees of the same communication problem. Group 2 faced significantly greater degree of discussion domination and discussion exclusion than Groups 1 and 3.

4.1 Discussion Dominance

In all three groups, there was a significant difference between the groups in terms of the number of successful interruptions made, $F(2, 9) = 11.50$, $p < .05$. Group 2 has significantly greater number of successful interruptions than Groups 1 and 3. Hence, there was a difference in the degree of discussion dominance between groups.

4.2 Communication Frequency

There was no significant difference between the groups in terms of the number of turns taken $F(2, 9) = 3.42$, $p > .05$. There was also no significant differences between the groups in terms of speaking duration, $F(2, 9) = 2.12$, $p > .05$.

4.3 Discussion Exclusion

There was a significant difference in discussion exclusion between the groups. Group 2 experienced significantly more instances of discussion exclusion than the other groups, $F(2, 9) = 4.65$, $p < .05$.

Table 2 gives a summary of the communication problems experienced in each group.

Table 2. Differences among three groups

	Group 1	Group 2	Group 3
	Same Cultural Background	Different Cultural Background (Ethnic)	Different Cultural Background (National)
Discussion Domination			
Successful Interruptions	--	*Significantly MORE*	--
Communication Frequency			
Speaking Time	No Difference	No Difference	No Difference
No. of Turns	No Difference	No Difference	No Difference
Discussion Exclusion			
No response	--	*Significantly MORE*	--

5 Discussion

Compared to past studies which show that groups composed of culturally different members experience more problems than groups composed of culturally similar members, this study shows that all groups experience communication problems, albeit different ones. The group composed of culturally similar members experienced more reduced communication frequency. The group composed of members with different ethnic background experienced more discussion exclusion and dominance. The group composed of members with different national background experienced reduced communication frequency.

It is observed in this study that the impact of nationality on group communication is different from the impact of ethnicity on group communication. As can be seen from Table 2, the different national background group (i.e., Group 3) experienced relatively equitable speaking time and turns. There was less discussion exclusion and interruptions than the different ethnic backgroun group (i.e., Group 2). In observation of the group discussions, it is found that members of the different national culture group spent more time than other groups in explaining their experiences with model airplanes and how they came up with the ideas that they were proposing to the group.

This suggests that nationality gives the common background necessary for coming to a common understanding of what we are trying to communicate. Heterogeneous groups with different national backgrounds experience communication problems related to trying to understand where an idea, phrase, or joke originates and how to make sense of it. This is a situation that heterogeneous groups with different ethnicities from the same country have less of a problem with, having been educated in similar conditions and having access to similar entertainment media and lifestyles. This suggests that future research should employ both qualitative and quantitative understanding of communication problems that heterogeneous groups face rather than rely on numerical measures alone.

As can be seen from Table 2, there are more instances of discussion exclusion and dominance in Group 2, where members are from different ethnicity, than in Groups 1 and 3, where group members come from the same ethnic group. Discussion exclusion and dominance are part and parcel of the dynamics of open conflict strategies during discussion. When group members openly challenge each others' ideas and allow disagreements to take place, interrupting another group member or ignoring another group member is bound to take place. This indicates that Group 2 utilized more open conflict strategies than Groups 1 and 3.

We propose that open conflict strategies were utilized more in Group 2 than Groups 1 and 3 because cultural differences in conflict strategies are more apparent between ethnicities than between nationalities. In other words, we propose that there are greater differences in conflict strategies between individuals from different ethnic group but same nationality than between individuals from different nationality but of the same ethnic group. Oetzel, Ting-Toomey, et al [20] conducted a cross-cultural comparison of Germans, Japanese, Mexicans and U.S. Americans in terms of the face and facework conflicts between parents and siblings. The authors found that individualistic, small power distance cultures use more dominating and integrating facework and less avoiding facework, giving evidence of cultural differences in conflict strategies. An interesting point for future similar-topic research would be to de-couple nationality from ethnicity. Oetzel, Ting-Toomey, et al [20] selected cases where ethnicity and nationality were the same, i.e., Japanese person is Japanese in ethnicity and Japanese in nationality. Based on our proposal, it would be interesting to conduct more research to affirm if there are indeed greater differences in conflict strategies between individuals from different ethnic group but same nationality than between individuals from different nationality but of the same ethnic group.

The above elucidates that ethnicity and nationality give rise to different communication problems. Therefore, future studies should be specific when defining hetero geneous groups in terms of whether they are manipulating differences in nationality, ethnicity or both. Different group compositions will give rise to different problems as well as different degrees of the same problem.

As this is an exploratory study, the sample size was small. There is much opportunity for future research to address its limitations. Future research can be done with more ethnic and nationality composition mixes to ascertain the different types of problems that different group compositions bring. In addition, more research is needed to understand why different ethnic and nationality composition result in different types and degrees in communication problems encountered by groups.

References

1. Becerra, M., Gupta, A.K.: Perceived trustworthiness within the organization: The moderating impact of communication frequency on trustor and trustee effects. Organization Science 14(1), 32–45 (2003)
2. Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., Kraut, R.: Using Social Psychology to Motivate Contributions to Online Communities. In: ACM CSCW 2004 Conference on Computer Supported Cooperative Work, Chicago, IL, pp. 212–221 (2004)
3. Blau, P.M.: Inequality and heterogeneity: A primitive theory of social structure. The Free Press, New York (1977)
4. Bochner, S., Hesketh, B.: Power distance, individualism-collectivism, and job related attitudes in a culturally diverse work group. Journal of Cross-Cultural Psychology 25(2), 233–257 (1994)
5. Chiu, M.L.: An organizational view of design communication in design collaboration. Design Studies 23(2), 187–210 (2002)
6. Cox, T.H., Lobel, S.A., McLeod, P.L.: Effects of Ethnic Group Cultural Differences on Cooperative and Competitive Behavior on a Group Task. Academy of Management Journal 34(4), 827–847 (1991)
7. Hirokawa, R.Y.: Group communication research: Considerations for the use of interaction analysis. In: Tardy, D.H. (ed.) A Handbook for the Study of Human Communication, Ablex, Norwood, NJ, pp. 118–134 (1987)
8. Hoffman, E.: The Effect of Race-Ratio Composition on the Frequency of Organizational Communication. Social Psychology Quarterly 48(1), 17–26 (1985)
9. Karau, S., Williams, K.: Social loafing: A meta-analytic review and theoretical integration. Journal of Personality and Social Psychology 65(4), 681–706 (1993)
10. Kraut, R., Egido, C., Galegher, J.: Patterns of contact and communication in scientific research collaboration. In: ACM Conference on Computer-Supported Cooperative Work (1998)
11. Larkey, L.K.: Toward a theory of communicative interactions in culturally diverse workgroups. Academy of Management Review 21(2), 463–491 (1996)
12. Latané, B., Williams, K., Harkins, S.: Many hands make light the work The causes and consequences of social loafing. Journal of Personality and Social Psychology 37(6), 822–832 (1979)
13. Loosemore, M., Lee, P.: Communication problems with Ethnic Minorities in the construction industry. International Journal of Project Management 20(7), 517–524 (2001)
14. March, J., Simon, H.: Organizations. John Wiley and Sons, New York (1958)

15. Mast, M.S.: Dominance as expressed and inferred through speaking time: A Meta Analysis. Human Communication Research 28(3), 420–450 (2002)
16. Morrison, A.M.: The new leaders: Guidelines on leadership diversity in America. Jossey-Bass, San Francisco (1992)
17. Oetzel, J.G.: Culturally homogeneous and heterogeneous groups: Explaining communication processes through individualism-collectivism and self-construal. International Journal of Intercultural Relations 22, 135–161 (1998)
18. Oetzel, J.G.: Self-Construals, Communication Processes, and Group Outcomes in Homogeneous and Heterogeneous Groups. Small Group Research 32(1), 19 (2001)
19. Oetzel, J.G., Burtis, T.E., Sanchez, M.I.C., Perez, F.G.: Investigating the Role of Communication in Culturally Diverse Work Groups: A Review and Synthesis. Lawrence Erlbaum Mahwah, NJ (2001)
20. Oetzel, J.G., Ting-Toomey, S., Chew-Sanchez, M.I., Harris, R., Wilcox, R., Stumpf, S.: Face and facework in conflict with parents and siblings: A Cross-cultural comparison of Germans, Japanese, Mexicans and US Americans. The Journal of Family Communication 3(2), 67–93 (2003)
21. Reid, S.A., Ng, S.H.: The dynamics of intragroup differentiation in an intergroup social context. Human Communication Research 32, 504–525 (2006)
22. Tajfel, H., Turner, J.C.: The social identity theory of intergroup conflict. In: Austin, W.G., Worchel, S. (eds.) Psychology of Intergroup Relations. Nelson-Hall, Chicago (1986)
23. Turner, J.C.: Rediscovering the social group: Self-categorization theory. Blackwell, Oxford (1987)
24. Watson, W.E., Kumar, K.: Differences in decision making regarding risk taking: A comparison of culturally diverse and culturally homogeneous task groups. International Journal of Intercultural Relations 16, 53–65 (1992)
25. Watson, W.E., Kumar, K., Michaelsen, L.K.: Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. Academy of Management Journal 36, 590–602 (1993)
26. Zenger, T.R., Lawrence, B.S.: Organisational demography: The differential effects of age and tenure distributions on technical communication. Academy of Management Journal 32(2), 353–376 (1989)
27. Zimmerman, D.H., West, C.: Sex roles, interruptions, and silences in conversation. In: Thorne, B., Henley, N. (eds.) Language and sex: Difference and dominance, Newbury House, Rowley, MA (1975)

Theoretical Model of User Acceptance: In the View of Measuring Success in Web Personalization

Mohd Afandi Md Amin and Richi Nayak

Faculty of Science and Technology,
Queensland University of Technology,
Brisbane, Australia
mohd.mdamin@student.qut.edu.au, r.nayak@qut.edu.au

Abstract. This paper attempts to develop a theoretical acceptance model for measuring Web personalization success. Key factors impacting Web personalization acceptance are identified from a detailed literature review. The final model is then cast in a structural equation modeling (SEM) framework comprising nineteen manifest variables, which are grouped into three focal behaviors of Web users. These variables could provide a framework for better understanding of numerous factors that contribute to the success measures of Web personalization technology. Especially, those concerning the quality of personalized features and how personalized information through personalized Website can be delivered to the user. The interrelationship between success constructs is also explained. Empirical validations of this theoretical model are expected on future research.

Keywords: Web personalization, measuring success, adoption, measures.

1 Introduction

Information overload [1], [2], and [3] is a degree of complexity information processing by the user in most collaboration technologies, as well in Web environment. The overload problem is resulting from the diffusion of the Web and the huge amount of information available online [4], and [5], since users have to find relevant, needed, useful and personalized information. Thus, the problem has given advance to the compelling required for Web systems able to assist users intelligently, when they browse through the Web. Web personalization offers this precious opportunity, representing one of the most influential technologies required by an ever increasing number of real-world applications. Consequently, Web personalization is momentous research issue in Web applications, as well as has been a key property for online news providers, ecommerce Websites, and technical data providers.

In addition, measuring success or effectiveness of the Web personalization system involves the defining metrics and feedback techniques, from two endeavour; computational intelligence (CI) and Information systems (IS). CI revealed on computational approaches that employed several algorithms and techniques [4], Whereby, IS focusing on the user's behavior of the systems that based-on the theoretical foundation

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 255–264, 2010.

from diverse disciplines such as psychology, computer sciences, business and statistics. Furthermore, IS studies remain on user's behavior of the Web personalization systems such as whether online firms can use personalization as a marketing strategy to attract new users [6], [7], and [8], personality traits and perception towards the influence of user's behavior on the Website.

This paper is structured as follows. Firstly, the theoretical underpinning used to develop the model is presented. The four prominent theories of measuring success are examined: DeLone and McLean Information Success model (IS-success), theory of acceptance Model (TAM), theory of planned behavior (TPB), and Unified theory of Acceptance and Use Technology (UTAUT). Then the possibilities of success measures in Web personalization are presented from those theories. Secondly, the TPB is preferred to be a framework for measuring success in Web personalization, followed by methodology to be used in the future research. Finally, we provide a discussion of the findings and the agenda for future research.

2 Measuring Success in Web Personalization

Websites today, have access to the incredible amount of data about the visitors and users to their Websites, the preferences and their behavior [9], [10], and [11]. There are numbers of different approaches and architectures has been employed in building support for such personalization systems, each of which has different strengths and weaknesses [12],[10], and [13].

The need for measuring success of the Web personalization is not only beneficial for vendors or providers but also has been remarkable to the users. For instance, the vendors need to know how success or effective their delivered personalized features on the Web, while feedback from the users can be an input for vendors to enhance their personalized services through Website. In a same way, the adoptions of personalized services by the users are also crucial, in terms of personalization strategies, especially for online vendors.

2.1 The Significance of Measuring Success

In this section, we define how measuring success in Web personalization is crucial from three perspectives: (1) the developers, (2) the Website owner, and (3) the user. Effective Web personalization has become a prominent issue due to the pervasiveness of e-commerce applications [14], [15], and [16]. Most of the Web services have multiple stakeholders, for example, developers, users and investors. Therefore, measuring success or effectiveness of the services has also multiple definitions from several views. For developers, success of the services may be one is completed in time or within the budget, with a complete set of features that are consistent with predefined specification and that functions correctly. On the other hand, from a Website owner, success for them is if the personalization services can reduce uncertainty outcomes, attracts a large, loyal and growing community of users. Additionally, the effective contents can be gained if the right person receives the right message at the right time and the right context [17]. In contrast, for users success is if personalization services able to improve their task and easy to use.

There is much significance of measuring success in Web personalization. Particularly, we summarize the significance of measuring success of Web personalization by considering of generalizing the adoption, implementation and the use of personalization techniques as presented in Figure 1 below.

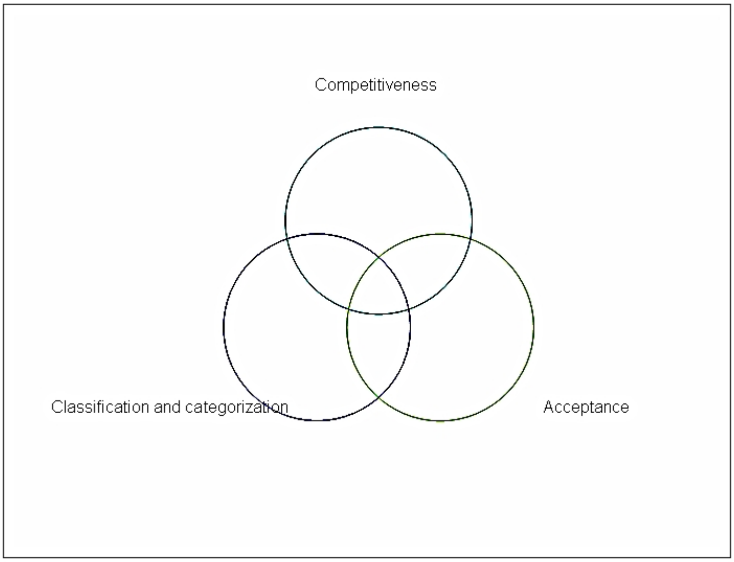


Fig. 1. Effectiveness of Web personalization

From the Figure 1, the significance of measuring success of Web personalization is proposed in three components:

2.1.1 The Ability to Classify and Categorize the Popular Features of the Website

The ability for classification and categorization of popular features is one of the most salient aspects to determine whether or not enterprises implement the Web personalization system [18]. It depends on personalization features on the Website, for example, the more personalization features will lead the effective personalization agent for classify and tailoring the contents or products to suit with user's preferences from user profile, explicit and implicitly. In terms of cost, the more sophisticated of personalized agents will increase the cost that needs to be allocated by the enterprises or Website owner. Therefore, measuring success is vital to acknowledge enterprises about the effectiveness of the personalized system and related with the cost benefit analysis.

Many studies have been conducted on measuring these features, which categorized in computational intelligence (CI) research. CI paradigms reveal to be potential tools to face under Web environment, which handle Web usage data and develop Web-based applications tailored on users' preferences [4]. Recently, classification and categorization of the usage data as well as Web contents, has been studied under the application of data mining techniques for Web data, namely Web mining. The most popular techniques have been used is clustering [4]. Clustering techniques look for a group of similar items among huge of data base on a general idea of distance. This technique computes the similarity between items [19].

2.1.2 The Competitiveness of the Website Provider

Measuring success in Web personalization, particularly in most e-commerce Websites is noteworthy since it reflects the competitiveness of the enterprises. E-commerce mainly focused on the sale of goods and with the advent of Web technology, it has been expanded to deal with all aspects of business interaction, at the individual and enterprise level. In terms of cost benefit analysis, the more competitive the market structure, the complicated the decision of whether or not to implement a personalization strategy, or proceed with the current personalization strategy on the Website. If the personalization strategy is not successful then the enterprises could suffer severe damage, in such having a negative value of the return of investment (ROI).

The studies revealed in comparing personalization strategy and marketing prospect, falls in marketing research, since personalization is the process of gathering information explicitly or implicitly about customer or user, which enable the enterprise to target products or recommendations that best match the user's preferences [20]. Research shows that the Web is particularly suited for personalized services [21] compare to other media (e.g. newspapers and television). For instance, recent empirical evidence indicates that about 80% of Internet users are interested in personalized services [22], a well as 56% of frequent online shoppers were more likely to make a purchase on a Website that offered personalization features [23].

2.1.3 Acceptance and Use of Web Personalization by the User

Acceptance refers to how users accept and favor the personalization features through personalization systems. The acceptance Web personalization by the user is needed to be measured, since it indicates to what extent that the personalization systems are use and continuing use by the user. Although the success evaluation is necessary, however, it is difficult to measure, since it is influenced by various factors such as customer usage, customer skills, and ease of use of the systems as well as usefulness of the systems.

Evaluation or measuring success can also be viewed as effectiveness and impact of the system that perceived by the users. According to [24], such evaluation of the system is based on design science research. Therefore, it concerns about evaluation of outputs, including theory and artefacts.

3 Theoretical Underpinning of Measuring Success

In this section, we introduce four prominent theories and model in Information Systems (IS) research for measuring success: IS-success model, theory of acceptance model (TAM), and theory of planned behavior (TPB), and unified theory of acceptance and use of technology (UTAUT).

3.1 Information Success Model (IS-Success)

The first theory examine is the Information Systems success model (IS-success), The IS-success model was the comprehensive IS success model for measuring IS-impact and has been introduced by [25], as a success framework with complex-dependent variables in IS research. The model shows an interrelationship between six IS success variables categories namely: (1) system quality, (2) information quality, (3) IS use, (4) user satisfaction, (5) individual impact, and (6) organization impact.

3.2 Technology Acceptance Model (TAM)

The second theory examine is technology acceptance model (TAM). The TAM was introduced by [26], is an outgrowth of the model of individual behavior as posited by [27], theory of Reasoned Action (TRA) and Theory of Planned Behavior (TPB), which showed the influence of the perceived ease of use and the perceived usefulness of a technology on the user's attitudes toward using the technology and subsequently on the actual usage thereof.

3.3 Theory of Planned Behavior (TPB)

Next, the TPB is examined. TPB is an extension of theory reasoned action (TRA) is the most influential theory in explaining and predicting behavior. The theory of planned behavior is one of the most influential theories in projecting human behavior across many settings [28] and has been validated by prior research. According to TPB, the direct root of any Behavior is its Behavioral Intentions; after all, people do what they plan to do. Behavioral intention is defined as "the strength of one's intention to perform a specified behavior [27]. According to [28], "human action is guided by three kinds of considerations: beliefs about the likely outcomes of the behavior and evaluations of these outcomes (behavioral beliefs), beliefs about the normative expectations of others and motivation to comply with these expectations (normative beliefs), and beliefs about the presence of factors that may facilitate or impede performance of the behavior and the perceived power of these factors (control beliefs)". Figure 2 below described the TPB.

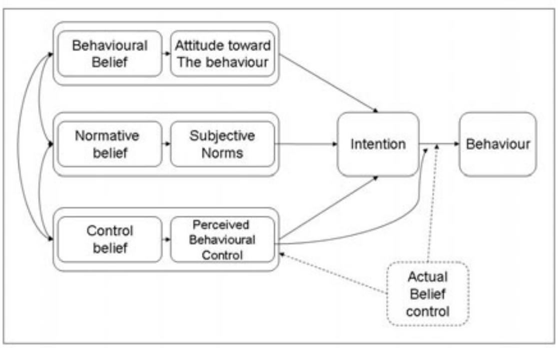


Fig. 2. The constructs of TPB [28]

3.4 Unified Theory of Acceptance and Use of Technology (UTAUT)

Finally, the composed model that integrated theories and model from unified view of user acceptance is examined. The latest work proposed by [7] integrates eight models from fragmented view to unified view that collaborated the major theories and models in user acceptance. Instead of three theories that have been discussed above, UTAUT utilized the following models: the Theory of Reasoned Action (TRA), the model of PC Utilization (MPCU), the motivation models (MM), the Innovation and Diffusion

Theory (IDT), and Social Cognitive Theory (SCT). In UTAUT, there are four antecedents used to describe user acceptance toward technology: performance expectancy (PE), effort expectancy (EE), social influence (SI), and facilitating conditions (FC). Four moderators i.e. gender, age, experience, and voluntaries of use are employed to describe various relationships between antecedents and behavioral intentions (BI), for example, gender are expected to have influences in PE, EE, and SI. Whereby, age expected to influence in all four antecedents.

4 Theoretical Model and Methodology

4.1 TPB as a Framework

From the four theories that defined success, TPB has been chosen as a framework for constructing the Web personalization success model. There are some reasons to select the TPB for the theoretical framework among other success theories. Firstly, the TPB has three elements that defined the cause toward user's behavior on the system: (a) attitude toward behavior, (b) subjective norms, and perceived behavioral control. These three elements described the behavior of the users toward intention to use of the Web personalization system. The first element described the idea that by using the personalization will increase performance of the user on the Website e.g. finding information, products, or articles. This feeling of the users will lead attitude to use the system. Secondly, subjective norms are related to user's decision on the system will be affected by other user's opinion to use personalization, and last element, perceived behavioral control defined by perception about how personalization system work as well as it depends on other factors such as time, experience, security and privacy manners that influence the use of personalization system by the users.

Thirdly, it was selected because it shows the relationships of the success measures of Web personalization, since we argue that success is multidimensional phenomenal of a user using a personalized Website, where they engage in a complex set of behavior in three behavioral intentions between acquiring information from a Website, giving information to a Website, and navigating through a Website. These multidimensional construct views are also agreed with the previous research on measuring success in several areas such as business modeling [29], and Information systems [30],[31],[32],[33],[34],[35],[36],[37], and [38].

4.2 Methodology

We expect a positive relationship for the three focal behaviors of Web users: (1) acquiring information, (2) giving information, and (3) navigating on the personalized Website. From the previous study, we argue that the effect of web personalization to the users are also related to personality traits on choice behavior [39], [40], and [6], which described how users interact with transaction driven personalization. In order to identify antecedents of web personalization adoption TPB is used. According to [41] and [28], the antecedents of attitudes, subjective norm, Perceived Behavioral Control (PBC) is a set of attitudinal, normative and control belief, respectively. However, normative beliefs were dropped due to it only specify the referent others (e.g. family, friends, or society), which is believed will not influence a user's attitude and control toward using Web personalization. Additionally, other researchers argued that in the context of using technologies, subjective norms did not seem to be a significant predictor of intentions [26, 42].

Moreover, several studies on the e-commerce Website have excluded subjective norms, focusing only on ease of use and usefulness. They argued that intention of browsing and use applications by an individual is not impacted from other individuals. This agree with finding about browsing a particular Website (e.g. personalized Website) is a private affair and not visible to peers or friends. Peers or friends influence might impact on Website in general, but not to revisit a Website. Thus, in this study, subjective norms are excluded to have influence in behavioral intentions of use a personalized Website.

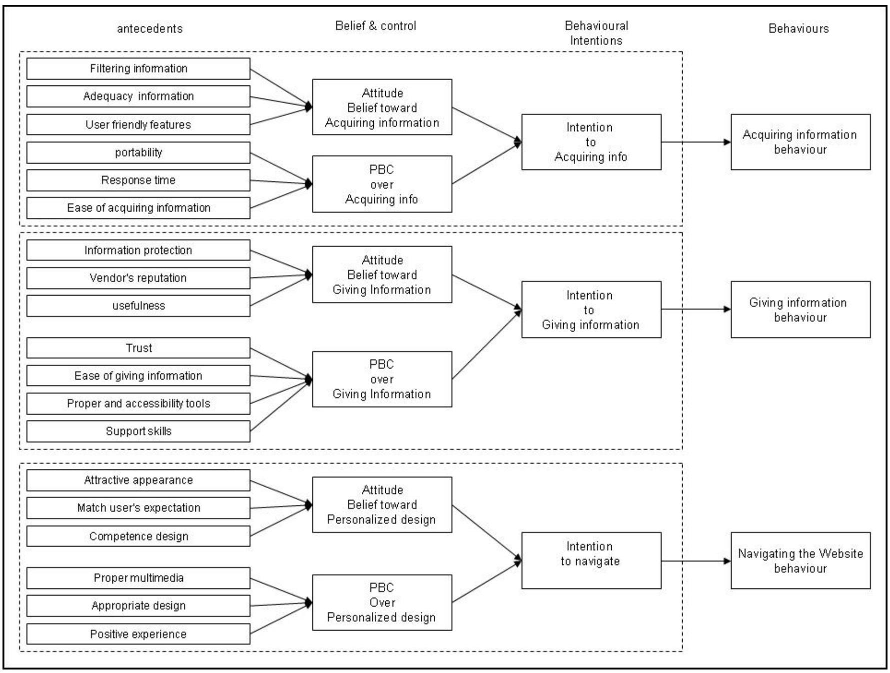


Fig. 3. Web personalization success model

Therefore, the Web personalization success model is proposed in Figure 3, as an extension of TPB. The model extends TPB from three focal user's behavior: acquiring information from personalized Website, giving information, and perceived on personalized designed.

As depicted in Figure 3, the success constructs are defined in antecedents' part, where belief and control, intentions, and user's behaviors are presented on the right side, respectively. There are nineteen (19) antecedents of success have been defined in three user's focal behavior dealing with personalized Website

According to the relationship between variables, the corresponding hypotheses among all variables will be developed. Then we further study the meaning of all variables and develop measurement scales for each variable. Once the full model testing is complete, the structural equation modeling (SEM) will be used to define the reliability and validity of the proposed model, particularly, by employing the partial least square path modeling (PLS-PM) for the structural and measurement model.

We summarized the success measures in Web personalization as presented in Figure 3, in Table 1 below:

Table 1. Success Measures

User's focal behavior	Constructs	Success Measures
Acquiring information	Attitude belief	Information filtering Intrusiveness User-friendly features
	Perceived behavioral control	Portability Response time Ease of acquiring information
Giving information	Attitude belief	Information protection Vendor's reputation Usefulness of giving information
	Perceived behavioral control	Trust Ease of giving information Proper and accessibility tools Support skills
Personalized design	Attitude belief	Attractive appearance Match user's expectation competence
	Perceived behavioral control	Proper multimedia capability Appropriate design Positive experience

5 Conclusions and Future Research

Web personalization is a prominent approach for tailoring user's need on the Web, since due to the uncertainty of information on the Web. This research is going to improve and define the recognition of user acceptance behavior by advancing and validating the theoretical model of measuring success in Web personalization.

The study postulates success measures in Web personalization, particularly from human contributions through user's acceptance and adoption of Web personalization technology. Therefore, this theoretical model can help personalized Web developers to define their strategy on delivering personalized information on the Web. This strategy is based on three user's focal behavior: acquiring information through personalized Website, giving personal details on the personalized Website (e.g. filling online form for registration on ecommerce Website), and navigating through personalized Website. The success strategy is reflected by good personalized design, including proper multimedia capability applications, competence and professional design, and as well as the personalized design match user's expectation about personalized information. In the future, all defined measures will be constructed in the form of survey questionnaires.

References

1. Kosala, R., Blockeel, H.: Web Mining Research: A Survey. In: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining (2000)
2. Paul, S., Nazareth, D.L.: Input information complexity, perceived time pressure, and information processing in GSS-based work groups: An experimental investigation using a decision schema to alleviate information overload conditions. Decision Support Systems, Corrected Proof (2010) (in Press)
3. Wheeldon, R., Levene, M., Keenoy, K.: Search and Navigation in Relational Databases (2003)
4. Castellano, G., Fanelli, A.M., Torsello, M.A.: Computational Intelligence techniques for Web personalization. Web Intelligent and Agent System 6, 253–272 (2008)
5. Castellano, G., Fanelli, A.M., Torsello, M.A., Jain, L.: Innovations in Web Personalization: Web Personalization in Intelligent Environments, pp. 1–26 (2009)
6. Ho, S.Y., Michael, J.D., Tam, K.Y.: Personalization and choice behavior: the role of personality traits. SIGMIS Database 39, 31–47 (2008)
7. Venkatesh, M., Morris, K., Davis, F.: User acceptance of information technology: Toward a unified view. MIS Quarterly 27, 425–478 (2003)
8. Venkatesh, M., Ramesh, K.: Web and Wireless Site Usability: Understanding Differences and Modeling Use. MIS Quarterly 30, 181–206 (2006)
9. Avery, C., Zeckhauser, R.: Recommender systems for evaluating computer messages. Communications of the ACM 40, 88–89 (1997)
10. Peyton, L.: Measuring and managing the effectiveness of personalization. In: Proceedings of the 5th International Conference on Electronic Commerce Pittsburgh. ACM, Pennsylvania (2003)
11. Shearin, S., Lieberman, H.: Intelligent profiling by example. In: Proceedings of the 6th International Conference on Intelligent User Interfaces, Santa Fe, New Mexico, United States, pp. 145–151 (2001)
12. Paganelli, L., Paterno, F.: Intelligent analysis of user interactions with web applications. In: Proceedings of the 7th International Conference on Intelligent user Interfaces. ACM, San Francisco (2002)
13. Scholtz, J., Laskowski, S., Downey, L.: Developing usability tools and techniques for designing and testing web sites. In: Proceedings HFWeb 1998, Basking Ridge, NJ (1998)
14. Goy, A., Ardissono, L., Petrone, G.: Personalization in E-Commerce Applications. In: The Adaptive Web, pp. 485–520 (2007)
15. VanderMeer, D., Kaushik, D., Anindya, D., Krithi, R., Shamkant, B.N.: Enabling scalable online personalization on the Web. In: Proceedings of the 2nd ACM Conference on Electronic Commerce Minneapolis. ACM, Minnesota (2000)
16. Yu, P.S.: Data Mining and Personalization Technologies. In: Proceedings of the Sixth International Conference on Database Systems for Advanced Applications. IEEE Computer Society, Los Alamitos (1999)
17. Kazienko, P., Adamski, M.: AdROSA- Adaptive personalization of Web advertising. International Journal of Information Sciences 177, 2269–2295 (2007)
18. Greer, Murtaza: Web personalization: the impact of perceived innovation characteristics on the intention to use personalization. Journal of Computer Information Systems 43, 50–55 (2003)
19. Vakali, A., Pokorny, J., Dalamagas, T.: An overview of web data clustering practices. In: Lindner, W., Mesiti, M., Türker, C., Tzitzikas, Y., Vakali, A.I. (eds.) EDBT 2004. LNCS, vol. 3268, pp. 597–606. Springer, Heidelberg (2004)

20. Nunes, P.F., Kambil, A.: Personalization? No Thanks. Harvard Business Review (2001)
21. Risch, D., Schubert, P., Leimstoll, U.: Classification of Personalization Functions for E-Commerce Applications. In: CollECTeR Adelaide (2006)
22. Kobsa, A.: Privacy-Enhanced Web Personalization. Communiaction of the ACM 30, 24–33 (2007)
23. Freedman, L.: Merchant views of personalization and lasting consumer relationships (2007)
24. Pries-Heje, J., Richard, B., John, V.: Strategies for design science research evaluation. In: Proceedings of the 16th European Conference on Information Systems (ECIS 2008), Ireland (2008)
25. DeLone, W.H., McLean, E.R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3, 60–95 (1992)
26. Davis, F.D., Bagozzi, R.P., Warshaw, P.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35, 982–1003 (1989)
27. Ajzen, I., Fishbein, M.: Attitudinal and normative variables as predictors of specific behavior. Journal of Personality and Social Psychology 27, 41–57 (1973)
28. Ajzen, I.: Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. Journal of Applied Social Psychology 32, 665–683 (2002)
29. Bandara, W.: Process Modeling Success Factors and Measures, in Information Systems. Ph.D Brisbane, Australia: Queensland University of Technology, p. 632 (2007)
30. Bhattacherjee, A.: Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly 25, 351–370 (2001)
31. Davison, J., Deeks, D.: Measuring the potential success of Information Systems Implementation. Journal of Measuring Business Excellent 11, 75–81 (2007)
32. DeLone, W.H., McLean, E.R.: The DeLone and McLean model of information systems success: a ten-year update. Journal of Management Information Systems 19, 9–30 (2003)
33. Gable, G.G., Sedera, D., Chan, T.: Enterprise System Success: A measurement Model. In: Twenty-Fourth International Conference on Information Systems (2003)
34. Ifinedo, P.: Impacts of Business Vision, Top Management Support, and External Expertise on ERP Success. Business Process Management Journal 14, 551–568 (2008)
35. Limayem, M., Cheung, C.M.K.: Understanding information systems continuance: The case of Internet-based learning technologies. Information & Management 45, 227–232 (2008)
36. Seddon, P., Sandy, S., Ravi, P., Matthew, B.: Dimensions of information systems success. Communications of AIS 2, 5 (1999)
37. Taylor-Cummings, A., Feeny, D.: The development and implementation of systems: bridging the user-IS gap: Managing IT as a Strategic Resource. In: Willcocks, L., Feeny, D., Islei, G. (eds.), pp. 71–202. McGraw Hill, Maidenhead (1997)
38. Wong, B., Arjpu, C.: A Study of How User Satisfaction and User Dissatisfaction Affect the Success of an Information System. In: Australasian Conference on Information Systems, Toowoomba, QLD, pp. 801–812 (2007)
39. Finn, A., Louviere, J.J.: Shopping center patronage models: fashioning a consideration set solution. Journal of Business Research 21, 259–275 (1990)
40. Gensch, D.H.: A Two-Stage Disaggregate Attribute Choice Model. Marketing Science 7, 299–310 (1987)
41. Ajzen, I., Fishbein, M.: Understanding attitudes and predicting social behaviour. Prentice-Hall, New Jersey (1980)
42. Loiacono, E.T., Watson, R.T., Goodhue, D.L.: WebQual: An Instrument for Consumer Evaluation of Web Sites. International Journal of Electronic Commerce 11, 51–87 (2007)

An Approach to Information Presentation Employing Game Principles and Physics Based Interaction

Mária Bieliková, Michal Lohnický, and Daniel Švoňava

Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava,
Ilkovičova 3, 842 16 Bratislava, Slovakia
{name.surname}@stuba.sk

Abstract. In this paper we propose a novel approach optimized towards the presentation of general sets of information entities that employs game-like features and social context to engage its users to explore the presented data and streamlines the navigation and overall usability in terms of several constraints that we established. The presentation environment is embedded within a physical engine that controls the movement of visualized entities, resolves user's interaction and takes part in the overall information throughput.

Keywords: web information entities visualization, exploratory navigation, game-like, physics based interaction.

1 Visualization Concept

The times when the web content was composed mostly of textual HTML-based websites are long gone. With some popular technologies like HTML5 the web has become a highly visual medium. However, the mere fact that the information is communicated by a visual presentation does not solve any of the major problems of crowded and boring 1.0 websites. We need to utilize the power of the new web to ease the information retrieval. The current information visualizations are often ontology based visualizations utilizing hierarchical navigation [1], graphs [2] and various aspects of information space [3]. Their disadvantages are excessive complexity or lack of scalability.

We propose a novel approach to presentation of general sets of information entities. We map the data entities on graph nodes, embedded in 3D space with edges visually encoded in spatial composition (see Fig 1). To engage the users we use game-like features [4] and to meet the navigating, filtering and clustering requirements we embed the presentation space with physical engine. The basic requirement of the dynamic information presentation is to simplify exploration of complex informational structures. To improve the clarity, it is necessary to extract and present the most important information and use presentation techniques that ease the information navigation. It is also important to offer an environment, which has serious character and forces user's imagination to improve the domain space in a user's personal way. Our idea is to visualize multimedia with attributes, where the focus lies on the most important attribute leading to the reduced complexity of domain space.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 265–268, 2010.

Fig. 1. Example visualization for UNICEF site – whole screen and detail

Following the p*reattentive visual patterns* [5] we created a pleasant interface that allows a user to comprehend the content quicker and more effortlessly. From five basic patterns to distinguish objects (colour, size, rotation, movement and shape) we have utilized the *colour* as an element to distinguish entities and the *movement* as a navigation element, which attracts users. The combination of the other three patterns creates various possibilities how the entity can be highlighted so the user is alerted on different actions in different way.

Considering *depth of presentation space* there are three possibilities that fit the requirements: the visualization made via *2D*, *2.5D* or *3D space*. Visualization of multimedia content implicitly requires a 2D surface or a 3D object with the multimedia displayed on its surface. The third dimension can be confusing but if it is properly used it can increase the clarity of visualization of a big amount of entities.

It is important to make the 3D environment look real and natural what ensures that the user feels comfortable. This can be realized by content emphasizes techniques using *perspective*, *kinetic depth*, *face-tracking* and *binocular vision*. Perspective is natural way how some objects can be highlighted without additional features. A large number of visualized objects require smaller objects or objects which are overlaid. The usage of intuitive object movement (kinetic depth) solves the problem of overlaid objects, because the user can easily see behind the objects. *Face*-tracking is natural usage of kinetic depth based on the user's head position in front of the screen. Giving good light conditions, it does not burden the user and so it is a good additional feature. Similar principle is binocular vision, which is the best way how to achieve clarity and simplicity of complex domain space. A disadvantage of this solution is the necessity to own quite expensive polarized glasses.

In our solution we have employed the clarity of 2D visualization and advantages of 3D. The data entities are in 3D space but multimedia content remains 2D. The media content is displayed on billboards that are perpendicular to the camera. To create an illusion of 3D space we made the back entities darker to look like dipped in shadow together with kinetic depth and natural perspective (front objects are about 3 times bigger than rear). The face-tracking is used as a bonus to add attractiveness.

1.1 Graph User Interaction

One plausible solution to information visualization and navigation in 3D is to distribute the information entities in the full volume of a 3D body and then let a user to change the position of the body, navigating through the data set.

We use the polar coordinate system to govern the 3D body movement, following the spherical layout explained later. We placed the center of the sphere at the origin and also positioned the camera at the ray in the certain distance from the origin. At the center of the sphere we placed an exchangeable data entity to highlight the point of rotation and consequently the entity itself. The user can look at the data set from any point of view and also from any distance along the ray. We opted to keep the rotation axes fixed relatively to the camera position and direction, to reduce the complexity.

In order to be able to interact with the specific visualized entities, the user needs to be able to pick the ones intended for the interaction. The position of the entities visualizes relations within the data, so it makes sense to select groups clustered together. To perform the selection, we allow the user to select a single entity and then until he does not let go of the mouse button, we gradually add the traversed nodes into the selected set, exploiting the spatial structure of the data space.

While it is possible to achieve a uniform initial distribution with fixed random layout of the objects in the 3D body, the layout itself is static and inflexible in later operations. Therefore we developed an interactive adaptation of force-based graph drawing algorithms. The iterative adjustment allows for the adaptation of node count change. Dynamic change of the force computing formulas lets the model to naturally regroup the nodes following a chosen similarity measure.

The scene of the model can also be augmented with so called virtual objects that help to take control of certain aspects of the movement of visualized entities. For example, we can insert an invisible and highly repulsive object in front of a node we want to highlight to repel nodes that block the view. The model creates an impression of living, playful "organism" behaving like a swarm of fish. The "viscosity" is controlled to accommodate the various users. We gather user input statistics as e.g. click rate, engagement in the game aspects of the presentation and feed them to a fuzzy controller that derives the speed settings.

We need to rotate the body around two perpendicular axes and change the middle node. To improve the mentioned effects, we embed the manipulations in the model itself. This is achieved by adding the virtual forces during certain iterations of the layout algorithm, allowing the whole system to compensate for the change of the middle node by shifting the nearby nodes and also creating naturally looking effects caused by inertia of the nodes apparent during the rotation, which shapes the normally spherically distributed data set into a disc-like object.

1.2 Applying Game Principles

We employ the game definition to abstract game-like features and use them in generic none-game information presentation in order to motivate users without the necessity to directly bring motivation through presented information. A game represents a formal system based on four elements: *goals*, *rules*, *challenges* and *interaction*.

Our solution employs short term goals via evaluating of effectiveness of interaction during a short period and via achieving ranks. Long term goal is applied in achieving a given level of entities number in presentation. The number of entities is dynamically changing according to the users' interaction. There is also necessary to define challenges which ensure that the way of achieving goals will be also important for user. We do so via scoring and budget manipulation, which is deeply related to tools usage. The rules of tools manipulation define how the goals can be reached and finally the environment interaction response based on our physics model allows us to attract user in a totally new and innovative way.

The game-like principles not only increase users' motivation, but also create many possibilities to implement wide variety of features (e.g. "welcomed-advertisement").

2 Evaluation and Conclusions

To evaluate the proposed concepts, we have developed so called WOWI (World of Web Information) framework. To demonstrate the versatility, we created WOWI presentations for UNICEF and our institution and performed several experiments aimed at evaluation of user experience. Microsoft WPF enabled us to render circa 100 entities in the web application.

Our contribution came from the unique combination of game principles catering on basic human instincts and the open presentation platform that is able to take a wide range of various types of data and transform it to a presentation appealing to broad audiences. Our method catches the attention of the user and sustains it, creates challenging and social environment that makes the user enthusiastic about learning and comprehending presented information and creates a wide range of possibilities for further augmentation of this space with business plans of "welcomed-advertisement".

Acknowledgement. This work was partially supported by VEGA 1/0508/09, KEGA 028-025STU-4/2010 and it is the partial result of the Research & Development Operational Programme for the project Support of Center of Excellence for Smart Technologies, Systems and Services, ITMS 26240120029, co-funded by the ERDF.

References

1. Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E.: Ontology visualization methods - a survey. ACM Comput. Surv. 39(4), 10 (2007)
2. Bai, X., White, D., Sundaram, D.: Visual intelligence density: definition, measurement, and implementation. In: Proc. of the 10th Int. Conf. NZ Chapter of the ACM SIG on HCI. CHINZ 2009, pp. 93–100. ACM, New York (2009)
3. Tvarožek, M., Bieliková, M.: Collaborative multi-paradigm exploratory search. In: Proc. of the Hypertext 2008 Workshop on Collaboration and Collective Intelligence, WebScience 2008, pp. 29–33. ACM, New York (2008)
4. Tsang, M., Fitzmaurice, G., Kurtenbach, G., Khan, A.: Game-like navigation and responsiveness in non-game applications. Commun. ACM 46(7), 56–61 (2003)
5. Deller, M., et al.: Preattentive visualization of information relevance. In: Proc. of the Int. Workshop on Human-Centered Multimedia, pp. 47–56. ACM, New York (2007)

Exploration in 3D Multimodal Virtual Environment for Nonvisual Spatial Recognition

Ying Ying Huang

HCI Dept. School of Computer Science and Communication,
Royal Institute of Technology
SE-10044, Stockholm, Sweden
yingying@csc.kth.se

Abstract. Visually impaired people have more problems on navigation both indoors and outdoors compairing with sighted people because of lacking the visual channel. Appropriate learning tools about spatial information could be used as a preparation for navigation before going to the real place. In this paper, a study on exploring access in non-visual spatial information by visually impaired people and the navigation and orientation with haptic and audio cues is presented. A 3D virtual simulation prototype of a real world environment was created for this purpose. Different navigation tools were designed in the prototype with haptic and audio cues. The main results from qualitative analysis of the study reported are: (a) the development of a virtual 3D environment prototype; (b) whether the spatial information for establishing the mental mapping of the space could be acquired by using touch and hearing channels ; (c) how haptic and audio cues can facilitate the navigation, mobility and orientation in such 3D virtual environments.

Keywords: Haptic, Audio, Virtual Reality, Navigation, Cognitive Mapping.

1 Introduction

Spatial information presented visually is not easily accessible for visually impaired users. Current access technologies, such as screen readers, can not convey spatial layout or structure. Virtual reality has been a popular paradigm in simulation based training, in the gaming and entertainment industries. It has also been used for rehabilitation and learning environments for people with sensory, physical, mental, and learning disabilities [1][2]. Particularly haptic technology offers alternative means of presenting spatial information to visually disabled users. Research on the implementation of haptic technologies within Virtual Environments (VEs) has reported on potential for supporting development of cognitive models of navigation and spatial knowledge with sighted people [3][4][5][6] and blind people [7][8]. However, users normally interact with the virtual world using a single "pen" probe or stylus. It is not very easy for them to interact with 3D models in the absence of visual cues because of the point-contact nature of

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 269–272, 2010.

the haptic device. There are two main problems as mentioned in the study by Walker et al. [10] One is that it can be hard to locate the data initially. The second is that it will be difficult to maintain contact with it after locating an object. As Sjostrom [5] reports: "For a blind person, locating an object with a point probe can be as hard as finding a needle in a haystack". Colwell et al. [9] describe users as becoming "lost in haptic space" when they are unfamiliar with this type of interaction, as they frequently lose contact with an object they are trying to explore.

A 3D haptic and audio virtual simulation environment of Kulturhuset (House of Culture) in Stockholm was developed for this study. Such touch-and-hearing tools could potentially be used by visually impaired users to learn about physical spaces that they are required to navigate (e.g., school, work place, public buildings). The prototype allows users to explore the virtual Kulturhuset building to find different locations. The Research questions of this study were: (a) the development of a virtual 3D environment helping visually impaired users to learn about real space where they are required to navigate (e.g. in schools, work places, public buildings), (b) whether the spatial information for establishing the mental mapping of the space could be acquired by using compensatory sensorial channels (e.g., touch and hearing), as an alternative to the visual channel; (c) how haptic and audio cues can enable the efficient navigation, mobility and orientation in such 3D virtual environments. Results from qualitative analysis regarding learning process and actual performance in the 3D virtual world are presented.

2 The Study

Five visually impaired participants took part in the study, one female 23 years old, males 28, 35, 40 and 52 years old. The participants tried the Kulturhuset prototype (Fig. 1, Fig. 2) individually. Fig. 3 shows the setting of the study environment. Participants should do two tasks (objects identification) under each condition(Haptic visual/ haptic and audio and visual) in a random order. We keep the visual channel available. In all conditions there are magnetic forces

Fig. 1. A virtual version of Kulturhuset

Fig. 2. Escalator

Fig. 3. User interacting with the haptic virtual environment using a PHANTOM Omni force feedback device (source: SensAble Technologies Inc)

designed for each object in order to attract the user when they move closer than 5 meters.

In the test, a Dell Precision Laptop with two dual core processors is used. One PHANTOM Omni, the haptic device, is connected serially to Laptop. For developing the application H3DAPI 2.0X3D, Python as well as Visual Studio .NET 2003, are employed. H3DAPI can be used both as an API to implement X3D applications, and as a loader of applications built in X3D and Python. A video camera is used for recording purpose and Camstudio for screen capturing during the tests. A video analysis was performed after the test sessions in this study. An interview was conducted for about 20 minutes after the test. Finally, the results with qualitative analysis were compared in order to obtain general findings.

3 Result and Discussion

This work presents an effort to explore the impacts of haptic and audio cues on non-visual spatial recognition for visually impaired users. Four issues were analyzed based the study, they are: the possibility to access a 3D virtual multimodal envrionment; Object Recognition; Spatial Recognition and the impacts of Haptic and Audio Aids on Navigation, Mobility and Orientation. The most important conclusions that can be drawn for the design of such environments are listed below:

· It is possible to access the virtual world of a real world by visually impaired users with haptic and audio cues. Both objects and spatial structure recognition will be faster, more accurate, and easier when both types of cues are available.

· Haptic cues alone will aid structure recognition, but audio cues alone will not. Once a node has been located, audio cues will provide more efficient means (speech or non-speech) of identifying it than haptic cues. Therefore, participants will rate haptic cues as more useful for identifying structure and audio cues as more useful for identifying objects.

· Integration of haptic and audio feedback provides more efficient navigation, mobility and orientation that either of them alone. It is better to have larger surfaces in all objects. The design of spatial architecture, fixed reference points like walls, floor, fixed objects and sensitive texture with haptic feedback are important to support orientation.

We hope that the study could be an investigation on the the possibility to access a 3D virtual multimodal envrionment for blind and visually impaired people to use haptic and audio virtual reality software for journey planning before they move into the real world. It could also serve as an input for visually impaired users to access online 3D virtual reality environments like Second Life with haptic and audio cues.

References

1. Schultheis, M.T., Rizzo, A.A.: The application of virtual reality technology for rehabilitation. Rehabilitation Psychology 46(3), 296–311 (2001)
2. Standen, P.J., Brown, D.J., Cromby, J.J.: The effective use of virtual environments in the education and rehabilitation of students with intellectual disabilities. British Journal of Education Technology 32(3), 289–299 (2001)
3. Hurmuzlu, Y., Ephanov, A., Stoianovici, D.: Effect of a Pneumatically Driven Haptic Interface on the Perceptional Capabilities of Human Operators. Teleoperators and Virtual Environments 7(3), 290–307 (1998)
4. Kuber, R., Yu, W., Mcallister, G.: Towards Developing Assistive Haptic Feedback for Visually Impaired Internet Users. In: Proceedings of the 2007 Conference on Human Factors in Computing Systems, San Jose, California, USA, pp. 1525–1534 (2007)
5. Sjostrom, C.: Designing Haptic Computer Interfaces for Blind People. In: Proc. ISSPA, vol. 1, pp. 68–71 (2001)
6. Bergamasco, M., Avizzano, C., Di Petri, G., Barbagli, F., Frisoli, A.: The museum of pure form: system architecture. In: Proceedings of 10th IEEE International Workshop on Robot and Human Interactive Communication, pp. 112–117. IEEE Press, Piscataway (2001)
7. Bergamasco, M., Prisco, G.: Design of an anthropomorphic haptic interface for the human arm. In: Shirai, Y., Hircse, S. (eds.) Robotic Research, the Eight International Symposiums, pp. 278–289. Springer, London (1998)
8. Frisoli, A., Simoncini, F., Bergamasco, M.: Mechanical design of a haptic interface for the hand. In: Proceedings of 2002 ASME DETC 27th Biennial Mechanisms and Robotics Conference, pp. 25–32. ASME Press, Montreal (2002)
9. Colwell, C., Petrie, H., Kornbrot, D.: Haptic virtual reality for blind computer users. Paper Presented at the Assets 1998 Conference, Los Angeles, CA (1998), http://phoenix.herts.ac.uk/sdru/pubs/VE/colwell.html
10. Walker, B.N., Lindsay, J.: Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. Human Factors 48(2), 265–278 (2006)

Experiments with Adaptable Interfaces for Elderly People*

Norbert Kuhn, Stefan Richter, Michael Schmidt, Andreas Truar, Markus Schwinn, Stefan Naumann, and Markus Dick

Institut für Softwaresysteme in Wirtschaft, Umwelt und Verwaltung
Fachhochschule Trier, Umwelt-Campus Birkenfeld,
Campusallee, 55761 Birkenfeld
{n.kuhn,s.richter,m.schmidt,a.truar,m.schwinn,s.naumann,m.dick}
@umwelt-campus.de

Abstract. The demographic change in European societies has led to raise attention for the needs of elderly people. Many of elderly people suffer from cognitive as well as from physical restrictions. This impacts their ability to participate and to communicate in their social environment. This could be substituted by constructing a virtual social environment based on services like videoconferences, chat rooms, e-mail and so on. However, elderly people often have difficulties to use IT technology because they encounter restrictions caused by their age. Thus, they need specific and adaptable human computer interfaces to be able to interact with IT systems. We present here some ideas how such interfaces can be constructed with reasonable effort and discuss some results from an evaluation of these ideas for an implementation of an interface to a personal information system for elderly and handicapped users.

Keywords: Adaptable HCI systems, Multi-modal User Interfaces, Accessibility, Ambient Assisted Living.

1 Introduction

Since many years a dramatic change in the age structure of the European nations may be observed. [1]. A further effect of the ageing society is the increasing percentage of people with disabilities in vision and hearing. [2].

In contrast to the generation of the "digital natives" elderly users often have rare experience e.g. in communication in social networks, or electronic communication at all. This means that classical criteria for human computer interfaces, like usability, clarity of the navigation structure etc are crucial for the acceptance of any application for elderly people.

Furthermore, we state that also techniques to ensure accessibility to IT systems are important to be considered. In [3][4] we presented some ideas to provide visual impaired users with access to printed documents.

---

* This work is supported by the German Ministry of Research and Technology under grant FKZ 1771X07.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 273–276, 2010.

In this paper however, we focus on the construction of the user interface. This should be able to embed an application into the experienced context of a particular user and allow him to use a wide rage of devices to interact with an application.

2 Constructing Dynamic Interfaces

2.1 Automatic Adaption of User Interfaces

Based on our evaluation with elderly people the process of automatic adaption of the user interface has to consider two main aspects. The first aspect concerns the presentation of the user interface, while the second aspect concerns the integration of different input devices. For the automatic adaption of the presentation we use a model-based approach. For the integration of different input devices we use the idea of abstract events.

2.2 Model-Based Design

In a model-based approach the user interface is usually described by a set of different models. For example, in [7], Calvary et. al. make use of four models, which are arranged in a hierarchical order. Each model describes the user interface on a different level of abstraction.

With this approach it is possible to adapt the user interface to different output devices, e.g. a PDA, a PC or a TV. In our context we want to specialize this approach to consider the input devices and the user's needs. Therefore, we introduce the concepts of a user model and a hardware model.

2.3 User/Hardware Model

The user model describes the requirements of a particular user concerning the user interface. For the group of elderly people we mentioned already the heterogeneity of requirements cased by a smooth but continuous degrade of their physical, perceptual, and cognitive abilities.

The hardware model contains information about the users hardware configuration. Like the user model the hardware model is different for every user. This is caused be the different input/output devices that are used by the intended user group.

For the realization of an automatically adaption of the user interface a model-based approach which considers the user model and the hardware model should be used.

2.4 Enhanced User Model for Seniors

To realize the dynamic approach, a model of the user interface is needed. Our framework extend the work of Calvary et. al.[8]

At the top of this hierarchy there is the Interaction Model. It describes the interactions between the user and the user interface in an abstract way. In addition to the reference framework of Calvary et. al. the Interaction Model is expanded with a user model. This model describes all the abilities and needs for a concrete user.

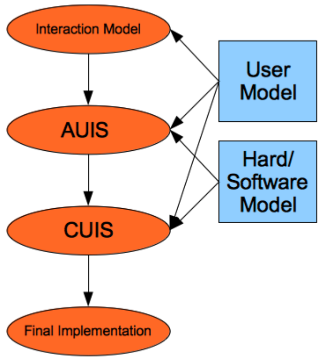


Fig. 1. Reference framework for the model based approach

The next level of hierarchy is the AUIS (Abstract User Interface for seniors). In this level the interactions of the Interaction Model are specified using abstract interface objects. The model consists of basic interactions that occur inside a user interface. As shown in Fig. 1 the user model and the hardware model are used while building the AUIS.

In the next stage of the framework there is the CUIS (Concrete User Interface for seniors). At this level the abstract user interface objects are mapped to concrete user interface objects. Like the two models before the CUIS also includes the user model and the hardware model. This is important because a specific user may change the mapping between the AUIS and the CUIS.

The last stage addresses the final implementation. Here, the CUIS is transformed to a final implementation, which runs on a specific hardware configuration.

In conjunction with the user model and the hardware model the framework will be able to automatically adapt the views of a user interface.

2.5 Adaption to Different Input Devices Using Abstract Events

Most applications are designed to use a mouse and/or a keyboard. In many cases it is very difficult or even impossible to use alternative input devices, such as a graphic tablet, a microphone or a remote control for basic functions. To overcome this limitation we introduce the idea of an abstract event. The abstract events are the same for every new device presented to the system. In this way the user interface has to support only these abstract events. To realize this, the special events of an input device are mapped to abstract events.

To implement this mapping we may use the Observer Design Pattern as described in the GOF Patterns ([5][6]).

This software architecture allows the user to use all input devices simultaneously. Furthermore, a new device can be easily integrated through a concrete implementation of the specified events.

The framework described here makes it possible to build the user interface at runtime. This ensures that the user interface can always adapt to the user's needs.

3 Conclusion and Future Work

The feedback of elderly or handicapped users to our demonstrators has underlined the need of adaptable interfaces. The approach we discussed here uses a model-based design and considers the user's needs, and the type of device he uses. It allows the implementation of an adaptable, dynamically generated user interface with special focus on elderly people.

To proof this concept we will extend our demonstrators to an information serving and document archiving application.

References

[1] U.S. Census Bureau: International Database, http://www.census.gov/ipc/www/idb/pyramids.html

[2] Schweizerische Eidgenossenschaft, Bundesamt für Statistik: State of Health – Severe disabilities in the population older than 15 years, state (2007) (in German), http://www.bfs.admin.ch/bfs/portal/de/index/themen/14/02/01/key/04.html

[3] Kuhn, N., Richter, S., Schmidt, M., Truar, A.: Improving Accessibility to Governmental Forms. In: IFIP 20th World Computer Congress, Proceedings of the 1st TC 13 Human-Computer Interaction Symposium (HCIS 2008), Milano, Italy, pp. 119–128 (2008), ISBN: 978-0-387-09677-3

[4] Richter, S., Kuhn, N., Naumann, S., Schmidt, M., Truar, A.: Enhancing Accessibility to E-Government Processes. International Journal of Informations Communication Technologies and Human Development, Special Issue: ICT and E-Governance 1(2), 28–47 (2009)

[5] Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns Elements of Reusable Object-Oriented Software, pp. 293–303. Addison-Wesley, Reading (2007) (35th Printing)

[6] Kuchana, P.: Software Architecture Design Patterns in Java, pp. 343–354. CRC Press LLC, Boca Raton (2004)

[7] Collignon, B., Vanderdonckt, J., Calvary, G.: Model-Driven Engineering of Multi-Target Plastic User Interfaces. In: Fourth International Conference on Autonomic and Autonomous Systems, ICAS (2008)

[8] Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A unifying reference framework for multi-target user interfaces. Interacting with Computers 15, 289–308 (June)

The Practice of Interaction Design

Ann Lantz

HCI-group,
The Royal Institute of Technology
S-100 44 Stockholm, Sweden
alz@kth.se

Abstract. The aim with this paper is to look into the role of interaction design and we investigate how interaction designers experience their work role, and, activities they perform. Semi-structured interviews were conducted with interaction designers. They do not have an overall view of the design process and are hired as internal consultants. The role of the interaction designer is included in the formal reports on how to conduct a development project. Often the interaction designer is used to work with a limited piece of the design process and the methods at hand. It is important to separate the role of the interaction designer from an old development tradition and move it to a more design-oriented way of thinking about interaction design and the work of interaction designers.

Keywords: Design process, role of interaction design, communication.

1 Introduction

What do we know about the practice of interaction designers and their understanding of the design processes they are a part of? Several publications focus on, for example, architects' practice [5] and theorists describe interaction designers' work roles and design processes [3;4] the view they present of the interaction designer is often idealized. But not much has been written on the practice of interaction design.

"We need more research that examines, uncovers, analyses, and interprets what interaction designers are already doing" [5, p.62]. Today researchers are describing and suggesting how the practice should be performed, in theory but sometimes without grounding it in practice. The gap between this and the actual performed practice, the existing practice, is called "rationality resonance".

Descriptions of interaction design often do not include the context. That is, they provide a generalized picture that is not connected to a specific case or situation and do not describe the kind of business of which the interaction design work is a part. According to [1;3] it is important to be aware that the context for the design work influences the work to be performed. The design process will differ depending on whether it is carried out by a group in a large product development organization, an in-house development group, or a design consultancy on contract to a client. Interaction design is a design area that is not necessarily based on design theory and methods for interaction design. HCI researchers can help construct these theories and methods that support the practice of interaction design:

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 277–280, 2010.

> *"In design you have to design the "whole"… In design methods and approaches you have to take the whole composition into account which creates distinct methodological requirements when it comes to testing and evaluating"* [6, p.7].

Design can be viewed as communication and common ground where collaborating with users are central. The perspective an interaction designer chose to take will have consequences for her role and for ways of handling users, responsibility, and a reduced or holistic view on design [7]. Design can be viewed as a process of unfolding the setting/problem solving pair. The role of the designer is also described as creating compositions on the basis of parts and smaller entities but also composing knowledge and information about details in a way that supports a holistic way of thinking [3].

In [2] interaction design practitioners describe what interaction design means to them; some represent a traditional HCI perspective where a designer works step by step, often with a single work input at one point in the system development process at a time. Others describe working as project leaders, having an overview of the design process and being able to influence it. For those in the latter category the focus is more on creating design than on evaluating a design made by others.

The aim with this paper is to look into the role of interaction designers in order to strengthen the role as well as developing the education of interaction designers. The following research questions are posed:

- Do interaction designers play an overall role in development projects?
- Do interaction designers have an overall view of the design process?
- What are the specific activities in which interaction designers take part?

2 Method

Semi-structured interviews were conducted with six interaction designers employed by the development organization of a Swedish government agency. The respondents were all educated in cognitive science except for one whom had a background in systems science. They had all worked at the agency for five to ten years except for one who had been recruited recently.

Each interview took about two hours, were tape-recorded, followed an interview guide whose main question areas were IT projects process, collaboration and procurement, communication and visualization, and interaction design. The interviews were transcribed verbatim, the data was sorted following the question areas of the interview guide and summarized for each question area in the interview guide. The respondents' names used below are all pseudonyms.

3 Results

In the Swedish HCI community, this particular government agency is generally known for its maturity in user-centred design (UCD). It has worked with HCI researchers over many years to develop its internal way of working with systems development. The agency uses the development method called Rational Unified Process; it has modified it for internal use. Here the role of usability designer is mentioned and described as one of the roles that must be included in all projects.

The group of usability designers at the agency works on projects that are run internally in the organization. The usability designers group functions as a pool of resources in the usability area and are assigned to projects on an hourly basis, like internal consultants. Often they perform small missions and it is rare for two of them to work on the same project.

The interaction designers are uncertain about where a procurement is coming from at first; they say that it comes from a high level in the organization, from procurers who work at headquarters. Some said that an IT project is constructed based on either a change in the Swedish law, or a development or change in the business because an old system needs to be improved. But projects can also be initiated within other groups in the organization. One respondent said:

"It should be that it is the procurer who has a need. Sometimes, it starts at the IT department and that is not the way it should be" (Anna).

A method is used to plan and perform projects and the effect goals and product goals that are supposed to be evaluate after the project ends is here brought forward. It should be possible to trace requirements in the project and to conduct a follow up, but several of the respondents said this is not done. The responsibility for making a system usable is said to be that of the procurers; it has to be procured by the business representatives but produced by the designers who have the knowledge about methods.

One of the respondents gave an example of the activities the designers perform:

"We work with conceptual design and evaluations, going through prototypes with user representatives early in the project, using personas; this is where you can get the requirement into the design. It is very flexible and later it is very much function" (Erik).

The interaction designers are performing one or more of the activities described below; rarely are they able to follow a project over time. The interaction designers visit users at their work site and conduct interviews, and observe their work and make user analyses. They also construct personas, and use cases, scenarios, and interfaces, and at the worksites they test prototypes with users. They make descriptions of GUIs for use cases and are responsible for ensuring that a prototype is constructed but they normally do not construct it themselves. They do work on help functions and manuals, and evaluate usability and the system's interactivity: what it will look like and how it will work.

The respondents say they work with interaction, with the surface, that is the GUI, and also with the context. On some longer projects an interaction designer may be assigned to spend as much as 80% of her work time on the project; then she will follow its process and engage in activities to ensure its usability. But this is fairly rare; more often, these assignments are quite brief. When designers were asked how the results from a project are followed up, they said it is not clear to them or that they think it is not always done. It is also clear that most often the target group for usability aspects is the employees working in the organization and not the citizens who will be the end-users of the e-service.

4 Discussion

The interaction designers do not have an overall view of the design process, or at least they cannot describe the process or the source of the procurement. It is encouraging that the role is included but that is far from enough. The interaction designers have the competence and knowledge but the system also has to be procured. There is a risk that this competence will not be used well if the procurers and project leaders are not aware of what to expect from an interaction designer. Lack of awareness can lead to designers getting caught in an engineering way of thinking of design. The designers themselves could benefit from seeing more of the overall picture since they are expected to perform interaction design and work with usability. It is important to separate the role of the designer from this old development tradition and move it to a more design-oriented way of thinking about interaction design.

In [5] the concept of *"rationality resonance"* [p.62] is promoted: researchers tend to describe how design should be performed rather than how it actually is performed. In this case study we can see that some descriptions about how design should be performed do not regard design as design but instead as engineering. As researchers, if we study how interaction design is performed and understood, we can instead support and influence the development of interaction design and the role of the designer and one way of doing this is to work actively with the education of interaction designers.

We suggest ways to facilitate a more holistic view on interaction design: a, allow the interaction designer to work more closely with the procurers during the requirements and problem defining phase, b, continue to make evaluations after the initiation and maintenance stages.

References

1. Erickson, T.: Notes on Design Practice: Stories and Prototypes as Catalysts for Communication (1995), `http://www.pliant.org/personal/Tom_Erickson/Stories.html` (last visited 2010-05-21)
2. Lantz, A., Artman, H., Ramberg, R.: Interaction design as experienced by practitioners. In: Proceedings of Nordic Design Research Conference, Copenhagen Denmark, May 29-31 (2005), `http://www.nordes.org/`
3. Löwgren, J., Stolterman, E.: Design av interaktionsteknik – materialet utan egenskaper. Studentlitteratur, Lund (1998) (in Swedish)
4. Nelson, H.G., Stolterman, E.: The design way. Educational Technology Publications, CITY, NJ (2003)
5. Stolterman, E.: The nature of design practice and implications for interaction design research. International Journal of Design 2(1), 55–65 (2008)
6. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI. In: Proceedings of CHI 2007, San Jose, CA, USA, April 28-May 3, pp. 493–502. ACM Press, New York (2007)
7. Gulliksen, J., Lantz, A.: Design vs. design. In: Lantz, A., Gulliksen, J. (eds.) Design versus Design: A Nordic perspective, Int. Journal of Human Computer Interaction, pp. 5–20 (2003)

Artefacts as a Cultural and Collaborative Probe in Interaction Design

Arminda Lopes

Instituto Politécnico de Castelo Branco, Portugal
aglopes@ipcb.pt

Abstract. This paper presents a summary analysis of observed case studies within two design network groups. Designers with different backgrounds created some artefacts which permitted to argument that design can be understood as a dialogue approach. The interaction among designers mediated by the artefact allowed to find cultural markers and a collaborative circle especially significant within the design process. A social methodological perspective followed by semioticians was undertaken for the artefacts analysis.

Keywords: Interaction Design, Culture, Collaboration, Artefacts/Arteworks.

1 Introduction

Interaction design has been associated with computing and technology, but the focus of interaction design must also be on designing the way people interact with any artefact, be it an object, a system, an environment, or being a consequence of the use of digital technologies or not. The aim is to support an interaction among designers that is mediated by the artefact. Much of what is understood about the design of digital artefacts is also applicable to non-digital artefacts. So, it seems that there is no reason for the discrimination between them.

This paper presents a synthesis of the analysis carried out within a research about designers with different backgrounds producing artefacts.

Designers create culture, as they create experiences and meaning for themselves and others. These experiences were reflected in the presented artefacts in the case studies and their analysis and interpretation was given by the objects meaning or what the objects meant in the different things to different people.

The collaborative circle consisted of designers who shared similar goals and who, through long periods of dialogues and collaboration, negotiated a common vision that guided the work. The vision, in general, consisted of a shared set of assumptions about the artefact/artwork, including what constituted good work, how to work, what subjects were worth working on, and how to think about them. The engagement with design was entertaining and expanded the notion of the very nature of participation, of taking part in and of itself.

Artefacts/artworks are expressions with different signs on different levels of design language. Just as semiotics, the study of signs, helps us to understand practices of description, which in turn reveal how meaning is communicated, so does it enable us

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 281–284, 2010.

to understand the creativity of the designer, design being a method of the communication of meaning. The conveyance of those meanings and the designer's input, but to consider what the designers actually said and how they behaved during the design process, and to describe the outcome they produced was regarded as of high import in this research.

2 The Study

The data was collected within two design network groups: the Leonardo Network group and the White Rose Network for Affective Communication in Consumer Product and Exhibition Design. There were about twenty institutions in the former and four in the latter; and about 55 artists and technologists that took part in the study. The methods used to collect data for this research were centred on a qualitative study, a combination of research methods was used to collect the data, literature review, including documentation, records of individual and group's experiences and behaviours, case studies, interviews and observation.

3 Artefacts Analysis

The approach followed, in the analysis, was a social semiotic one as well as a multimodal social semiotics one. From a methodological standpoint, social semioticians analyse images according to three main metafunctions. Halliday recognise three main kinds of semiotic analyses that are always performed simultaneously [1]. This idea was also extended to images, using a somewhat different terminology: representational, interactive and compositional [2]. The approach allows deconstructing texts into these three main types of meaning.

Figures or texts were analysed within this perspective and also describing a 'thing' and its qualities and the 'context of the things'. The qualities referred in this context are related to the object properties such as colour, shape, weight and size.

The analysis to follow is an exploratory process involving a visual analysis of a dialogue through a series of drawings that were made in a dialogue context. Far from being a subjective experience, the researcher considers to be a profound dialogical achievement. During the visual analysis of the designers' interaction, the form of dialogue observed was: vocal and sketching in that they engaged in dialogue via both words and images.

The artefacts developed within some case studies were drawings and words, from this point, all called texts [2]. ïThe artefacts analysis could be carried out in different deepening degrees. We due to the large amount of data selected some examples and did a 'snapshot analysis'. However the goal was attained because it could be argued that each artefact reflected the dialogue that occurred during the design process.

3.1 Artefacts Description and Interpretation

In this study, two types of interaction were understood: interaction of designers with the artefact and that among designers mediated by the artefact, this paper refers to the latter one. Table 1 presents the main features considered for the analysis.

Table 1. Meaning Analysis

Representational Meaning	Compositional Meaning	Interactive Meaning
Syntax (qualities of the artwork - lines, shapes, colours, textures) and **Materials**	**Semantic** (forms, purpose, and meaning)	**Pragmatic** (relationships).

In each group composition a diversity of designers' background was observed, this being the case study that contained the highest variety of participants from different disciplines. This had a reflection on the designers' behaviour: they were very engaged in the design process, however they were more critical. Some of them considered that they had gone for systems solutions rather than personalisation and they concluded that they were approaching products as systems in affective design in some way. Also, concerning culture, this group was formed by designers from a variety of national cultures which influenced in the way they talked and in the considerable amount of ideas each team presented until a consensus was achieved defining the one to follow. Referring to creativity, the challenge was conducive to creativity in the way that the designers found solutions.

All the drawings were executed within a limited time and also using a limited range of media in order to make their productions as simple as spontaneous as possible. They were executed using paper and pens, with only black, blue and red ink. Some "drawings" were cut out magazine pictures, some sketches were accompanied with words.

In the Human Beans challenge, teams were assigned scenarios and asked to identify the top three emotions to design environments that would mitigate these emotions. From another case study groups were challenged to identify interactive technologies that in some way could respond to a problem set by another team. This took a form of relating to a social issue, a context or setting and a technology in order to produce a bid. In the Chindogu case study, the irony of the design solution arose from the dialogue established between designers during the design process. The proposal was to develop an artefact that could be a remote wild animal interaction device but following the chindogu tenets and using the available materials. Throughout the design process stages, communicative practice, collaboration, the demonstration of different levels of creativity and the expression of the designers' attitudes, interests, beliefs (culture) could all be seen to have had an effect on the conception of this humorous solution.

All the artefacts described and analysed had a common denominator: social interaction. Under this encompassing umbrella, in the artworks from the Human Beans workshops, subjects included were: online networks to facilitate social cohesion: exploration of how technology could facilitate more appropriate communication between people and their loved ones; the use of technology to preserve cultural memory; comfort associated with concepts of production; affective communication in design.

The resultant artworks of the Chindogu challenge must be understood within the context of uselessness. However, in the three resulting outputs, interaction was the overall concept and the preoccupation with communication between humans/humans or humans/animals were the underlying aspect.

Both conversations and artefacts/artwork practice were understood in this research as forms of dialogue intimately linked through meaning and, as meaning and interpretation are inherent in the relationship between the verbal and the visual, the practical and theoretical elements involved in the investigative process of the research were mutually dependent.

4 Conclusions

Culture was an important component of the design process. During the design of the artefacts/artworks, the opportunity to integrate cultural and geographic qualities into designed objects was provided. There was a cultural context of design and it was this that gave meaning, and also provided the values reflected in the objects' form and function.

The influence of the designers' own cultural dimensions of values in the design of objects was noted as being very important during the observed design processes. The design of artefacts and artworks was addressed as a response to a need, a desire or a challenge. This response in turn was influenced by the interplay of several factors including designers, the context and other participants in the design process. Also diverse cultural values, such as where the designer was born and has lived, influenced the way s/he designed an object during the design process. Designers' relationships were frequently made within an atmosphere of collaboration which means that those involved work together, shared an interest in the design goals and chosen problems, either to find a common solution or to exchange experiences aiming at a better understanding of the situations. Design as dialogue involved experiences: experiences of designers, experiences of the sharing, and experiences during dialogue.

The semiotic analysis of the artworks, which contained many messages – some more obvious than others, was complex but very interesting, beneficial and also surprising. It was clear that artworks are much more than designed objects - they are structures of meaning and "*an intersecting point in a network of relations*" [3].

References

1. Halliday, M.A.K.: Language as Social Semiotics. Edward Arnold, London (1978)
2. Kress, G., van Leewwen, T.: Reading images: A grammar of visual design. Routledge, London (1996)
3. Hjelmslev, L.: Prolegomena to a Theory of Language. In: 1953 trans. by Francis J. Whitfield, University of Wisconsin Press, Madison (1943), Rev. ed. (1961)

An Easy to Use Augmented Reality Authoring Tool for Use in Examination Purpose

Ming-Jen Wang, Chien-Hao Tseng, and Cherng-Yeu Shen

No. 7, R&D Rd. VI Hsinchu Science Park, Hsinchu, Taiwan (R.O.C.)
{renny,c00how00,c00cys00}@nchc.org.tw

Abstract. This paper introduces an augmented reality authoring tool that allows users to edit and publish an examination application with the Augmented Reality (AR) interface. The AR authoring system in this paper is constructed with the help of ARToolKit, a widely used open source for AR. This unique system can help users with no programming knowledge to build AR applications quickly and efficiently. 2D text, 3D content editing, and finger-based interaction are the three major components found in the AR authoring environment. AR content created by the user is stored in a separate txt file for easy sharing, modification, and reusability. A user testing of the published examination application was also conducted, and participants in the testing offered generally positive feedback.

Keywords: Augmented Reality, education, Human Computer Interaction.

1 Background

In order to introduce AR technology to the public and to content developers, many AR development tools have been proposed. These tools can generally be divided into two categories.

The first major category is AR-related library or software framework. Since the development of AR applications is technologically complex, using such systems as computer vision, computer graphics, image recognition, and so on, developing an AR application at a low-level programming environment often consumes considerable amounts of time. Therefore, in order to shorten AR development time, many AR related libraries and software frameworks have been proposed. Some of the most well known libraries and frameworks are ARToolKit [1], osgART [2], DWARF [3], MRT(Mixed Reality Toolkit) [4], ARTag [5], etc. Among these libraries, perhaps the most widely used is ARToolKit, a C/C++ based open source tracking library. The ARToolKit library is so popular that it has even been transferred into other computing languages, such as NyARToolKit (Java) and FLARToolKit (Flash ActionScript). Other than tracking issue, some libraries, such as OSGART, focuses on rendering issues related to AR. However, despite the many efforts expensed in the development of AR tools, most if not all of these tools are still only accessible and practically useful to professional programmers. The remainder of the universe of potential AR users are still waiting for tools which will provide them with a competitive amount of AR developing power.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 285–288, 2010.

The second category is the GUI-based AR authoring tool. As opposed to text-command based interfaces, GUI's provide more intuitive interaction for users. Before the GUI-based AR tool, a finished AR application might require programming and then de-bugging thousands of lines of code. Even with AR related open source libraries, developers still needed to invest significant time to learn how to use them. By introducing the GUI to the tool, developers now have a tool allowing them to complete AR applications on a point-and-click basis. AMIRE [6] is one of the GUI-based AR authoring tools. With it, users create AR applications without having to write any code. DART [7] is a plug-in for Adobe Director software which allows one to author AR content as well. ComposAR [8] provides another GUI-based tool for authoring AR content. Outside of the GUI solution, ComposAR provides an additional scripting interface for advanced users with computer programming experience who desire more control over the application design process. ATOMIC Authoring Tool [9] is among the examples of AR authoring interfaces made extremely simple.

We notice that most of the above mentioned tools only focus on building the relationship between AR Markers and 3D models. When the above tools are used to build complex AR applications, they either fail entirely or require further scripting to achieve the goal. For a non-programmer, AR application development is therefore highly restricted if only these tools are available. In this paper we propose a GUI based AR authoring tool that allows users to build AR applications with examination mechanisms and 3D model display functions free of any further coding. The user can answer pertinent questions at decision nodes by pointing to Yes or No Markers.

2 System Design

The system design has four key features:

First, we embed a Yes/No question examination mechanism in the AR authoring tool. From a pedagogical standpoint, examination is an important method in understanding the efficacy of a system of learning. However the examination process is often dry and not very attractive to students, especially younger ones. Also, a question might not be only text-based. It might be a graphical question that contains a 3D virtual model. Using a Marker-based AR application, users can interact with 3D models in a way that is very similar to how one interacts with real objects. Hence we hybridize the AR interface and examination to best take advantage of the most desirable traits of each system. To do so, a question and answer input interface is needed. In the authoring interface, a user can type in the question and choose the correct answer to the question in the question input section. If the question requires 3D virtual modeling in order to gain additional information, a user can also select a 3D model that has been previously made from the local computer to assist. If the 3D model's settings are not appropriate, the system also provides some basic functions to adjust the 3D model's size, translation and rotation parameters. Testing revealed that developing a 20 Yes/No question-based AR examination application using the system frequently required no more than ten minutes provided any 3D models and questions have been prepared beforehand.

Second, the reusability of the question data is considered. The system stores the question data in an external text file. All the questions and their answers, 3D models' saving paths, and Marker information are saved with predefined rules in a single txt file. There are two important purposes in doing so. First, saving the question data in an external file allows authors to exchange question data more easily via import/export protocols. This feature reduces the development time one step further when question data is readily obtainable. Second, the generation of questions is basically relying on the text file. Authors can readily understand most of the rules of the text file once they have examined it using the authoring system. This means the author can modify the questions in the text file directly without using the authoring system. This feature gives authors an easy and fast way to create or modify an AR examination application.

Third, the system is divided into two different modes, the authoring mode and the viewing mode, to provide the author and students the various functions they need. The system structure is shown in Fig. 1.

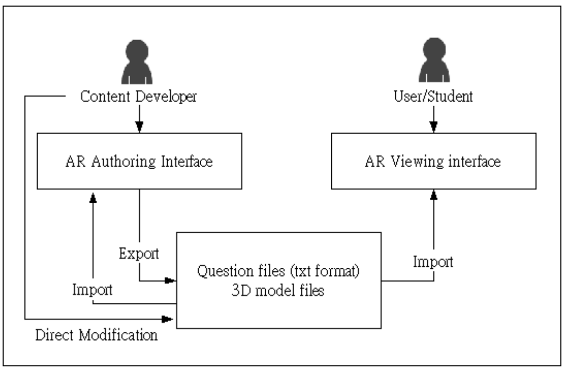


Fig. 1. The Authoring tool system structure of authoring and viewing modes

Fourth, we utilize a complete graphical user interface in the system. Target users of the system include teachers and students in elementary/high/senior high schools. Because the target users' knowledge of computer programming is assumed to be limited, a GUI-based interface will be more suitable for them. With our system, a few clicks and the input of some questions suffice for the author wishing to rapidly complete an application.

3 Prior User Testing and Discussion

We conducted a user testing to gather feedback and responses after using this system. The participants are all elementary students from the central region of Taiwan. There are a total of 24 students, all aged around 8 years old, who participated the experiment. To begin, we setup an AR examination application which consists of 10 questions using the authoring program. In the prior testing, we discovered that they weren't sure how long to "touch" the Yes/No marker to verify the choices of their answers. After being instructed with the time needed to press down on the marker in order to make the selection, all the students in this experiment were able to practice the interaction very quickly. Some of the students even finished all the 10 questions within 2 minutes. To conclude, 21 out of 24 students consider that the system is interesting and convenient to use and are willing to use the system as their future learning tool.

In this paper, we propose an AR developing tool that could assist content developers to create an interactive examination system with AR interface within a short amount of time. Comparing with the other AR authoring tools, the system is targeting at a specific purpose which is examination/education. For those tools that only support simple 3D models and Marker information linkage, the system we have developed could offer an easier and faster way for the content developers to create an examination with AR application. We also propose two different models for various users, content developers and as well as students. The data of the questions is also saved where it could be easily shared, exchanged or modified later on. In this user testing, we discover that most of the students offer positive feedback on the AR application built with the authoring tool. Nevertheless, the system still has room for further improvements. For one, the system could only support the Yes/No question-based examinations. If an author wants to develop a more complex examination or non-examination project using the AR application, it cannot be done by using this system and that is one of the limitations of this system. In the future, we are going to modularize the system functions. With more AR interaction modules being developed, the program users would be given more freedom and choices in creating examinations with a simple authoring process.

References

1. Kato, H., Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (1999)
2. Looser, J., Grasset, R., Seichter, H., Billinghurst, M.: OSGART – a pragmatic approach to MR. In: Industrial Workshop at ISMAR (2006)
3. Bauer, M., Bruegge, B., Klinker, G., MacWilliams, A., Reicher, T., Riss, S., Sandor, C., Wagner, M.: Design of a component-basedaugmented reality framework. In: Proceedings of the International Symposium on Augmented Reality (2001)
4. Freeman, R.: Mixed Reality Toolkit. MSc VIVE Final Year Project Report, Department of Computer Science, University College London (2004)
5. Fiala, M.: ARTag Revision 1, A Fiducial Marker System Using Digital Techniques NRC Technical Report (NRC 47419), National Research Council of Canada (2004)
6. Grimm, P., Haller, M., Paelke, V., Reinhold, S., Reimann, C., Zauner, J.: AMIRE - Authoring Mixed Reality. In: First IEEE International Augmented Reality Toolkit Workshop, Darmstadt,Germany (2002)
7. MacIntyre, B., Gandy, M., Dow, S., Bolter, J.D.: DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In: Conference on User Interface Software and Technology (UIST 2004), Sante Fe, New Mexico (2004)
8. Seichter, H., Looser, J., Billinghurst, M.: ComposAR: An Intuitive Tool for Authoring AR Applications. In: Mixed and Augmented Reality, 7th IEEE/ACM International Symposium, Cambridge, UK (2008)
9. ATOMIC Authoring Tool, `http://www.sologicolibre.org/projects/atomic/en/`

From Microsoft Word 2003 to Microsoft Word 2007: Design Heuristics, Design Flaws and Lessons Learnt

Yin-Leng Theng, Eng Kiat Ting, and Xuehong Tao

Wee Kim Wee School of Communication and Information
Nanyang Technological University, Singappore 637718

Abstract. By tracing and comparing the number of key-presses taken to complete the functions in Word 2007 and Word 2003, the paper identifies eight design heuristics Word 2003 and Word 2007 could have violated, creating design flaws leading to potential usability problems. This paper concludes with design implications for software.

Keywords: Microsoft Office Word 2003, Word 2007, design heuristics.

1 Introduction

Designers are often faced with the huge challenge of designing software and websites for a large number of different users, and do so by making the functionality of the software and websites as extensive as possible. In this one-size-fits-all approach, common goals of different types of users are identified. When introducing new interfaces or upgrades of existing software and websites, initial user confusion and poor adaptation are common user problems identified. The recent migration from Microsoft Office Word 2003 to Microsoft Office Word 2007 in late 2006 [3] is one such example that has generated poor reviews. This was evident from the numerous articles and blogs which have sprouted in the internet discussing this migration [4,6].

In this paper, we carried out an investigation of the learning problems faced by novice and expert users of Word 2003 when they migrating to Word 2007. We employed a heuristic evaluation of Microsoft Office Word 2003 and Office Word 2007 carried out to identify new or existing functions in Word 2007 that might cause usability problems based on established design heuristics [5].

2 Study and Findings

In our study, the *Word Ribbon Mapping Workbook* [2] was used. This workbook, aimed to facilitate users with the transition, listed the *Word 2003* commands along with their corresponding *Word 2007* commands (if any). The changes in locations/forms of these commands were then noted. The few hundred commands listed in the workbook were analysed one at a time. The changes in locations of these commands were noted. Those changes of locations of commands which could cause

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 289–292, 2010.

confusion for Word 2003 users were selected. Using the *Word Ribbon Mapping Workbook* [3] provided by Microsoft, we went through the 338 commands (menus and *Standard toolbar*) listed. Twenty-seven of such commands that could cause potential usability problems for *Word 2003* users were selected and classified into eight types of potential usability problems as shown in Table 1. Table 2 summaries each of the eight violations.

Due to space constraints, we show one example to illustrate the principle of each group. Table 2 (row1) shows the steps taken to perform function of creating a new *word* document in *Word 2003* (column 2) and *Word 2007* (column 3). The main source of confusion is that the sequence of clicks for *Word 2003* was located under the *File menu* whereas in *Word 2007*, it was located under the *Office Button* (see Figure 1). As such, *Word 2003* users could experience trouble finding the command to create a new document in *Word 2007*. Type 1 potential usability problem is also about changing associated icon (double jeopardy) for locating file function, and hence creating "cannot-find-the-command" user experience.

Table 1. Eight Types of Violations against Design Principles

THE EIGHT PRINCIPLES				
Principle	**Principle Description**	**Example Ref**	**Microsoft Word 2003 Command Paths**	**Microsoft Word 2007 Command Paths**
1	Same command but located under different top menu.	1a	[File > New > Blank document]	[Office button > New > Blank document > Create button]
		1a	[File > New > Close button]	[Office button > New > Blank document > Cancel button]
		1b	[Tools > Macro > Macros]	[View > Macros > Macros > View Macros]
		8a	[Format > Columns > OK button]	[Page Layout > Page Setup > Columns > More Columns > OK button]
		8a	[Format > Columns > Cancel button]	[Page Layout > Page Setup > Columns > More Columns > Cancel button]
		10a	[Tools > Spelling and Grammar]	[Review > Proofing > Spelling & Grammar]
		10b	[Edit > Paste Special > Unformatted Text > OK button]	[Home > Clipboard > Paste > Paste Special > Unformatted Text > OK button]
		10b	[Edit > Paste Special > Unformatted Text > Cancel button]	[Home > Clipboard > Paste > Paste Special > Unformatted Text > Cancel button]
		12a	[Tools > AutoCorrect Options]	[Office button > Word Options > Proofing > AutoCorrect Options]
		12b	[Tools > Customize]	[Office Button > Word Options > Customize]
		14	[Window > Split]	[View > Window > Split]
2	Change of visual search direction.	2a	Vertical search direction: [Insert > Symbol]	Horizontal search direction: [Insert > Symbols > Symbol > More Symbols]
		2b	Vertical search direction: [Tools > Word Count]	Horizontal search direction: [Review > Proofing > Word Count]
3	Change of top menu's position.	3a	[View] at 3rd position	[View] at 7th position
		3b	[Insert] at 4th position	[Insert] at 2nd position
		15	Help as a small text menu next to the main menu group.	Help as a tiny icon on the far right, away from the main menu group.
4	Cannot locate toolbars in Word 2007 and so user has to learn the corresponding commands.	4a	User depended on [Standard > Print] toolbar	Instead use: [Office button > Print > Quick Print]
		4b	User depended on [Standard > Zoom] toolbar	Instead use: [Status Bar > Zoom Slider]
		7a	User depended on [Drawing > Autoshapes] toolbar	Instead use: [Insert > Illustrations > Shapes]
		7b	User depended on [Standard > Insert Table] toolbar	Instead use: [Insert > Tables > Table > Insert Table]
5	Same top menu, same terminal menu, but different paths.	5a	[Insert > Date & Time > OK button]	[Insert > Text > Date & Time > OK button]
		5a	[Insert > Date & Time > Cancel button]	[Insert > Text > Date & Time > Cancel button]
		5b	[Insert > Field > OK button]	[Insert > Text > Quick Parts > Field > OK button]
		5b	[Insert > Field > Cancel button]	[Insert > Text > Quick Parts > Field > Cancel button]
6	Multiple ways for the same task in Word 2007.	6a	[Insert > Text Box]	First Way: [Insert > Text > Text Box > Draw Text Box]
		6b		Second Way: [Text Box Tools > Format > Text > Draw Text Box]
		13a	[Table > Draw Table]	[Home > Paragraph > Borders > Draw Table]
		13b		[Insert > Tables > Table > Draw Table]
7	Some commands need to be set up before use.	9a	[Format > Frames > New Frames Page]	[Office button > Word Options > Customize > All Commands > New Frames Page]
		9b	[Edit > Clear > Contents]	[Office button > Word Options > Customize > All Commands > Contents]
8	Certain group functions have been removed.	11a	[Tools > Online Collaboration > Meet Now]	Removed from Word 2007
		11b	[Tools > Letters and Mailings > Letter Wizard]	Removed from Word 2007

Table 2. Eight Types of Violations Against Design Heuristics[1]

Type	*Word 2003*	*Word 2007*	*Type of Potential Usability Problem*
1	File | New | Blank document	Office Button | New | Create button	Change in location
2	Insert | Symbol | | Insert button | Close button	Insert | Symbol |	Change in visual search direction
3	View menu located at the third left	View tab located at the seventh left	Change of top menu's position
4	Standard Toolbar (Insert Table) | <Col x Row> Table	Insert | Table | <Col x Row> Table	Cannot locate corresponding toolbars
5	Insert | Field | CreateDate | <Date> | OK button	Insert | Quick Parts | Field | CreateDate | <Date> | OK button	Same top menu, same terminal but different paths
6	Table | Draw Table | <Table>[1]	First Way: Insert | Table | Draw Table | <Table>[1] Second Way: Home | Borders | Draw Table | <Table>[1]	Multiple ways for doing the same task in Word 2007
7	Edit | Clear | Contents	Office Button | Word Options | Customize | All Commands | Contents | Add button | OK button	Some commands need to be set up before use
8	Tools | Letters and Mailings | Letter Wizard | OK button	Removed from product	Some menu-driven commands have been removed

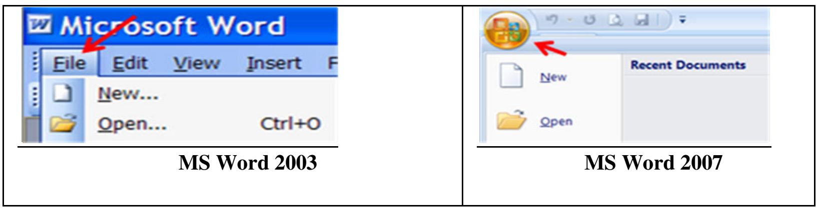


Fig. 1. Type 1 Violation

3 Discussion and Conclusion

The study reported in this paper was motivated by two concerns on users and usability research work on interactive systems [7]. Firstly, as evaluation techniques evolve to be more and more complicated and comprehensive, the impression designers/developers have towards usability engineering is often blocked by intimidation barriers such as "difficulty to conduct", "too time-consuming", "tedious" and "expensive". Secondly, there is a lack of practical evaluation techniques that designers can use to design and build more usable systems while meeting the time constraints of the design process.

Using Microsoft Word 2007 as a case in point of committing serious design violations against well-known design heuristics, this paper argues that ignoring common sense design heuristics could lead to detrimental outcomes such as user confusion and usability problems.

Despite much work done on improving software accessibility, usability can still be improved as shown in Word 2007 [1], and is worsened with new technologies and

[1] Push, hold, drag, and release left mouse button to define size of table.

applications. Many argue that it can be too slow and costly for the financial and time constraints of the design process. It is not good enough to design an interactive system without subjecting it to some form of evaluation, because it is impossible to design an optimal user interface in the first try. Even if one has used the best methodology and model in the design of usable interactive system, one still needs to assess the design and test the system to ensure that it behaves as expected and meets end-users' requirements. Nielsen's [5] advice is that designers should simply conduct some form of testing. If designers were to do things well at the start, these types of usability problems could be ameliorated and/or prevented. Otherwise, remedial work after the new interfaces or upgrades have been done could be costly and time-consuming.

References

1. Colazzo, L., Molinari, A., Tomasini, S.: Is new necessarily good? Testing usability of the new Office 2007 user interface. In: Luca, J., Weippl, E. (eds.) Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008, pp. 1371–1379. AACE, Chesapeake (2008), `http://www.editlib.org/p/28562`
2. Microsoft, Locations of Word 2003 commands in Word 2007, retrieved on 14 April (2009), `http://office.microsoft.com/en-us/word/HA100625841033.aspx#3`
3. Microsoft, Word, to Word, interactive command reference guide (2003), `http://www.microsoft.com/downloads/details.aspx?familyid=9044790b-4e24-4277-b714-66d7b18d0aa1&displaylang=en` (retrieved on April 14, 2009)
4. MSDN Blogs, Easing the Transition to Word 2007 (2007), `http://blogs.msdn.com/microsoft_office_word/archive/2007/06/07/easing-the-transition-to-word-2007.aspx` (retrieved on April 14, 2009)
5. Nielsen, J.: Usability problems through heuristic evaluation usability walkthroughs. In: Proceedings of ACM CHI 1992 Conference on Human Factors in Computing Systems, pp. 373–380 (1992)
6. TechRepublic, 10+ ways to help your users transition to Word 2007, Susan Harkins (2007), `http://blogs.techrepublic.com.com/10things/?p=407` (retrieved on April 14, 2009)
7. Theng, Y.L., Chan, M.Y., Khoo, A.L., Buddharaju, R.: Quantitative and Qualitative Evaluations of the Singapore National Library's Digital Library. In: Theng, Y.L., Foo, S. (eds.) Design and Usability of Digital Libraries: Case Studies in the Asia Pacific, pp. 334–349. Idea Group Publishing, PA (2005)

The Effect of Age, Gender, and Previous Gaming Experience on Game Play Performance

Mona Erfani[1], Magy Seif El-Nasr[1], David Milam[1], Bardia Aghabeigi[1], Beth Aileen Lameman[1], Bernhard E. Riecke[1], Hamid Maygoli[1], and Sang Mah[2]

[1] Simon Fraser University
Surrey, BC
{mea16,magy,dma35,baa17,ber1}@sfu.ca, beth@bethaileen.com, tonymaygoli@yahoo.com
[2] Bardel Entertainment
sang@badel.ca

Abstract. It is common sense that people don't play games that are too difficult for them. Thus Game developers need to understand the performance abilities of players. Several studies suggest a clear dissimilarity in video game playing abilities between different genders and age groups. In this paper, we report on a study investigating impact of age, gender and previous gaming experience on gameplay performance. The study explored the performance of 60 kids 6-16 years old within three video games: Rock Band 2, Lego Star Wars and Kameo. The paper outlines clear impact of age and gender and less prior gaming experience on performance parameters: score and game progression.

Keywords: Game play performance, player modeling, game design.

1 Introduction

One of the enjoyable aspects of games is challenge [1]. There is evidence showing that users enjoy games that they perform well in, but have not yet mastered [2-5]. Keeping the right balance in game is increasingly difficult due to the divide between gamers, casual gamers, and non-gamers. Thus developers need to know their users' behavioral abilities. Few researchers developed models that explain what players do in different game environments [6-8]. Some studies explored users' behaviors within virtual worlds and Massively Multiplayer Online Role Playing Games [9, 10] or mobile learning environments [11]. Other have argued the influence of gender, age, family race/ethnicity, and socioeconomic status on game play [12-16]. Chan found males play video games more than females [12]. Klimmt and Hartmann found boys and girls often prefer different games, females dislike games with violent content, lacked meaningful social interaction and had characters that were sexual stereotypes [14]. Other studies [18, 19] reported girls are motivated by 2D and easy, while boys by 3D and challenging. Also games with high level of competitiveness and task fulfillment are known to not attract female players as much as male [20]. In this paper, we explore performance as we believe performance is a key to motivation. As Lucas

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 293–296, 2010.

and Sherry [15] argue it is not gender that determines performance, but there are some underlying behavioral abilities that need to be examined.

2 Study Design

We had 60 participants: 18 females (average age=9.81), 42 male (average age=10.4) in 26 sessions. We chose cooperative popular games due to the results of our questionnaire which revealed kids in the age 6-16 often play together. Thus all sessions were cooperatively played a rhythm based game: *Rock Band 2 (RB)* and two action/adventure games: *Lego Star Wars (LSW)* and *Kameo* with 2-4 players (friend/family). We videotaped the sessions and interviewed players after each game using a structured questionnaire. We annotated videos with a set of defined metrics that evaluate player performance including difficulty level, level completed, number of points earned, number of deaths, etc. Analysis measured skills at the individual level as well as influence of the partner. We defined 'previous gaming experience' as a quantitative measure associated with the amount of times for the types of games they reported play. We used 14 genre classifications namely First Person Shooter, Role Playing Games, Sports/Racing, Music/Rhythm, etc. Questionnaire asked for their favorite games, genre of games they enjoy most, rate of play (hours/week), and length of a typical play session. This is limited as participants underestimate their time of play [17]. Questionnaire's results indicated *females spent less time playing video games ($p<.05$).* We found no significant age difference between genders.

3 Result

Findings regarding *Rock Band 2* (Rhythm based game). 46 participants ($n_{female}=11$) completed the song, mean age:10.21, while 14 ($n_{female}=7$) failed, mean age:8.28. Those failed only tried the Easy mode. ANOVA analysis revealed *age played a significant role for males ($p<.05$)* but not for females. Also there was a gender difference for those completed the song ($p<.05$), but not for those failed. For those completed the song *females scored significantly lower than males.* Analysis showed *there was a significant strong positive correlation between score and age for females $r=.746(7)$, p(two-tailed)$<.05$ and a weak positive correlation for males $r=.478(33)$, p(two-tailed)$<.01$.* On the other hand, for participants who failed the song, *percentage of completeness was higher for females. There was a significant positive correlation between age and percentage of song completeness for males $r=.896(5)$, p(two-tailed)$<.01$, but not for females.* Regarding prior gaming experience, *there was a positive correlation between time spent playing games and score but only for males.*

Findings regarding *Lego Star Wars* (3D action/adventure game). 54 participants ($n_{female}=16$; $n_{male}=38$) played for average of 10:01 minutes. *More males finished the level than females ($p<.05$).* To keep consistency we isolated the participants who played the game before; 28 ($n_{female}=4$) participants have played the game before, while 26 ($n_{female}=12$) did not. Analysis indicated older participants without *LSW* experience could complete the level. However, there were several cases which had experience with *LSW* but didn't complete the level. Analysis of exceptions indicated partners' experience level greatly affected the females, but not males. *We concluded for males, age and previous experience were factors affecting their game performance, while for females, partners also had an effect.* There were 12 obstacles, required participants to (a) move objects using force ability, (b) pass platforms using special abilities e.g., high jumps. The mean for number of obstacles solved was 9.4 and for number of deaths was 6.4. In these two parameters, experience with *LSW* as well as gender played significant roles ($p<.05$); *females solving fewer obstacles and dying frequently*. Regarding the impact of age on number of obstacles solved, *there is a significant positive correlation for males r=.396(38), p(two-tailed)<.05, but not for females.* Completeness time of obstacles was not appropriate. They involved fighting, solving, and collecting items. Players spent time on different activities. Also analysis showed no significant differences for time on game genre and obstacle resolution.

Findings regarding *Kameo* (3D action/adventure game). 56 ($n_{female}=16$; $n_{male}=40$) participants played same level for an average of 9:17 minutes. Only 7 male could finish the level. Overall *older participants finished the level*. Similarly, we isolated participants who played the game before; 9 (all male) played it before, while 47 ($n_{female}=16$; $n_{male}=31$) did not. There were 9 obstacles required participants to (a) perform character shifts, (b) understand capabilities of characters, (c) shift to appropriate character, and (d) use environment to jump. The mean for number of obstacles solved was 4.37 and for number of deaths was 1.23. *Prior experience with Kameo increased the number of obstacles solved (p<.05) but didn't have an effect on the number of deaths (p>.05).* Regarding gender, *females cleared less obstacles than males (p<.05) while no significant difference for the number of deaths (p>.05).* Regarding age and number of obstacles solved, both genders showed significant positive correlation: males r=.587(40), p(two-tailed)<.01 and females r=.655(16), p(two-tailed)<.01. However only males showed positive correlation between number of deaths and age r=.571(40), p(two-tailed)<.01. This meant *older participants solved more obstacles, and thus had more encounters with enemies.* Similarly, analysis showed no significant differences for time on game genre and obstacle resolution.

4 Conclusion

In this paper, we reported the impact of age, gender, and prior gaming experience on game play performance of 60 kids. Not surprisingly, we found significant gender and age influence on performance. For all games used within the study, the percentages of males who completed the songs (*RB*) or levels (*LSW* or *Kameo*) were higher than females. Females scored lower than males (*RB*) and solved smaller number of obstacles (*Kameo* and *LSW).* In *RB* age had significant impact on score for females, but not for males while in *Kameo*, it had impact on number of obstacles solved for both genders and in *LSW* only for male. It is interesting that in *LSW*, the number of obstacles solved by females was affected by their partners' skills and performance. Also time spent playing video games had effect on scores of males in *RB*. This was surprising to us as we expected to see more impact of previous gaming experience. These results play a significant role within the game design and development process, as they give designers some guidelines for more gender and age inclusive designs. However, more studies with more and mature players in each game genre are needed to investigate the conclusiveness of prior gaming experience.

Acknowledgments. The study was funded by MITACS (Mathematics of Information Technology and Complex Systems), a Canadian Network Center of Excellence (NCE), and Bardel Entertainment, a virtual worlds company in Vancouver, BC.

References

1. Malone, T.W., Lepper, M.R.: Making learning fun: A taxonomy of intrinsic motivations for learning, vol. 3, pp. 223–253. Erlbaum, Hillsdale (1987)
2. Koster, R.: A theory of fun for game design. Paraglyph Press (2005)
3. Sweetser, P., Wyeth, P.: Gameflow: a model for evaluating player enjoyment in games. Comput. Entertain. 3 (2005)
4. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper Perennial (1991)
5. Bateman, C.: Beyond Game Design: Nine Steps Towards Creating Better Videogames. Charles River Media (2009)
6. Bateman, C., Boon, R.: 21st Century Game Design. Charles River Media, California (2005)
7. Bartle, R.: Hearts, clubs, diamonds, spades: Players who suit muds. International Journal of Virtual Reality 6, 11–16 (1996)
8. Drachen, A., Canossa, A.: Analyzing Spatial User Behavior in Computer Games using Geographic Information Systems. In: MindTrek (2009)
9. Reid, D., Fitzpatrick, G.: Acting your age in Second Life. In: Fun and Games (2008)
10. Taylor, T.L.: Play Between Worlds: Exploring Online Game Culture. MIT Press, Cambridge (2006)
11. Fitzpatrick, G., Hooper, G., Weal, M.: Does it matter who is holding the PDA in a mobile learning experience? In: IADIS Mobile Learning (2008)
12. Chan, E.: Girls Playing Games: The Effect of Gender Stereotypes on Video Game Playing Motivation and Performance. In: Meaningful Play 2008, East Lansing, MI (2008)
13. Jenson, J., de Castell, S., Fisher, S.: Girls playing games: rethinking stereotypes. In: Future Play, Social, ethical and cultural perspectives on games (2007)
14. Hartmann, T., Klimmt, C.: Gender and computer games: Exploring females' dislikes. Journal of Computer-Mediated Communication 11, 910–993 (2006)
15. Lucas, K., Sherry, J.L.: Sex Differences in Video Game Play: A Communication-based Explanation. Communication Research 31 (2004)
16. Sandberg, D., Meyer-Bahlburg, H.: Variability in middle childhood play behavior: Effects of gender, age, and family background. Archives of Sexual Behavior 23 (1994)
17. Williams, D., Consalvo, M., Caplan, S., Yee, N.: Looking for gender (LFG): Gender roles and behaviors among online gamers. Journal of Communication (2009)
18. Ziemek, T.K.: Two-D or not Two-D: Gender Implications of Visual Cognition in Electronic Games. In: Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games, pp. 183–190 (2006)
19. Ziemek, T.K.: Electronic games: 2D or not 2D? International Conference on Computer Graphics and Interactive Techniques. In: ACM SIGGRAPH (2005)
20. Agosto, D.E.: Girls and gaming: a summary of the research with implications for practice. Teacher Librarian 31(3), 8–14 (2004)

New Attitude to Learning in Virtual Environments - Mining Physiological Data for Automated Feedback

Zdena Lustigova[1], Aude Dufresne[2], and François Courtemanche[3]

[1] Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
[2] Department of Communication, University of Montreal, Montreal, Canada
[3] Department of Computer Science and Operations Research, University of Montreal, Canada
lustigo@plk.mff.cuni.cz,
{aude.dufresne,François.Courtemanche}@umontreal.ca

Abstract. We present the great potential of real time data, obtained from 1/ mouse and keyboard events' monitoring, 2/ physiological measurements of the individualized reaction on the learning process of the learning subject in virtual learning environments. We emphasized simple, non invasive, easily available methods like eyes tracking, blink rate and blink speed measurements, electrodermal activities measurements, and heart and/or respiration rate. Those methods have big potential to reflect decreasing attention, increasing visual or cognitive information load, task difficulty, tension, stress and fatigue. We compared the 'real time' data with records obtained from screen captivate SW, video and audio records, records of external observers and learners (volunteers) interviews. We highlight the advantages and constraints of different data acquisition approaches, as well as constraints, done by hardware and software limits, and discuss the future potential for automated learners' feedback within VLE.

Keywords: virtual learning environment, automated feedback, learning, physiological measurements, data mining.

1 Introduction

Within the progress of information technologies for education, and the progress of physiological measuring tools, the chance to reflect real and immediate learning and emotional needs of a subject, who learns, has emerged in an inconceivable way.

Simple, noninvasive, low cost measurements of eyes blinks, galvanic skin response (GSR) or heart rate are easily available and with the progress and miniaturization in technologies, transferable to the 'out of laboratory' conditions – into the real learning environments – home, classroom, internet coffees, etc.

In the following, we present selected partial experiments (including examples of obtained data sets and their description), which reflect various cognitive brain activities and also emotional stage of the learner, and discuss their advantages and constraints with the goal to reveal their potential for real time individualized reaction on the learning process of the subject during computer based learning within virtual learning environment.

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 297–300, 2010.

2 Mouse and Keyboard Events Monitoring and Reflection

The keyboard and mouse activity monitoring is rarely used for educational and learning purposes in virtual learning environments, although it can bring a full range of new information, highlight some parts of the learning process and help to improve virtual learning environment itself. For our purposes we did not used monitoring tools downloadable from the Internet. We incorporated 'spying' software onto Java applets/physlets, which are traditionally used for teaching physics at our secondary schools. For additional monitoring and reflection we used also 'real time' screen captivating SW, video and audio records and records of external observers. The time synchronization (soft real time) of all monitoring tools and methods (mentioned above and bellow) is one of the biggest problems, we still solve.

3 Eye Tracking and Blinking Patterns

Among many different data, obtained from computer connected eye tracking systems, the changes in spontaneous blink rate seem to be the most easily detected significant factor for cognitive and visual information load. Blink rate (BR) is low when information memory is operating. Blinking is suspended during certain cognitive activities to avoid disrupting these processes [1], [2].

Wong, Wan and Kaye [2] during their research among eight surgeons found the significant reduction in the average blink rate between two conditions (casual conversation and operating, using the microscope -16.69/min and 4.75/min, p= 0.0002 paired t test). We verified their results on a group of eight students (age:19-28). Surgeons, operating, and students trying to read the values from the onscreen graph (Fig.: 1), both decreased the average blink rate in a significant way (see table 1).

Table 1. Average blink rates (No/min) for students (age 19-22) during periods of casual conversation and while focused on solving physics problem presented as Java applets

Student	1	2	3	4	5	6	7	8	Mean
Casual conversation	11,8	17,7	7,2	17,6	13,8	22,5	11,2	13,8	14,45
Solving problem	0,9	6,4	0,5	6,4	4,7	7,6	1,3	3,8	3,95

Blinking rate personal pattern is relatively stable in time (Barbato et al. [3]) According some researchers (e.g. [4]) blink rate and especially blink speed reflect also the level of fatigue. Blinking rate is easy to detect comparing to other eyes features detection (saccades, fixations, gaze position, etc).

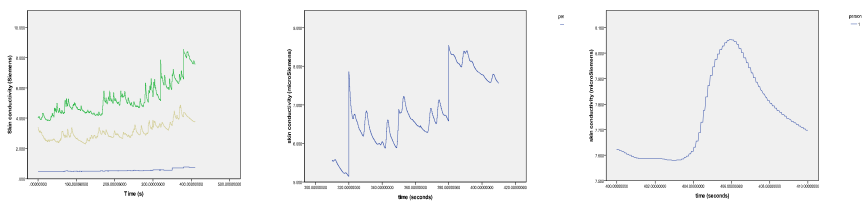
BR measuring also brings some constraints. According Patel at al. [5] observations, for example, visual tasks requiring concentration result in a similar decrease in average blink rate; we obtained from our experiments on cognitive information load. Other researchers [6] also hesitate about BR results, measured under different baseline conditions.

4 Galvanic Skin Response

GSR - galvanic skin response is one of the several electrodermal responses. GSR consists of two components: tonic and phasic. The tonic component is a low frequency baseline conductivity level, which can oscillate over the course of days. Each person has a different tonic conductivity, generally in-between 2 and 20 μS. The phasic component rides on top of the tonic component, is of higher frequency, and generally increases when a person is aroused.

Tonic skin response rises in anticipation of performing a variety of tasks and during the performance of these tasks (mental arithmetic, attention tasks, discussing social issues).

The low frequency baseline conductivity level (tonic component), which can oscillate over the course of days and generally slowly increases with the level of the person's arousal, is characterized by large individual differences. See Fig. 2a



a b c

Fig. 1. a/The individual differences in both phasic and tonic components of the skin conductance during the exposition to the same web side content. b/ Problems with automated data interpretation, person badly connected, leads to misinterpretation (scaled 0-1 μS). c/ Typical response, illustrating the time parameters for rise time and half recovery time. (scaled 7,5-8,1 μS). Data obtained by authors.

Real time data processing and interpretation of GSR signal is not simple. Phasic component is usually evaluated on the base of amplitude and latent period. For the evaluation of the relative amplitude of the phasic reaction in percent relative to the tonic level is often used. The signals of both, the tonic and phasic component, are not periodic; their spectra are continuous and overlap. For phasic component data interpretation signal processing methods are rarely used, although some methods, e.g. FFT, seem to bring new information [7]. Usually the amplitude in μS , latitude, rise time and half recovery time (typical values in the same order are 3s, 1-3s, 2-10s), are measured [7]. For illustration see Fig. 3c. Typical problem with computer based real time data interpretation is illustrated on Fig. 2a.

5 Heart Rate, EEG, Respiration Rate

While GSR changes are mostly connected with emotional arousal and the level of person's involvement and BR is highly influenced by the level of fatigue and visual acuity, the heart rate and EEG measurements undoubtedly reflect mainly mental activities.

Evidence, since the beginning of 20th century, indicates that alpha power is inversely related to mental. Amplitude decrease in the alpha rhythm, called “alpha blocking," or “alpha desynchronization “ has been reported for several cognitive tasks such as mental arithmetic, tasks taken from IQ tests, and creative problems (e.g. [8], [9], [10], [11]). Generally the heart measurements and especially EEG were provided mainly in laboratory conditions. Recently, EEG and HR together with respiration rate measurements improved a lot, and according our experiments seem to have a great potential for the “out of laboratory” large scale use.

References

1. Holland, M.K., Tarlow, G.: Blinking and thinking. Psychological Reports 41, 403–406 (1975)
2. Wong, K.K., Wan, W.Y., Kaye, S.B.: Blinking and operating: cognition versus vision. British Journal of Ophthalmology 86(4), 471–479 (2002)
3. Barbato, G., Ficca, G., Muscettola, G., Fichele, M., Beatrice, M., Rinaldi, F.: Diurnal variation in spontaneous eye-blink rate. Psychiatry Research 93, 145–151 (2000)
4. Bittner, R., Smrcka, P., Pavelka, M., Vysoky, P., Pousek, L.: Fatigue Indicators of Drowsy Drivers Based on Analysis of Physiological Signals. In: Crespo, J.L., Maojo, V., Martin, F. (eds.) ISMDA 2001. LNCS, vol. 2199, pp. 62–68. Springer, Heidelberg (2001)
5. Patel, S., Henderson, R., Bradley, L.: Effect of visual display unit use on blink rate and tear stability. Optom. Vis. Sci. 68, 888–892 (1991)
6. Cho, P., Sheng, C., Chan, C., et al.: Baseline blink rates and the effect of visual task difficulty and position of gaze. Current Eye Research 20, 64–70 (2000)
7. Ishchenko, A.N., Shev’ev, P.P.: Automated complex for multiparameter analysis of the galvanic skin response signal. In: Biomedical Engineering, vol. 23(3), pp. 1573–8256. Springer, New York (May 1989), Print, 0006-3398, Online, 1573-8256
8. Donchin, E., Kutas, M., McCarthy, G.: Electro cortical indices of hemispheric specialization. In: Lateralization in the Nervous System, pp. 339–384. Academic Press, NewYork (1977)
9. Lustigova, Z.: New e-learning environments for teaching and learning science. In: 20th IFIP World Computer Congress on Learning to Live in the Knowledge Society, pp. 363–364. Springer, Berlin (2008)
10. Martindale, C.: Biological bases of creativity. In: Handbook of Creativity, Cambridge Univ. Press, New York (1999)
11. Nunez, P.L.: Mind, brain, and electroencephalography. In: Nunez, P.L. (ed.) Neocortical Dynamics and Human EEG Rhythms, pp. 133–194. Oxford Univ. Press, New York (1995)

Personalized Sightseeing Tours Support Using Mobile Devices

Ricardo Anacleto[1], Nuno Luz[1], and Lino Figueiredo[1,2]

[1] GECAD, Knowledge Engineering and Decision Support Research Center
[2] ISEP, Instituto Superior de Engenharia do Porto, Porto, Portugal
{rmao,nmal,lbf}@isep.ipp.pt

Abstract. In this paper, we present PSiS (Personalized Sightseeing Tours Recommendation System) Mobile. PSiS Mobile is our proposal to a mobile recommendation and planning support system, which is designed to provide effective support during the tourist visit with context-aware information and recommendations about places of interest (POI), exploiting tourist preferences and context.

Keywords: Mobile Recommendation System, Sight Information Provider, Context-Aware, Client-Server Application.

1 Introduction

It is well known that the task of planning where to go and what to do, in the limited amount of time available, are common problems encountered by tourists when visiting a city for the first time.

In effect, cities are large information spaces, and in order to navigate these spaces visitors often require numerous guide books and maps that provide large amounts of information. Although the amount of information allows tourists to select more appropriate points of interest, it also turns the process so complex that the tourist might not be able to assimilate all this information adequately.

Mobile devices applications can be used to provide an effective support to tourists in tour planning process. Mobile terminals are embedded systems with very limited capabilities (limited display size, resolution, power consumption, processing capabilities, low memory and networking capacity) compared to a traditional computer. These limitations need to be considered because of possible technical, ergonomic or economic implications for the mobile user. So, mobile client-server applications performance is a crucial aspect that depends on many factors, such as bandwidth, connectivity, positioning capability and support for the paradigms of interaction, the user interface and security issues.

2 PSiS Mobile

PSiS (Personalized Sightseeing Tours Recommendation System) [1] is a tour planning support, it aims to define and adapt a visit plan combining, in a tour, the most

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 301–304, 2010.

adequate tourism products, namely interesting places to visit, attractions, restaurants and accommodations, according to the tourists specific profile (which includes interests, personal values, wishes, constraints and disabilities) and available transportation system between different locations. To ensure a good visit plan, working schedules as well as transportation schedules are considered.

The system gathers knowledge about the tourist's profiles, creating groups and stereotypes with specific interests and features, allowing characteristic inheritance. Tourists travel history is stored, which leads to accumulated knowledge about personal profiles. This knowledge, together with tourist stereotypes offers a mean of learning about general and specific interests of tourists. Also, it is possible for the tourist, to provide feedback on accomplished tours.

Currently, PSiS only interacts with tourists through a web application accessible only from a browser, but it's indispensable to have a tool to assist tourists "on the field". Thus, we are studying and developing a mobile tool to be integrated in the PSiS project, called PSiS Mobile. This tool also takes into account the tourist current context and nearby sights context.

In a preliminary phase, PSIS Mobile will be limited to data from the city of Porto, Portugal. But it will be designed so that no data or user restrictions are imposed. It is composed by two pieces, the server-side and the mobile client. All the main information like user profiles, history and similarity values, is compiled on the server. In other words, all the recommendation aspects are on the server, since it classifies sights with a rate to that specific user. There is a complete database with all information about points of interest in a certain city/region, and a complete portfolio of users as well as their visit history.

The mobile client is a very important piece in all system, because it interacts with the user. With a PDA, the user can see his generated route and provide information about his context with this the system can offer a more effective recommendation about places to visit and can do the re-planning of the original visit, in real time. The system interacts with the user providing information about nearby sights to see, these points of interest are recommended according to user profile and context; Shows trip planning for current day, that can be re-arranged according to current context, for example, if tourist is behind schedule a planning algorithm is executed to do a re-plan; Shows favorite sights stored on the system.

With PDA we know user's current context, i.e. its location, day/time information and traveling speed. And with this information we can get even more information, for example, we can know what is the weather forecast for that location at that moment, to refine the recommendation (to not suggest outdoor spaces to visit). With traveling speed, planning can be made more effective, taking that speed and calculate the time that takes to get from one to another point of interest. Besides the already mentioned information, on or behind schedule.

Our mobile application will manage some basic recommendation routines only. What we mean with this is that it will not classify (or rate) points of interest, but only show the results to the user. For example, if a user likes Chinese food, certainly a Chinese restaurant has a higher classification value according to the user preferences (might not happen if classification is given using collaborative filtering, since the restaurant might have a negative classification by similar users).

Our mobile client will show the points of interest, for that user and for a specific category, ordered by classification (downward; higher classification appears first). After visiting some points of interest, the user can provide feedback about the visited place.

When the user is going to see a point of interest, the application will show detailed information about it. With this, the tourist knows more about what he is going to see, for example, the history of a museum.

We also want to implement Augmented Reality in our system, meaning that the user will have the options of pointing the PDA to the point of interest direction and access to its detailed information. These details will include information like pictures of the point of interest in other seasons of the year (i.e., covered in snow). This application will offer built-in social networking too, so the user can share his pictures with the community in a matter of seconds. Despite all the features to be implemented in our application, we want a smooth and easy to understand interface to the user. To facilitate navigation throughout the application it is essential that the number of clicks between various features is kept at the minimum.

To conclude, we want to make a real application that really helps people on seeing what they expect to see, or going where they like. It is important to develop an optimized communication mechanism to ensure that a tourist does not waste too much time just to gather the necessary information.

In our case scenario we have Microsoft server side technology, and all the recommendation system is working under .NET framework. The database, that is present in the same physical server, is implemented on SQL Server 2008.

At user side, we have a PDA running the Android OS. The problem is that Android uses Java technology. So, there are two different modules, implemented with different technologies that need to communicate. Another issue is the low RAM memory capacity: only 288MB for the whole system, so we need to be very careful with the mobile application development.

Since we have two different technologies communicating with each other, and the base system is already implemented, we must create a middleware that bridges communications between these two technologies. This means that the mobile middleware will play a crucial role on the system.

Some of the important features of the middleware include security, data synchronization, device management, and the necessary support for multiple devices. Because this will be an occasionally connected application (Smart client), a temporary database is used on the mobile device to permit access to parts of the data without constant traffic consumption over the network, and to allow the application to work without internet connection (with multiple limitations, like no access to new points of interest and recommendations).

First of all, after requesting a recommendation for a trip, all the necessary data is transferred from the server and stored on the mobile device. We find this to be necessary, because of the low Internet speed rates on mobile phones and its possible unavailability. When we say necessary data, we mean, the information of all the nearby points of interest that will be on the planning schedule, and other points of interest nearby the first ones. This approach is useful if the tourist wants to re-plan the schedule in real time.

All the collected data, photos, user context and others will be stored on the device, to then be sent to the server.

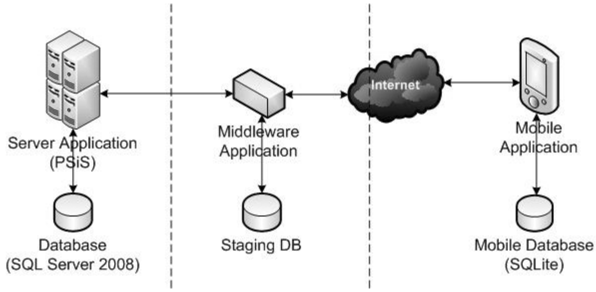


Fig. 1. Mobile Architecture Overview

This architecture from Figure 1 can be summarized saying that:

- The existent system will does not need to be changed;
- The middleware application is a component that will reside on the server side and will be developed on .NET Framework, with directives to permit the communication between the existing system and the mobile application;
- The mobile application runs on Android devices and is used to capture/send data from/to the field. The application also has a synchronization component to synchronize the handheld data with the server database;
- Internet connection is used to retrieve/update itinerary information, sites information and personal preferences. Data is uploaded and downloaded automatically without user intervention.

3 Conclusion

We have introduced the main concerns present in the development of a mobile application in a client-server environment. Although mobile devices have many constraints, with PSiS Mobile we pretend to provide a good user experience, giving tourists a fast and user friendly tool including context-aware adaption, a route planning system, augmented reality and built-in social networking features, to provide the user with important and significant details about what he is seeing or is about to see.

Reference

1. Almeida, A.: Personalized Sightseeing Tours Recommendation System. In: The 13th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2009, Florida, USA (2009)

Reaction on a Change of User Context in Complex Mobile User Adaptive System

Ondrej Krejcar

VSB Technical University of Ostrava, Center for Applied Cybernetics, Department of measurement and control, 17. Listopadu 15, 70833 Ostrava Poruba, Czech Republic
Ondrej.Krejcar@remoteworld.net

Abstract. People life is today surrounded by a digital world environment which enables to interact with user and to react on user needs not only after user's requests, but also before the user needs grow. Paper deals with a concept of User Adaptive System (UAS) with two areas of user context along with reaction on them. A Predictive Data Push Technology (PDPT) Framework concept is mentioned as a possible solution fo user location context changes as well as a Biotelemetrical Monitoring System (BMS) is suitable for user biomedical data changes reaction on.

Keywords: User Adaptive System; Mobile Device; Biotelemetry; Prebuffering.

1 Introduction

The idea of User Adaptive Systems (UAS) grown from interaction between user and system (e.g. throws his mobile device). Such interaction can behold in the reaction on user's non declared requests. These requests are based on current user environment and biological or emotional state (e.g where I am?, what I feel?, am I ok?, etc.). Such user questions can be answered by sensors on user body or inside the user devices. By the help of user mobile device, we can get a user location (e.g. user current position, user future-predicted position, his movement and tracking, etc.). Biomedical sensors on user body can detect several important biomedical data, which can be used for determination of user emotional state in the environment around.

By the combination of user's requests (known or predicted) in conjunction with other sources of user's knowledge and behaviors, the sophisticated information system can be developed as UAS. The impact of UAS can be seen in the increased user comfort when accessing these mobile UAS. In ideal case, everything what user can imagine to have in his mobile UAS is there.

A one specific kind of problems is based in increased data amount in new mobile systems. In current cases, the user need to specify a data to be downloaded to his mobile device and he need to wait for data downloading and displaying. Due to a several limitations in hardware of current mobile devices, the use of such large amount data has result in lower user comfort. The needs of any techniques to reduce such large data amount or to preload them before user's needs, is still growing up. We created a Predictive Data Push Technology (PDPT) Framework to solve these

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 305–308, 2010.

problems by data prebuffering. Our idea can be applied on a variety of current and future wireless network systems. More usability of PDPT grows from definition of area to be prebuffered as well as from evaluation of artifacts or other user's behavior sources. Additional will be presented in sections [2], [3].

The second area of problems which we would like to solve is based on a users biomedical data inputs and a wide area of their possible utility. Current body sensors allow a monitoring of a huge number of biomedical data information (e.g. use a special t-shirt equipped with an ECG, temperature, pressure or pulse sensors). Current hi-tech mobile devices are equipped with a large scale display, provide a large memory capabilities and a wide spectrum of network standards plus embedded GPS module (e.g. HTC Touch HD, HD2). These devices have built-in also a special accelerometer which can be used to determine a user’s body situation (user is staying or lying). Last but not least equipment is a light sensor which can be used not only to brightness regulation.

2 Reaction on a Change of Location

We can imagine the usage of such described UAS in the information systems area of botanical or zoological gardens. In such areas there has been a big potential of usage of a continual localization by use of GPS or wireless networks (in case the GPS has not a sufficient signal – e.g. in urban centers or neighborhoods with high buildings, forest parks or in deep valleys). There is also a possibility to compute a current and predicted user track, so we can predict a position of user in near future (e.g. 25 meters north in one minute). Usability of these information sources is uncountable.

One of possible use of user predicted position is for a determination of a data, which will be needed by user of mobile UAS in near future. Such data (data artifacts) can be preloaded to user’s device memory for future requests. The need of preloaded artifacts grown from a need of up to date data context of dynamic online system. Of course when static offline system is used, there is a possibility to load a needed data before usage (e.g store artifacts at SD Card with a size limit to several GB). When user request info about his location in context of zoo or garden (turn-on the device is only needed by user), the client application will respond with a map of near surroundings and a prebuffered data artifacts. User can select a documentary about animals or vegetation around him which can be viewed or played. User can act with direct requests to selected kinds of these. These preferred kinds will be taken into account to evaluate future objects/artifacts and preloaded only the most important ones for a user. The type of artifact is also evaluable as well as his size because the user may not want to look at too long or micro presentation. From several statistic info obtained from users tracks, the most frequented ways in gardens can be also found. By the help of some mentioned info sources a very sophisticated dynamical area definition can be developed.

As client devices of online UAS, the mobile wireless devices like PDA or Smart phones are commonly used with an internet connection. The connection speed of the two most common standards GPRS and WiFi varies from hundreds of kilobits to several megabits per second. In case of online UAS or some other types of facility management, zoological or botanical gardens, libraries or museums information systems, the WiFi infrastructure network is often used to connect mobile device clients to a server. Unfortunately, the low performance hardware components are used in PDAs or SmartPhones due to a very limited space. By this fact a theoretical maximum connection speed is not achievable on these devices. The limited connection speed represents a problem for online systems using large artifacts data files. It is not possible to preload these artifacts before the mobile device is used in a remote access state. This problem was found as a very important. The goal is to combine the data networking capabilities of RF wireless LANs [4], [5] with accurate user location and tracking capabilities for user needed data prebuffering. These properties are used as an information base for an extension of existing UAS or to create special new ones. Information about location is used to determine an actual and a future position of a user. The following section describes also the conceptual and technical details about PDPT Framework.

3 Reaction on a Change of Biomedical Data

We are using principles of UAS in area of biomedical data processing, where we try to predict some kind of problems by patient data analysis. We developed a context-aware Biotelemetrical Monitoring System (BMS) [6] as a part of the UAS and PDPT Framework project facilitates the following:

- Real-time collection of the patient vital signs (e.g. ECG, EEG) by means of a Body Area Network (BAN) or direct wireless connection to PDA device monitoring station.
- Real-time transmission of data using the wireless connectivity to the healthcare professionals through a complete architecture including a server database, web services, doctors web access to patients collected and preprocessed data.
- Seamless handover over different wireless communication technologies such as BlueTooth, WiFi, GPRS or UMTS.
- Context-aware infrastructure to sense the context (e.g. location, availability, activity, role) of the patients and Emergency Response Services (ERSs) to provide assistance to the patient in case of an emergency. An ERS could be fixed (e.g. hospital) or mobile (e.g. caregiver). A mobile ERS is published in the BMS.

Classical access to patients request are made by reactive flowchart, where a patient is equipped with a classical offline measuring devices with some type of alarms. Every violated alarm need to be a carried out by doctor decision. Such access is very time-consuming. Second proposed access is based on a proactive principle, where the patient is equipped with an online measuring devices with an online connection to some kind of superior system (in our case the BMS is presented). In this case, a patient's measured data are processed on mobile monitoring station or at server. An alert will invoke when the anomaly data are founded in patient's records. Consequently the doctor is responsible to make a decision to invoke other ERSs or to remove Alarm (in case of false detection of anomaly). Such kind of behavior is based on UAS. In many of events a predicted and solved problems can save a life. The predicted patient's problems are in most cases minor in compare to a major problems detected in time where occurred.

4 Conclusions

A two main directions of UAS as a reaction a change of location and reaction on a change of biomedical data were described with main focus on use of mobile devices. Coexistence of proposed solutions is in unnumbered areas and the results of complex solution are better than expected. The developed UAS can be stocked on a wide range of wireless mobile devices for its main issue at increased downlink speed. The localization part of PDPT framework is currently used in another project of biotelemetrical system for home care agencies to make a patient's life safer [6]. Several areas for PDPT stocking was founded in projects of hydro-information system "Transcat" [1], [2]. Several of UAS principles can be used there also. The last but not least was found in social dimension of UAS where the focus is now moved from explicit to implicit collaboration [7]. These possibilities will be investigated in future.

Acknowledgement. This work was supported by the Ministry of Education of the Czech Republic under Project 1M0567.

References

1. Horak, J., Unucka, J., Stromsky, J., Marsik, V., Orlik, A.: TRANSCAT DSS architecture and modelling services. Journal: Control and Cybernetics 35, 47–71 (2006)
2. Horak, J., Orlik, A., Stromsky, J.: Web services for distributed and interoperable hydro-information systems. Journal: Hydrology and Earth System Sciences 12, 635–644 (2008)
3. Krejcar, O.: Problem Solving of Low Data Throughput on Mobile Devices by Artefacts Prebuffering. EURASIP Journal on Wireless Communications and Networking, Article ID 802523 8 (2010)
4. Brida, P., Duha, J., Krasnovsky, M.: On the accuracy of weighted proximity based localization in wireless sensor networks. In: Personal Wireless Communications. IFIP, vol. 245, pp. 423–432 (2007)
5. Krejcar, O.: Large Multimedia Artifacts Prebuffering in Mobile Information Systems as Location Context Awareness. In: 4th International Symposium on Wireless Pervasive Computing, ISWPC 2009, Melbourne, Australia, February 11-13, pp. 216–220 (2009), doi:10.1109/ISWPC.2009.4800591
6. Krejcar, O., Janckulik, D., Motalova, L.: Complex Biomedical System with Biotelemetric Monitoring of Life Functions. In: Proceedings of the IEEE Eurocon 2009, St. Petersburg, Russia, May 18-23, pp. 138–141 (2009), doi:10.1109/EURCON.2009.5167618
7. Longo, L., Barrett, S., Dondio, P.: TOWARD SOCIAL SEARCH From Explicit to Implicit Collaboration to Predict Users' Interests. In: WEBIST 2009, 5th International Conference on Web Information Systems and Technologies Inst. Syst. & Technol. Informat., Control & Commun., Lisbon, Portugal, March 23-26, pp. 693–696 (2009)

Augmented Reality for Deaf Students: Can Mobile Devices Make It Possible?

Becky Sue Parton, Robert Hancock, and John Dawson

Southeastern Louisiana University

Abstract. Digital and real world events can be combined to create powerful learning opportunities for students, but time, tools, and expertise have been traditional barriers to teacher-created enhancements. This paper provides a rationale for using emerging, teacher-friendly tools, to merge real space and virtual space through video and 2D barcodes. The results of three pilot studies combine to illustrate the potential for using these tools. Results indicate that cell phones have the potential to facilitate augmented reality experiences for deaf students and adults.

Keywords: Augmented Reality, Mobile Devices, Accessibility, Deaf.

1 Introduction

According to Burik and Kelly [1], active learning has long been considered a proven method for increasing attention, motivation, and retention of concepts especially among deaf students. Although there are many definitions for interactive multimedia, Akhtar [2] described it as using many different media (print, audio, video, etc) to present more comprehensive information than any medium alone can, accommodating children with different learning styles, and employing interactivity to stimulate children to become active, motivated learners. Beyond the commercial and researcher-oriented instructional design projects, perhaps those with the most practical impact are the ones created by individual instructors. As the design tools become easier to use and more affordable, teachers can craft specially designed interactive software in a reasonable amount of time [3]. Every year more interactive material is developed by the teachers and that in turn has led to students participating in the design process themselves. For example, at the Laurent Clerc National Deaf Education Center, students collaborated to build an interactive video dictionary [4].

In taking the concept of interactive multimedia to an even higher plane, researchers have coined the term ‘physical world hyperlinking’ to capture the notion of connecting a tangible object to an on-line resource. The idea is to converge ‘realspace’ and ‘mobilespace’. Tangible technology is a term that refers to directly linking computer-based activities with real-world physical objects and events [5, 6, 7]. When digital information is presented to users that enhances their real world surroundings the term Augmented Reality is often used. For example, popular applications for the iPhone such as Wikitude allow users to focus their phone camera on their surroundings and

P. Forbrig, F. Paternó, and A. Mark-Pejtersen (Eds.): HCIS 2010, IFIP AICT 332, pp. 309–312, 2010.

see data, like restaurant locations, layered on top of the image. For Deaf students, the augmented information could be sign translations.

The ultimate purpose of this paper is to explore how 2D barcodes and You Tube integrated digital video can be combined by a classroom teacher to provide augmented reality experiences accessible through cell phones for Deaf students. In order to provide a rationale for using those tools, it is necessary to look at a pilot study that focused on Global Positioning Satellite (GPS) technology.

2 Augmented Reality Experience One: Using GPS

A pilot set of materials were developed in late 2007 for the purpose of enhancing an on-site fieldtrip experience at Tiger Creek Wildlife Refuge with digital content. Tiger Creek is home to about forty big cats, most of whom have been rescued from abusive situations. The sanctuary provided the background stories for each tiger, and the lead author recorded video clips of a certified interpreter retelling the stories in American Sign Language (ASL). The video had to be captured and then compressed for web use – a total of 25 were created. Finally, a website was constructed to contain the videos, along with still photos, of each tiger. The research team then took GPS readings near each tiger exhibit area, recorded them in mapping software (Microsoft Streets and Trips), and coded the various flag points to launch web pages.

A group of seven high school students from the Jean Massieu Academy for the Deaf volunteered to participate in the pilot. In advance of the fieldtrip, a three hour training session was held for the purpose of showing the students how to use a micro laptop (a Sony Vaio UX) and GPS unit. At Tiger Creek each student had approximately forty-five minutes to locate an exhibit and launch the interpreted story.

The pilot project provided valuable insight into the concept of linking the digital and physical world. In written feedback and individual post-study interviews, all the students expressed positive comments regarding the value of the video component and the enjoyment level of the experience as a whole. For example one student said, “It’s so awesome – just to have fun with the computer and enjoy seeing the real tigers”. The researchers examined the screen capture footage from the laptop and determined that participants spent an average of 11 minutes and 8 seconds actively using the laptop to find flags, launch web pages, and watch video stories. The rest of the time the student was either watching the live tigers, taking photos/notes, or discussing events with peers or guides.

However, there were many limitations to the experience. First, from the perspective of the teachers who accompanied the group, the event was not duplicable. Specialized skills were required to make the videos, post the videos on websites, and code the GPS points – all of which was time consuming. In addition, the hardware was expensive and not readily available in the school. From the perspective of the research team, the system appeared cumbersome in regards to the GPS because a signal was often hard to obtain due to the tree coverage. The battery life of the micro laptop was also an issue. Thus a major finding of the study was the conflict of student satisfaction with the practical feasibility for teacher produced fieldtrips.

3 Augmented Reality Experience Two: Using 2D Barcodes

Based up on the pilot study outlined above, the researchers set out to find an alternative approach that was less labor-intensive and not cost-prohibitive but equally enjoyable for students. A process that is currently in widespread use in Japan, and to a limited extent in other countries, is called 2D barcode camera phone scanning. Using this technology, a person walking around a Cineplex might notice movie posters that had a 2D barcode image and use his/her cell phone to take a picture of that image which would launch the trailer video. Free software is available to users who want to create their own QR codes at http://qrcode.kaywa.com. This technology appeared to be a viable option for hybrid fieldtrips, so two studies were designed: 1) a survey to find out if teachers had access cell phones capable of running the barcode software and 2) to determine the best hardware for launching the barcodes.

In order to read the images, users need a camera phone with the reader software installed. In late 2008, an on-line survey was sent to 547 K-12 educators. The educators had self-selected themselves into the group receiving the survey because of their previous use with Hotchalk learning environments, an ad-based alternative to Blackboard. One hundred and twenty three teachers completed the survey and all participants indicated that they owned a cell phone. Two questions were asked on the survey: 1) Can you view web pages from your cell phone? And 2) Does your cell phone have a built-in camera? These two features are required in order for barcode software to work, although not all cell phones capable of performing these two actions have barcode software available. Data revealed that 49% of the teachers had web viewing capability and 81% had photo taking capability on the phones.Although not part of the survey study, it should be noted that as each day passes, there are more phones on the market, such as Apple's iPhone that have this capability.

Beyond the issue of whether teachers have cell phones capable of reading 2D Barcodes, is the issue of how easy or difficult it is for people to activate the barcode launch mechanism. Therefore, the research team setup a second study. At a state-level technology conference a booth was setup to test a participant’s ability to retrieve the linked data from two encoded barcodes with two different cell phones. The sample size was thirty one people who voluntarily chose to participate, but three surveys were removed from the analysis because they were only partially completed.

The first barcode linked to a short URL, whereas the second barcode linked to a much longer URL thus requiring more bits of data to be encoded. The two cell phones used for the study were a Motorola Q with a 1.3 mega pixel camera and an AT&T Tilt with a 3.0 mega pixel camera. Subjects tried to read both barcodes with both phones and then responded to a survey. They were asked to mark whether a successful launch of the linked information occurred for each combination and then to rate the ease of use from one to seven with seven being the friendliest. Factors influencing ease of use may have included the number of tries (i.e. length of time) it took for the camera to recognize the code, the steadiness required to read the codes, the distance that the user had to maintain from the code, and the phone’s interface for maneuvering between the code reading screen and the web browser.

Three patterns emerge from the data: 1) All the participants were successful in being able to use the barcode reader to launch content, 2) A cell phone with better resolution (i.e. more megapixels) was easier to use, and 3) Barcodes that were encoded were more complex data were harder to read with low resolution cameras, but the difference became less noticeable with the higher resolution camera. Specifically, only 35.7% of the subjects rated the 1.3mp camera easy to use (a score of 6 or 7) in regards to the complex URL; whereas, 100% of the subjects rated the 3.0mp camera easy to use (a score of 6 or 7) in regards to the complex URL. Based upon these findings, the research team decided to re-design the field trip experience.

Although 2D barcodes could have been linked to the web pages created previously during the Tiger Creek pilot, there was a concern about the embedded video being universally playable across multiple types of phones. Additionally, the research team wanted to continue making the concept of merged realties more attainable for the average teacher so a decision was made to link directly to You tube videos rather than websites. Many user-friendly camcorders are capable of creating and posting you tube videos without any video editing skills or equipment such as the Flip camcorder and even the iPhone itself. For the Tiger Creek presentations, however, the researchers still created the You tube videos by mixing still photos and the previously captured interpreted story videos. The resulting videos were uploaded to You tube and the barcodes were generated. A working barcode for the tiger, Mowgli, can be found at: http://www.lifted-up.org/awe/TigerWeb/barcodes.htm. For readers who do not have a barcode reader, the image is also directly hyperlinked to the video. In actual use, the barcodes are printed and distributed to participants. The response to the revised materials was positive enough to warrant construction of a set of cards for the New Orleans Zoo where a formal study is anticipated this year.

Students want interactive, multimedia experiences. Until recently, teachers have been unable in many cases to develop that type of enhanced material because of the skills, equipment, and time required. This paper has outlined how a once complex hybrid fieldtrip can now be created by a teacher with limited access to a computer.

References

1. Burik, L., Kelly, W.: Active learning through technology-creating a technology- infused environment to engage Deaf students in the learning process. Paper Presented at the Technology and Disabled Persons Conference, Las Angeles, CA (2003)
2. Akhtar, A.: A study of interactive media for Deaf learners in post 16 education. Paper Presented at the Technology and Education of the Deaf Symposium, Rochester, NY (June 2003)
3. Rawlings, L. (n.d.): Clicker4 and More – multimedia and multiple support. Unpublished Document Retrieved November 10 (2003), http://education.qld.gov.au
4. Stifter, R.: Integrating technology and literacy: Digital video dictionary. Paper Presented at the Technology and Education of the Deaf Symposium, Rochester, NY (2001)
5. Chipman, G., Druin, A., Beer, D., Fails, J., Guha, M., Simms, S.: A case study of tangible flags: a collaborative technology to enhance field trips. Paper Presented at the 5th International Conference for Interactive Design and Children (IDC), Tampere, Finland (2006)
6. Price, S.: A representation approach to conceptualizing tangible learning environments. Paper Presented at the 2nd International Conference on Tangible and Embedded Interaction (TEI), Bonn, Germany (2008)
7. Sung, J., Levisohn, A., Song, J., Tomassetti, B., Mazalek, A.: Shadow box: an interactive learning toy for children. Paper Presented at the International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), Jhongli City, Taiwan (2007)

Author Index