Department of Computer Science
AALBORG UNIVERSITY

Proceedings of the Fourth Danish Human-Computer Interaction Research Symposium

16th November 2004

Edited by
Jesper Kjeldskov, Mikael B. Skov and Jan Stage
Department of Computer Science, Aalborg University

HCI Lab Technical Report no. 2004/1

Centertryk
Aalborg University • Sohngaardsholmsvej 57
DK-9000 Aalborg • Denmark

Proceedings of the Fourth Danish Human-Computer Interaction Research Symposium

16th November 2004

Aalborg University, Denmark

Organizing Committee

Benedikte Skibsted Als, Department of Computer Science, Aalborg University
Rune Thaarup Høegh, Department of Computer Science, Aalborg University
Jesper Kjeldskov, Department of Computer Science, Aalborg University
Jeni Paay, Department of Information Systems, The University of Melbourne
Mikael B. Skov, Department of Computer Science, Aalborg University
Jan Stage, Department of Computer Science, Aalborg University

Proceedings Editors

Jesper Kjeldskov, Mikael B. Skov and Jan Stage
Department of Computer Science, Aalborg University

HCI Lab Technical Report no. 2004/1

Preface

The Fourth Danish Human-Computer Interaction Research Symposium was held at Aalborg University's Department of Computer Science on November 16th 2004. The aim of the symposium was to stimulate interactions among HCI researchers from academia as well as industry through a mix of oral presentations, discussions of posters, and keynote presentations.

We received 20 paper contributions for the symposium, which are all included in these proceedings. 12 papers were presented orally at the symposium and 8 papers were presented as posters.

Jesper Kjeldskov, Mikael B. Skov and Jan Stage
Aalborg University, 16 November 2004

Table of contents

Fourth Danish HCI Research Symposium

November 16, 2004, Aalborg University

Introduction

Aalborg University is happy to host the Fourth Danish HCI Research Symposium. The aim of the symposium is to stimulate interactions among HCI researchers from academia as well as industry through a mix of oral presentations, discussions of posters, and keynote presentations. The focus of the symposium is on HCI research, e.g. usability work; novel interfaces; web design; affective computing; psychological models; computing in music, creative arts, design, and architecture; design of input devices; support of collaborative work; e-government; speech input; information navigation; mobile devices; learnware; visualization; and home computing. However, reflections on and challenges of HCI work based on industrial experiences are highly welcomed as well.

Programme

09:00 - 09:30 Registration and coffee

09:30 - 09:45 Welcome

09:45 - 10:30 Keynote Speaker

Ellen Christiansen: HCI Research Challenges: A U-turn and a Change in the Driver's Seat

10:45 - 11:45 Paper session 1

- Janni Nielsen and Carsten Yssing: What Kind of Information does an HCI expert want? - on concurrent usability testing
- Nette Schultz, Dan Saugstrup and Lene Sørensen: A User-Centred Scenario Framework Using Creative Workshops
- Jesper Kjeldskov, Mikael B. Skov and Jan Stage: Evaluating IT Systems for the Healthcare Domain: Longitudinal Usability Studies and Rapid Analysis Techniques

12:00 - 13:00 Paper session 2

- Christina Brodersen, Bent G. Christensen, Kaj Grønbæk Christian Dindler and Ole Sejer Iversen: eBag – the digital school bag
- Lars Bo Larsen: The Automatic Pool Trainer - a Platform for Experiments with Multi Modal User Interaction
- Morten Hertzum: Remembering Multiple Passwords by Way of Minimal-Feedback Hints: Replication and Further Analysis

13:00 - 14:00 Lunch

14:00 - 15:00 Paper session 3

- Torkil Clemmensen and Shivam Goyal: Studying cross cultural think-aloud usability testing – some suggestions for an experimental paradigm

- Benedikte Skibsted Als, Janne J. Jensen and Mikael B. Skov: Usability Evaluation with Children
- Amalia de Götzen, Stefania Serafin, Juraj Kojs and Rolf Nordahl: Fitts' law in an audio perspective

15:15 - 16:00 Panel Discussion: Validity in HCI Research

- Lene Nielsen (Copenhagen Business School)
- Nette Schultz (Technical University of Denmark)
- Christina Brodersen (University of Århus)
- Morten Hertzum (Roskilde University)

16:00 - 16:15 Poster introduction

16:15 - 17:00 Coffee and posters

- Christina Brodersen, Christian Dindler and Ole Sejer Iversen: eCell – making room for collaboration
- Malene Damborg and Brith Klarborg: PDA as a 'Cognitive Prosthesis'
- Eva Hornecker: A Framework for the Design of Tangible Interaction for Collaborative Use
- Anker Helms Jørgensen: How Can HCI/Usability and Computer Games Enrich Each Other: A First Look
- Lene Nielsen: Design Through Engagement
- Jeni Paay: Studying User Experience in Public Places
- Stefania Serafin, Juraj Kojs, Amalia De Goetzen and Rolf Nordahl: Multimodal interactions in physically modeled sound synthesis
- Lori Webb: Design Souveniring: An Experimental Approach Aimed at Improving Cross-Cultural Awareness in Global User Interface Design

17:00 - 18:00 Paper session 4

- Erik Frøkjær and Kasper Hornbæk: Studying usability evaluation to improve its practical utility
- Rune T. Høegh and Jan Stage: Examining the Use of Usability Results in a Software Development Company
- Georg Strom: How usability practitioners can get their suggestions implemented in industrial software design

18:00 - 18:30 Closing

18:30 - 19:15 Demonstration: The Automatic Pool Trainer

19:30 - 21:30 Dinner

Place
Aalborg University, Niels Jernes Vej 10, 9220 Aalborg East

Sponsors
The symposium is kindly sponsored by Aalborg University, Department of Computer Science

Organizing Committee
Benedikte Skibsted Als, Department of Computer Science, Aalborg University, als@cs.aau.dk
Rune Thaarup Høegh, Department of Computer Science, Aalborg University, runethh@cs.aau.dk
Jesper Kjeldskov, Department of Computer Science, Aalborg University, jesper@cs.aau.dk
Jeni Paay, Department of Information Systems, the University of Melbourne, jpaay@unimelb.edu.au
Mikael B. Skov, Department of Computer Science, Aalborg University, dubois@cs.aau.dk
Jan Stage, Department of Computer Science, Aalborg University, jans@cs.aau.dk

What Kind of Information does an HCI expert want? - on concurrent usability testing

Janni Nielsen
HCI Research Group
Department of Informatics
Copenhagen Business School
Telephone: +45 3815 2417
janni.nielsen@cbs.dk

Carsten Yssing
HCI Research Group
Department of Informatics
Copenhagen Business School
Telephone: +45 3815 2443
cy.inf@cbs.dk

ABSTRACT
Working with usability techniques, with focus on how a given technique enables data capture made us ask the question: What kind of information is it an HCI expert want from the user when conducting a usability test? We answer the question by discussing two techniques both relying on concurrent data. Think Aloud is one of the most frequently used techniques and almost an institution in itself. Eye-tracking is new in usability testing and still at an experimental level in HCI. We reflect critically upon the two obtrusive techniques. We discuss the usability of concurrent data capture, suggesting participatory analysis and retrospective verbalisation as a possible step in usability testing.

Keywords
Concurrent data, usability test, Think Aloud, Eye-tracking, mind, participatory analysis and retrospective verbalisation

1. INTRODUCTION
It has been pointed out that most methods often are taken "off the shelves" rather than chosen on the basis of pre-analysis and grounded decisions (Bødker and Sejer Iversen, 2002). In a survey on user centred design (Mao et al, 2001) the claim is that the most frequently applied methods/techniques are: simple prototyping, heuristic evaluation and usability test. Usability testing focus on user's interaction with the computer, and our interest is the way the user's eye travels around the graphic dynamic interfaces, the actions of the user (the navigation) and the way the user experience the interaction: What does the user see, why does the user do what she does and what does she think? In our search for techniques we have come to wonder about what information a usability expert is looking for when conducting iterative user testing. Among usability test it is especially the concurrent techniques, which are of interest, and there are two techniques that stand out, but at each end of the usability line: Think Aloud (TA) and Eye-tracking. Think Aloud (TA) is old in the HCI business, it is one of the most popular techniques, used equally in industry and academia, and it is taken for granted that it gives access to user's thoughts. Eye-tracking is new to HCI, it is being applied rather exploratory but with great expectations and promises access to the user's visual interaction.

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

The short paper is organised around these two techniques. We describe the use of TA, which promises access to cognitive processes, and introduce Ericsson and Simons classical distinction between Talk Aloud, Think Aloud and Retrospective Verbalisation. We discuss the data capture and point out that the interface is visual, not verbal. Any usability technique should be able to capture the visual interaction, and we turn to Eye-tracking (ET) as it gets beyond the verbalisation requirement and promises access to visual data capture. We describe ET and point out that logging of cursor, fixation of cursor, paths of cursor, saccades of cursor and the different interfaces capture do not tell us what the users feel, think experience. We raise the question: What kind of data does one gets access to when conducting concurrent testing? We suggest that other approaches are considered, and discuss the data capture that retrospective verbalisation enables. We propose a next step in our work, which is cursor tracking and interface capture followed by participatory analysis.

2. THINKING ALOUD
Think Aloud (TA) is the most popular usability test, often referred to as the usability method. It is used equally by industry and academia (Boren and Ramey, 2000, Clemmensen and Leisner 2002). TA is cheap, does not require heavy investments in technology can even be conducted by non-usability experts and only requires 5-8 users. It seen as a straightforward technique, ready to use with proper handling (Hackos and Redish 1998, Molich, 1994), and has been given the credit of simplicity (Dix, Finlay, Abowd and Beale, 1997). Especially Jakob Nielsen has been a tireless promoter (1994, http://www.useit.com).

The understanding embedded in most studies is that the techniques allows us access to the cognitive processes, to mental behaviour and gives us insight into thinking.
By recording the verbal protocol, you will be able to "…detect cognitive activities that may not be visible at all" (Hackos & Redish 1998, p. 259). But caution has been voiced because the technique puts a cognitive load on the user, which may interfere with the cognitive requirements of the interaction hence disturb

the task. It has been argued that concurrent verbalisation is problematic because TA adds strain and cognitive load on the users ((Branch 1999) that users have difficulties in speaking and to speak aloud feels awkward (Preece, Rogers and Sharp 2002). Silence is the likely outcome of the situation, hence the need for the investigator to encourage the user to "keep talking".

The basis for TA in usability testing is the classic text on protocol analysis from 1984 by Ericsson and Simon. They discussed the use of introspective data in the study of task directed cognitive behaviour. It is the work of Ericsson and Simon, which have reinstalled verbal data as a valid resource for understanding human cognitive processes. They did this by introducing the technique of Think Aloud. Their understanding was that most performance measures rely on responses that are psychologically indistinguishable from a verbal report, because some kind of verbal reporting is usually necessary to understand people's actions, even in very simple tasks.

Ericsson and Simon distinguished between three kinds of cognitive processes in their model;

1. *Talk Aloud* is direct oral expressions of thoughts which already exist in verbal form
2. *Think Aloud* is verbalisation of a sequence of thoughts, that are held in memory in some other form, e.g. visually
3. *Retrospective verbalisations* of retrospective reports or thoughts not held in short term memory, i.e. explanations and descriptions.

They argue that a sentence is the verbal realisation of thought. The assumption is that everything we know has, at some point, gone through our short-term memory (STM), and we have been conscious of it. We can verbalise what we are learning while in the process of learning, and we can verbalise what we know if questioned shortly after the process of learning has taken place. This is because it is still retained in our short-term memory (talk aloud and think aloud). However, if there is a time span between learning and being requested to recall, we will produce descriptions and explanations (retrospective verbalisations) - not a report of our immediate thought, because the information from STM is lost. Ericsson and Simon were only interested in Talk aloud and Think Aloud, whereas they considered retrospective verbalisations too error prone due to the time lap and the reliance on users recall.

2.1 Critical Issues in TA

But what is it we get access to when asking users to think aloud? Does it really give us access to what goes on in people mind? Teaching graduate computer science students the TA test techniques, and requiring them to reflect on their experience raised a number of issues. Students complain that 1) they think faster than they can speak, 2) thought processes are much more complex than can be verbalised, 3) having to think aloud interferes with their interaction with the interfaces and the task and 4) thinking aloud does not come naturally. (Nielsen, Clemmensen and Yssing 2002a).

The assumption that performance measure has to rely on some kind of verbal data, and that a sentence is the verbal realisation of the thought has been questioned (Nielsen, Clemmensen and Yssing 2002b). The sentence that the user speaks is a verbal realisation of thought, but there is not a 1:1 relationship between thoughts, actions and the spoken words. Using TA requires the user to shift focus in attention from giving sense to that which is perceived and does not exist in verbal form - to that of constructing sentences or words, and then expressing them aloud. TA requires perception and actions to be transformed to talk. Even if the speech is immediate and runs concurrently with the thoughts – user's attention has to shift focus from understanding to verbalisation (Nielsen and Yssing, 2003). As a consequence the process of understanding is interrupted, because attention keeps changing object. TA may result in verbal overshadowing (Schooler, Uhlsson and Brooks 1993) and we do not get access to mind.

Besides the user is interacting with net-based multi modal interfaces. Colours, layout, forms, animations, video clip and endless jumps through links all interact and it is visualisations, which is the main feature. Hence the interaction is mental and based in visual perception, and thoughts are not mainly verbal and directly accessible in oral speech, but percepts, which are, to a large extent, tacit (Polanyi 1967).

3. EYE-TRACKING

In our search for techniques we took a closer look at Eye-tracking. It seems to get around the verbalisation problem, yet it captures concurrent data and the visual interaction. Eye-tracking builds on a mind-eye hypothesis and expectations are high. However, the technique is still at an experimental level within the HCI field. Where it has been used by to investigate cognitive workload and scanning behaviour (Salvucci 1999), interface and screen design (Ellis and Candrea 1998), to determine the position of visual feedback on the screen. (Rauterberg and Cachin 1993), and visual search on pull down menus (Byrne et al 1999).

There are different ways of conducting eye-tracking. One is the well known head mounted system with cameras that the user has to wear. One camera shows the scene that the user is looking at; the other camera is tracking the eye movements. A less obtrusive system is a remote eye-tracking system where the user works directly on the computer. A camera lens, mounted besides or under the computer, is focused upon the user's eye tracking the pupil, most often with an infrared source.

Eye tracking makes it possible to follow the visual fixations and scan-paths of the user. It captures a user's visual focus of attention on a visual display unit – through special hardware and software. It registers x/y location and pupil size/border line information. It is especially the jerky movements of the eyes (saccades), which are followed by fixations (x,y location) and combined into scan-paths, and the smooth pursuit of eyes tracking slowly moving objects, which are analysed (Goldberg and Wichansky 2003, Ellis and Candrea, 1998, Heyhoe, Shrivastava, Mruzcek and Pelz 2003).

3.1 Critical Issues in eye-tracking

The cost of acquiring, learning to operate and the maintenance of an eye tracking system have to be considered. Prices range from just around 18.000 Euro to 100.000 euro for a complete eye tracking system. Besides the need for calibrations are high, even with a remote tracking system. Hence the user is instructed to maintain a relatively stable head position because movements during tests require recalibration. But users have difficulties in keeping their head still, and the recommendation is to conduct "recalibration every few minutes" (Goldberg and Wichansky 2003). There are problems with pupil/borderline registration because the pupils contract and expand in response to light, there are large individual differences in eye tracking results, and users with glasses and contact lenses are often excluded. Even the analysis done by the computer software creates problems and "the investigator is strongly encouraged to review the (software) created fixations against images of viewed displays to ensure that the fixations are valid". Besides, the raw data has to be " aggregated off-line by the investigator into meaningful behavioural units of fixations and saccades", and the amount of data is enormous and the task is extremely time consuming.

But what do we get access to with eye-tracking? Ellis and Candrea (1998) used eye tracking to test a website with a two column lay-out with both text blocks and images. Links were embedded in the text blocks, but also images could be links to video sequences. They redesigned the web page in three variations: One version had many links, in the second version they replaced graphics with text blocks and in the third version they made the page look like a book page. They name it "dense-text". Their analysis of eye movement tracking showed that the dense text version scored highest on their usability test. But they conclude, somewhat surprised, " despite it's potentially superior usability, dense-text was the lowest rated of all the lay-outs by our testers". This data was not registered by the eye-tracking system, but was collected after the test when the users were asked to rate the different interfaces.

The surprise that Ellis and Candrea voice points to a serious problem with eye-tracking. It does not give access to mind. Interestingly, the title of their paper is "Windows to the soul"? qualifying it with "eyes reveal a great deal about a person's feeling and behaviour". But registration of eye movement does not tell us anything about the soul. Eye-tracking only register movement – not what goes on inside the human being. It does not give us access to mind, and the capturing does not tell us anything about users intentions. However, Ellis' and Candrea's data show that we need to go beyond the mere concurrent testing and follow up with further investigations. Though the hypothesis lying behind eye-tracking is the eye-mind assumption – we only have logging of cursor, fixation of cursor, paths of cursor, saccades of cursor and the different interfaces capture. We do not have access to mind and we do not know what the users feel, think experience.

4. FROM CONCURRENT TO RETROSPECTIVE USABILITY TESTING

Our initial attempt to solve the problem with the obtrusive concurrent techniques was to look for an unobtrusive data capture technique, which would also allow us to deal with the multimodal interfaces. The solution was a software tool which enables concurrent data capture with cursor tracking an interface capture. In this way we can look at what the use looked at, and we can see how the user interacted. However, cursor motion does not necessarily track where user's visual attention is, s/he may forget to move the cursor because something on the multimodal interface disturbs or pleases her, or even be lost in daydreaming.

This brought our attention to the third level in Ericsson and Simons model: retrospective reporting which are thoughts not held in short term memory, i.e. explanations and descriptions. Ericsson's and Simon's argument is that if there is a time span between learning and being requested to recall, the user will produce retrospective verbalisations - not reports of their immediate thought, because the information from STM is lost. And retrospective reporting is more error prone because it relies on user's subjective recall – not on "hard facts", and subjective verbal data are not considered valid.

4.1 Retrospective verbalisation and participatory analysis

User's mind cannot be observed or registered. The only way to get access to user's experience is by probing the user: What does s/he see, why does s/he do what s/he does and what does s/he think? Our suggestion is to combine the capture of cursor tracking and interface with retrospective reporting in a participatory analysis. This technique will get us beyond the "total subjective recall". Because one of the unique advantages with cursor tracking and interface capture is that there is a recording of the actual actions of the user. It can be replayed and shown to the user. What the user sees is what s/he saw while working on the test; only the screen capture includes user's movements with cursor. The recording of the screen and mouse can be stopped/resumed at any given point. During stops a recording of the user's comments and reflections unfolding as a consequence of the probing will be captured on top of the frozen image. Thus, in the final analysis, the investigator is in possession of concurrent data in the form of 1) an uninterrupted recording of user's interaction with the interface during the test, and a retrospective reporting in the form of 2) a recording of the user's interaction with the interface with interruptions, and with a voice over.

4.2 Closing comments

The capture of interface with cursor tracking combined with participatory analysis seems promising because the processes of insight that runs associatively while the user interacts with the computer application may become partly explicit, and not be a total subjective recall. We call this technique Mindtape (Nielsen and Christiansen) and the replay triggers a running commentary at the same time as the events take place on the capture. These images may enhance the user's access to, and help recall, the thought processes that took place. The verbalization flows easy with the actual sequence of events structuring – not the users memory. This is important; a Mindtape is structured by the actual user-computer inter-actions as they unfolded during the

test session. It is not the users memory which controls the recall, but the actual events.
Naming it eye/cursor movements, or eye-mouse correlation escape the fact that it is the hand the user has to move – represented on the interface through the cursor, hence coordination with the eye.

ACKNOWLEDGEMENTS

Thank you to the HCI research group at Copenhagen Business School. This work was made possible through the Danish Research Council grant ref. 9900403.

REFERENCES

1. Boren M. Ted and Judith Ramey (2000) Thinking Aloud: Reconciling Theory and Practice, |IEEE Transactions on Professional Communication, vol. 43, no. 3, September, p. 261-278
2. Branch, J.L. (1999) Investigating the information-seeking processes for adolescents, Library & Information Science Research, vol.22, 4, 371-392
3. Byrne Michael D., Anderson John R., Douglass Scott and Michael Matessa (1999) Eye Tracking the Visual Search on click-Down Menus, Proceedings of HI99, Pittsburgh, 402-409
4. Clemmensen, T and P. Leisner (2002) Community knowledge in an emerging online professional community, Proceedings of IRIS25, Denmark,
5. Dix A., Finlay J., Abowd G. and R. Beale (1997) Human-Computer Interaction, Prentice Hall, 1997
6. Ellis Steve and Ron Candrea (1998) Windows to the soul? What eye movements tell us about Software Usability, Proceedings of the 7th Annual conference of the Usability Professional Association, Washington DC, 151-156
7. Ericsson, K.A. and Simon, H.A. s984) Protocol Analysis. Verbal reports as data, Cambridge, Massachusetts
8. Goldberg, Joseph H. And Anna M. Wichansky (2003) Eye Tracking in usability Evaluation: A Practitioner's Guide, in Hyöna J., Racdach R. And Deubel H. (eds.) The mind's eye: Cognitive and Applied aspects of eye movement research, Elsevier Science BV, chapter 23, 493 - 516
9. Hackos, J.T. and Redish, J.C. (1998) User and Task Analysis for Interface Design, Wiley, USA
10. Hayhoe Mary M., Shrivastava Anurag, Mruczek Ryan and Jeff B. Pelz (2003) Visual memory and motor planning in a natural task, Journal of Vision (2993), 3, 49-63
11. Molich Rolf (1994) Brugervenlige edb-systemer, Nyt Teknisk Forlag, København
12. Nielsen, Jakob: http://www.useit.com/
13. Nielsen, Jakob (1994) Estimating the Number of subjects Needed for a Thinking aloud Test, International Journal of Human-Computer Studies, 41, 3, 385-397
14. Nielsen Janni and Nina Christiansen (2000) Mind tape: A Tool for Reflection in Participatory Design, in Proceedings, Conference Participatory Design, New York 2000
15. Nielsen Janni, Clemmensen Torkil and Carsten Yssing (2002a) Getting access to what goes on in people's Heads? NordiCHI proceedings, Denmark, 101-110
16. Nielsen Janni, Clemmensen Torkil and Carsten Yssing (2002b) People's Heads – People's Minds, Proceedings of the APCHI 2002 (5th Asia Pacific Conference on Computer Human Interaction): User Interaction technology in the 21st Century, 2002, China, Science Press, Beijing (897-906)
17. Nielsen Janni and Carsten Yssing (2003) Getting beyond the disruptive effect of Think Aloud, in Hertzum Morten and Simon Heilesen (eds.) Proceedings of the third Danish Human-Computer Interaction Research symposium. Datalogiske Skrifter, Roskilde University, 67-71.
18. Preece J., Rogers Y. And Sharp H. (2002) Interaction Design: beyond Human-Computer Interaction, John Wiley & Son
19. Polanyi, Michael (1967) Logic and Psychology, American Psychologist, no.23, 27-43
20. Rauterberg Matthias and Christian Cachin (1993) Locating the primary attention focus of the user, in Grechenig T. and M. Tscheligi (eds.) Human Computer Interaction (Lecture Notes in Computer Science 733, 129—140) Berlin, Springer
21. Salvucci Dario d. (1999) Inferring Intent in Eye-based Interfaces: Tracing Eye Movements with Process Models, Proceedings of CHI 99, Pittsburgh, USA, 254-261
22. Schooler Jonathan W. Uhlsson Stellan and Kevin Brooks (1993) Thoughts beyond Words, Journal of Experimental Psychology, vol. 122, 2, 166-183

A User-Centred Scenario Framework Using Creative Workshops

Nette Schultz
Assistant Professor
Center for Tele-Information
Technical University of Denmark
Ph. +45 4525 5173, DK-2800 Kgs. Lyngby
nette@cti.dtu.dk

Dan Saugstrup
Ph.D. student
Center for Tele-Information
Technical University of Denmark
Ph. +45 4525 5211, DK-2800 Kgs. Lyngby
saugstrup@cti.dtu.dk

Lene Sørensen
Associate Professor
Center for Tele-Information
Technical University of Denmark
Ph. +45 4525 5170, DK-2800 Kgs. Lyngby
lene@cti.dtu.dk

ABSTRACT
This paper describes the methodology, definition and construction of a user-centred scenario framework, integrating end-user requirements with Personal Area Network and Personal Network (PAN/PN) system requirements. Based on a creative end-user workshop four different scenarios were derived in relation to a diabetes case study. These served as input to an expert workshop, where the main purpose was to discuss end-user scenarios in terms of technical requirements and parameters for the establishment of a PAN/PN architecture.

The paper has a strong user-centred approach to ICT service and application development based on the widely accepted fact that future services and applications need to be developed with a strong focus on the end-users' needs and demands.

Overall the paper is based on ongoing work in the EU-project MAGNET (WP1, Task 3) where a scenario methodology framework has been developed to secure a more user oriented approach to converting end-user requirements into specific system requirements in ICT development.

Categories and Subject Descriptors
D.2.1 [**Software Engineering**]: Requirements/specifications - *Elicitation methods*; H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User interfaces - *User-centered design.*

General Terms
Management, theory and methodology, HCI, interaction design

Keywords
Creativity, scenarios, workshops, end-users, participatory design.

1. INTRODUCTION
The user-centred scenario methodology framework presented here is based on a joint work between Aalborg University (AaU) and the Technical University of Denmark (DTU) that run under the European Union Sixth Framework Project MAGNET (My personal Adaptive Global NET). The project is about development of future information and communication technologies (ICT), it addresses research issues in personal distributed environments, where the users interact with various entities, devices and systems, without having to pay attention to the physical location of these. In the MAGNET context, Personal Networks (PNs) encompass potentially all of a person's devices, being networking enabled and capable of connecting to a network, physical or wireless – based on a service or ad-hoc related setup construction.

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

1.1 Personal Networks
The concept of a PN is a central theme in the MAGNET project. The main idea is to make use of all available networks to bind together the entire sphere of interest for the end-user, e.g. the user's Personal Area Network (PAN) - immediate vicinity - and more remote (global) or distributed "network islands" containing work environment and location of friends, family members and other personal contacts.

In the MAGNET context, the PNs are user centric, which means that these PNs should be able to provide personalised services for the end-users and adapt the content to their context environment.

PNs aim to support the user's professional and private activities, without being obtrusive, and while safeguarding their privacy and security. In order to do this, the PN services must bridge the PN's resources (both hardware and software) and the users. Determining the services framework is the first important step. It is essential to examine a number of different user scenarios in terms of services and context, and hereby get an overview of the PN sphere. By grouping the PN services into several types of themes based on "quick and dirty" user scenarios it has been possible to include a large number of user needs. This grouping has been further elaborated on and defined into eleven themes [1], such as health care, shopping, transportation, etc.

Within MAGNET a PAN/PN based health care service for diabetics, based on the DiasNet case study from AaU, is further developed [1]. In short, the DiasNet (Diabetes Advisory System) case study develops a web-application for support of the individual diabetics in disease self-management. Currently, DiasNet runs non-commercially and is tested amongst a group of diabetics in the Northern part of Denmark. This case study represents the first of a number of cases on discussing PAN/PN support to end-users by 2010. This paper focuses on how end-user requirements in the diabetes case were derived and translated into high-level technical, system parameters for establishing a valid PN architecture.

1.2 End-User Requirements

It is a challenge to translate the needs from the user's context to a structured language and method. And it is also difficult to define the specific technical parameters for the general theme level, because those parameters such as bandwidth and radio coverage radius may be more related to the specific user activities. Therefore the study area must be narrowed from the theme level down to the case level, i.e. focus on the specific user scenario, and follow the top-level user requirements structure [2] to analyze and group the user's needs and demands.

Through the work, the following top-level user requirements structure has been used: *Personal user requirements* are focusing on what the end user perceives experiences and does, such as usability, user experience and personalization; *Environmental requirements* are focusing on the surrounding society, such as security, privacy and ethics; *Industry domain requirements* are focusing on technical constraints in ICT development, and on policies and procedures within and between the PN stakeholders, such as billing and pricing. These requirements have been included in the user centric scenario construction process.

2. THE SCENARIO METHODOLOGY

From a methodological point of view, the main task has been to derive a scenario template, which can be used to link the user perspectives and requirements present in all MAGNET cases with the technology and components of a future PAN/PN architecture.

2.1 The scenario template

Scenarios were here defined as [3]: *"- a descriptive set of plausible but structurally different futures"* .

The futures are conceived through a process of causal, rather than probabilistic thinking, reflecting different interpretations of the phenomena that drive the underlying structure of the surrounding environment. The scenarios are linked to the year 2010. In practice the scenarios developed are based on a socio-technical perspective, which focus on the user requirements in terms of interoperability and global coverage. The scenarios, therefore, address; different user requirements and groups, user needs; user interfaces and device interaction; and discuss the social implications of use and development of PAN/PNs.

In order to develop consistent and solid scenarios, the scenario template used here is derived from well recognised references. By combining [3] and [4] a seven step template has been derived. Central for this template is the expression of the focal issue (the aim) of the scenario development, the analysis of the current and future situation of both local factors as well as general external, environmental factors which all may influence the scenarios. A detailed description of the template steps can be found in [5].

Figure 1 shows the practical interpretation of the scenario template. All actions are linked to provide a foundation for construction of a variety of PAN/PN architectures supporting different job and domestic situations in the year 2010, where there is given high consideration to what end-users may wish for and need in the future. The scenario framework encompasses two workshops addressing the user requirements and needs and translates these into the PAN/PN concept. These are the so-called creative end-user workshop and the MAGNET expert workshop. The end-user workshop addresses both the current situation for the users as an outset to identify future possibilities, needs and results in end-user scenarios. The MAGNET expert workshop addresses the future end-user needs (and end-user scenarios) and translates these into overall technical, system requirements. The results from these two workshops, together with an analysis of the general environmental trends and expectations for the year 2010, serve as input to the scenario representation in the end.

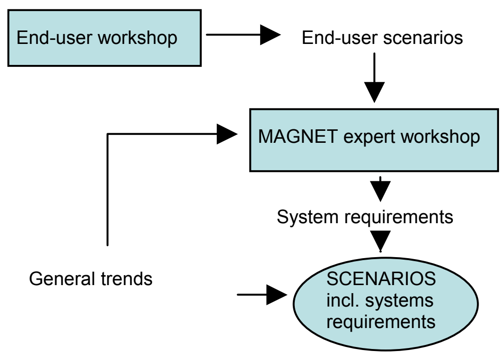


Figure 1: The practical input to the scenario template

Throughout the rest of the paper, the main focus is related to the two workshops, as these are the main steps in the process of developing the scenario frameworks. This includes the end-user workshop, the development of end-user scenarios, the MAGNET expert workshop and the translation of system requirements. The final scenario development (including final system requirements) is still to be carried out and is therefore left out of this presentation.

2.2 Workshops

The purpose of the workshops is to involve the end-users in the design process, not only in the evaluation/test phase but also in the design phase. Creative workshops have, therefore, been developed as a participatory design method, dedicated to the MAGNET framework.

The MAGNET framework has two prominent challenges. The first challenge is that end-users generally are not able to understand a PAN or PN requirements specification. Therefore, when defining future PAN and PN requirements, possible future applications and services, user requirements and scenarios must be defined together with the end-users. The second challenge is that MAGNET consists of many different partners with different competences. Collaboration is essential when deriving future PAN and PN requirements and scenarios.

For development of the scenarios there have been scheduled two successive workshops for each individual case study: A first creative workshop with end-users and a second PN architecture workshop with MAGNET-participants. In the first end-user workshop, end-users interact and participate in a creative design process. The creative design process focuses on the end-users and their needs, requirements and wishes for technology to support them in a particular situation. In the second workshop experts from the MAGNET project participate. The purpose of this second workshop is to relate the end-user results to components in a PN architecture and discuss different options for technologically supporting the end-user results.

In the first workshop users' address, comment and envision about technological concepts they may not know much about. The underlying framework of the workshop is therefore based on a number of creativity and participatory design theories. The stimulation of creativity is central to envisioning and getting new ideas for usage and creation of new devices and should be used when thinking about the future in order to get some new and significantly different ideas of technology and to secure that we address a different situation than today. Creativity is not something that can be expected coming immediately from the end-users, so it is important that special considerations are made to stimulate the creative process.

For stimulation of the creative process Gardner's multiple intelligences is used [6]. These intelligences are stimulated as part of a design process where pictures and words describing technologies, direct themes and user requirements central for the end-users are applied in the workshop. To ensure a user-centred process the participatory design tools with inspiration from the techniques originated in PICTIVE [7] were applied. The prototyping tools used were a subset of end-user related image elements developed to fit the specific case. The image elements included words and image metaphors representing technologies, themes (identified in [1]) and specific situations defining the exercise as meaningful as possible.

In the second workshop, MAGNET participants identified and described the industry domain requirements based on the result from the first workshop. During this workshop the personal user requirements and environmental requirements, which were represented in the end-user scenarios created after the first workshop, were "translated" into industry domain requirements. The underlying principle for this second workshop is collaboration and teamwork between the different sciences and technology areas represented in the MAGNET project. Only through such teamwork, all aspects of a PAN/PN architecture can be discussed. Further details on the two workshops can be found in [5].

3. WORKSHOP RESULTS

The following results are related to construction of end-user scenarios in the diabetes case. The primary focus here will be the process and results from the end-user workshop.

3.1 The End-User Workshop

The workshop participants were a mixture of end-users and MAGNET partners from AAU and DTU. A few characteristics should be made on the invited end-users: The diabetes patients, doctors and nurses were all adult, working persons covering ages from mid 20's to mid 60's. The diabetes patients had different educational backgrounds but had all been diabetics for a long period of their lives.

The workshop took place over a whole day where all workshop participants were present. First a shared user evaluation of diabetic helping devices available today was used as a pleasant start-up. Then a guided fantasy journey was used to introduce the participants to the diabetes life-cycle landscape and give a kick-start to begin the creative design process. The participants were divided into two groups and placed at two different tables loaded with paper, various elements representing PN components (such as computers, PDAs, etc.), scissors, play dough, glue, pens, etc. These tools were used to create a landscape with pictures, photographs, words, etc., representing ideas and situations for the diabetic person. First each group was asked to make a picture of their current situation as a diabetic person. The workshop participants were then asked to focus on 4 different life cycles: baby/child, teenager, adult and elderly stages, and relate discussions/visions to 2010. Each life cycle stage is characterised by different needs and requirements. Using the picture of the current situation representing an adult diabetic person, the groups moved into envisioning how it could be (preferably) in the near future (again representing the adult person). After this, focus moved to discussing the same issues for the older diabetes person, the teenagers and the babies/infants. In all life cycles, various image elements were used as input, as well as post-it notes to create notes and drawings. Later the groups were asked about the more technology specific questions related to the identified user requirements from [1]. The questions were discussed and did create an offset for discussing more technology specific requirements seen from the user side.

After the workshop, four different end-user scenarios for 2010 were developed (relating to the four life cycle stages). These scenarios are shortly presented below. The complete scenarios can be found in [5].

3.1.1 The baby/child scenario

Main focus of the baby/child stage was surveillance where parents are overall responsible for the treatment of the disease. This was made possible through three central devices; two measuring and injection handling devices on the child communication together and a third parental device for monitoring, communication and data storage. Communication could take place with other devices situated in kinder garden, at the hospital or other places. Extra features added were an alarm for emergency situations and a GPS for locating the child.

3.1.2 The teenager scenario

Central for teenagers is that they often express their rebellious minds in an urge to ignore the disease in order to be just like every other teenager. In order to solve this problem, an intelligent device (camouflaged, easy to wear and use and with a trendy design) could be used. This device would be able to regulate the diabetes automatically, while at the same time function as a template for independent contacts to the hospital and other central places. Furthermore, the device would include facilities for getting information on the disease for non-diabetics and for emergency calls.

3.1.3 The adult scenario

Central for the adult working diabetic is "freedom" and "self-control" in the sense that the diabetic can treat and monitor his disease himself (at least to a certain extent). This was made possible in a combination of two devices carried by the diabetic himself. These devices were able to monitor and regulate the disease. One of these devices, the intelligent device, could automatically or on manual request transmit the treatment data into the computer system at the pc and at the hospital. Additionally, the intelligent device could be used for communication purposes and for checking the meeting schedule at the doctor.

3.1.4 The elderly scenario

For elderly diabetic persons, monitoring and surveillance are central issues. This is the case both when the elderly person is at home but also when travelling. This requires an intelligent device carried by the diabetic which can monitor the disease and treat it, can transfer data to the hospital for further monitoring, and serve as a basis for making emergency calls when needed. Additionally, the device should be able to exchange information with foreign hospitals when travelling, and it should be solid and specific enough for the elderly to be able to swim for example.

These scenarios were used as template for the expert workshop participants' to discuss the relevant PAN/PN perspectives supporting the end-user scenarios.

3.2 Expert Workshop

The overall purpose of the expert workshop is to discuss the diabetes end-user scenarios in terms of technical system requirements and related parameters regarding the establishment of a supporting personal network and personal area network. The workshop participants were split up in two groups, each allocated with two different scenarios but given the same issues and questions to discuss regarding the allocated scenarios. Overall the issues and questions to be discussed were spilt up in industry domain and environmental requirements, thereby linking the end-user requirements and perspectives with the specification requirements for establishing solid PN/PAN structures.

The workshop results consist of two complementary, child/teenager and adult/elderly, overall PAN/PN scenario structures with corresponding high-level system requirements. A full description of the results can be seen in [5]. An important side-effect of this workshop was the collaboration and team spirit among the MAGNET partners that evolved through the workshop-day.

4. CONCLUSIONS AND FUTURE WORK

A user-centred scenario methodology framework has been developed to fit the purpose of the MAGNET project – derive and translate end-user requirement into technical PAN/PN requirements. In this procedure the workshops have played a central role in the user centred design process applied in the scenario development. Creativity has been the underlying framework for securing that visions and ideas of the year 2010 were addressed in the end-user workshop. Teamwork has been the underlying framework for securing that the different aspects from all the different participants were represented in the MAGNET expert workshop.

From the end-user workshop, four end-user scenarios have been constructed representing four different life-stages in a diabetic's life. From the MAGNET expert workshop, these four end-user scenarios have been discussed in terms of technical (industry domain) requirements and environmental requirements (as defined in [1]). Overall PAN/PN specifications have been identified mixing existing and new standardisation requirements. These will serve as input to other work packages within the MAGNET project for further system development.

Overall creativity has been used as a central template in the participatory design process carrying the scenario development. The practical way Gardner's intelligences have been applied in the workshop enabled the end-users to actually discuss technologies and systems they themselves would not have thought of otherwise. This means that the end-users views and ideas are represented directly in the MAGNET project, and as such has created a sound foundation for pointing towards PAN and PN architectures that in time will be needed and therefore, hopefully, will be successful.

In order to strengthen and further develop the diabetes end-user scenarios, future work must be conducted to find other studies, trends and details. Secondly the scenario methodology (and workshops) outlined here will be further developed, challenged and modified when addressing other cases and acknowledging comments and suggestions from partners.

5. ACKNOWLEDGMENTS

The work presented in this paper is supported by the European Union Sixth Framework Project "My Personal Adaptive Global NET (MAGNET)", contract no. 507102. Particular acknowledgements should be given to the DiasNet research (and user) group at the Institute of Health Science and Technology and the Department of Communication Technology, AaU.

6. REFERENCES

[1] MAGNET deliverable D1.1.1a, *Draft user requirements for PN to drive the definition of a valid architecture*, Mar. 2004.

[2] Jiang, B., Kaldanis, V., Markopoulos, A., Monti, M., Prasad, R., Saugstrup, D., Schultz, N. and Skouby, K.E., User requirements & Demand for Services and Applications in PNs. In *Proceedings of 13th IST Mobile & Wireless Communications Conference* /Summit2004, Lyon, France, 27-30 June, 2004, Pages 176-180.

[3] van der Heijden, *The Art of Strategic Conversation.* John Wiley & Sons, Chichester, 1996.

[4] Schwartz, P., *The Art of the Long View.* Currency Doubleday, New York, 1991.

[5] MAGNET deliverable D1.1.3.1a, Draft User Centric Scenarios for PNs of a valid architecture. Sept. 2004.

[6] Gardner, H. *Multiple Intelligences: The theory in practice.* Basic books, 1993.

[7] Muller, M.J. *PICTIVE – An exploration in participatory design.* Proceedings of CHI'91, 1991.

Evaluating IT Systems for the Healthcare Domain: Longitudinal Usability Studies and Rapid Analysis Techniques

Jesper Kjeldskov, Mikael B. Skov, Jan Stage
Aalborg University, Department of Computer Science
Fredrik Bajers Vej 7, DK-9000 Aalborg, Denmark
{jesper, dubois, jans}@cs.aau.dk

ABSTRACT
This paper describes two studies evaluating the usability of IT systems for the healthcare domain. First we describe a longitudinal study of an Electronic Patient Record system, secondly, we describe a new technique for rapidly analyzing usability data.

Author Keywords
Usability, novices versus experts, rapid analysis

ACM Classification Keywords
H.5.2 User Interfaces: Evaluation/methodology

INTRODUCTION
As a part of the Digital Northern Jutland project, we have conducted a series of usability studies of IT systems for the healthcare domain between 2002 and 2004 in collaboration with Sygehus Vendsyssel and Virtual Centre for Health Informatics at Aalborg University. This paper describes and outlines some of our findings from two of these studies; a longitudinal usability study of Electronic Patient Records and a comparison between a rapid analysis technique (Instant Data Analysis) and traditional video data analysis.

A LONGITUDINAL USABILITY STUDY
In our first study, we investigated into the way novice and expert users experience the usability of an Electronic Patient Record system (EPR). Electronic Patient Records allow nurses and medial doctors to access and register information such as state, diagnosis, treatment, and medication of patients on a computer rather than on paper. The basic design of the study was to conduct two usability evaluations of the same EPR system with the same users one year apart in time. The first evaluation took place in May 2002 when the system was being taken into use at the hospital. The second evaluation took place in August 2003 when the users had used the system in their daily work for more than a year.

The Evaluated System
The EPR system used in our study was IBM IPJ 2.3, which is being used at Sygehus Vendsyssel. For the purpose of the usability studies, a test version of IPJ 2.3 was installed in our usability laboratory and configured to match the system used at the hospital

Novice Users
The first usability evaluation involved eight trained nurses. All eight nurses were women, aged between 31 and 54 years, their experience as nurses varied between 2 and 31 years. Prior to the first evaluation they had received between 14 and 30 hours of training in the IPJ system. They characterized themselves as novices or beginners in the IPJ system and in IT in general.

Expert Users
The purpose of the second evaluation was to study the usability of the EPR system after one year of use. Seven of the eight nurses in the first study were able to participate in the second evaluation. A participant with the same characteristics replaced the eighth nurse. Before the second evaluation, all the nurses had used the system in their daily work for about 15 months. They indicated that they on average used the system 10 to 20 times a day, amounting to a total time of use of about 2 hours per day. Therefore, we now characterized them as expert users.

The Two Usability Evaluations
Preparations: Prior to the first evaluation, we visited the hospital and had a number of meetings and discussions with the two persons who trained the nurses in using the IPJ system and dealt with the practical deployment of it. The purpose of this was to understand the work in the hospital wards related to the patient record and get an overview of the system and its parts. Based on this we made a number of overall scenarios for the use of the system and generated realistic test data.

Tasks: The purpose of the usability evaluations was to inquire into the extent to which the IPJ system supports nurses in solving work tasks that are typical for the hospital. Based on our scenarios, we designed three tasks centered on

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

the core purpose of the system such as retrieving information about patients, registering information about treatments, making notes, and entering measurements. The draft tasks were evaluated by nurses responsible for the training program.

Settings: All test sessions were conducted in our usability laboratory using a standard PC with a 19" screen matching the setup at the hospital.

Procedure: The test sessions were based on the think-aloud protocol as described by Rubin [5] and Nielsen [3]. In both evaluations, the eight test sessions were conducted over two days. The order of the nurses was random. One of the authors of this article was test monitor throughout all sixteen test sessions.

Data Analysis: The data analysis was done in August 2004, one year after the second evaluation. The two authors who did not serve as test monitor analyzed all sixteen videos. Each video was given a code that prevented the evaluator from identifying the year and test subject. The videos were assigned to the evaluators in a random and different order.

The evaluators produced two individual lists of usability problems. For each problem in the list there was a precise description. A usability problem was defined as a specific characteristic of the system that prevents task solving, frustrates the user, or is not understood by the user, as defined by Molich [2] and Nielsen [3]. Each evaluator also made a severity assessment of the usability problems as cosmetic, serious or critical [2].

The individual problem lists from the two evaluators were merged into one common list of usability problems. This was done in a negotiation process where the problems were considered one at a time. The evaluators also produced a log file of between two and four pages for each of the sixteen test sessions with the exact times and descriptions of the steps that the user goes through in order to solve each task. The log file also described whether the user solves each task, and to what extent the test monitor provides assistance.

RESULTS

Table 1 summarizes key results of problem identification for the novices and experts. Based on our analysis, we identified a total number of 103 usability problems. The novices experienced 83 of these 103 usability problems whereas the expert subjects experienced 63 of the 103 usability problems and a contingency analysis shows that this difference is significant (χ^2=8.489, *df*=1, *p*=0.0036).

Attributing severity to the identified usability problems, the highest experienced severity for each problem is used. We found that the novice subjects experienced significantly more serious problems than the experts (χ^2=4.296, *df*=1, *p*=0.0382), however no significant differences were found for the critical or cosmetic problems.

	Novice (N=8)	Expert (N=8)	Total (N=16)	χ^2	*p*
Critical	25 (21)	19 (17)	27 (21)	3.068 (2.487)	0.0798 (0.1148)
Serious	45 (29)	34 (23)	56 (32)	4.296 (2.564)	0.0382 (0.1093)
Cosmetic	13 (6)	10 (5)	20 (8)	0.409 (0.291)	0.5224 (0.5896)
All	83 (56)	63 (45)	103 (61)	8.489 (5.752)	0.0036 (0.0165)

Table 1. Total numbers of identified usability problems for the novices and experts. Numbers in parentheses show non-unique problems; problems experienced by at least two subjects.

Out of the total number of 103 usability problems, 64 were identified by both evaluators, 17 only by evaluator 1, and 22 only by evaluator 2. The overlap between problems identified by the two evaluators suggests a low presence of the evaluator effect [1] and thus a high reliability of the merged list of problems.

We also sought to explore differences and similarities in the problems identified by the two sets of subjects. Figure 1 outlines problems unique to the novice subjects, problems unique to the expert subjects, and problems experienced by both novices and experts. 40 of the 103 identified problems were experienced by the novice subjects only and most of these problems concerned simple data entry tasks such as typing in values for patients. 43 of the 103 identified problems were experienced by both novice and expert subjects and they typically concerned advanced data entry or solving judgment questions. 20 problems were identified for experts only and these mainly concern functionality and services that were not applied in the novice sessions, e.g. work task lists for nurses.

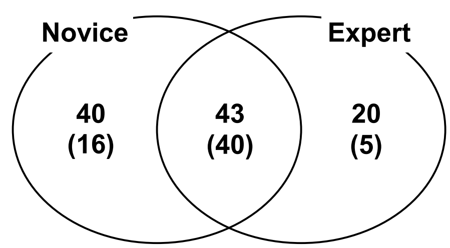


Figure 1. Distribution of the identified problems for the novices and experts. Numbers in parentheses show total numbers of problems subtracted unique problems.

Removing unique problems from the distribution, we see that most of the usability problems (40 of the 61) were identified in both the novice sessions and expert sessions. Further, the experts experienced 5 non-unique problems not experienced by any novice subjects and none of these 5 problems were critical.

INSTANT DATA ANALYSIS

Motivated by the challenges of analyzing usability evaluation data from our studies of Electronic Patient Record systems, we have developed a technique for reducing the efforts spent on analyzing data from usability evaluations: Instant Data Analysis (IDA). The aim of applying this technique is to make it possible to conduct an entire usability evaluation in one single day.

The IDA technique adopts the assumption that identifying the highest number of critical usability problems of a software product can lead to improved quality through redesign. The IDA technique is designed to be combined with the use of the well-established think-aloud protocol for user-based usability evaluations as described in for example [4], [5]. The technique can be applied to both laboratory-based and field-based think-aloud evaluations. The IDA technique exploits the fact that think-aloud usability evaluations typically involve a test monitor and a data logger with high level usability expertise. When conducting the evaluation, the test monitor and the data logger typically gain a strong insight into the evaluated system's key usability problems very quickly. While some of these problems may be captured by taking notes, much of this insight is often lost and needs to be reconstructed during later video data analysis. Rather than loosing this valuable moment of insight, the IDA technique extends the think-aloud sessions with a joint data analysis session.

Procedure

The use of the IDA technique follows immediately after think-aloud usability test sessions. Aiming at conducting the entire evaluation in one single day, 4 to 6 think-aloud sessions should provide a proper foundation for the analysis. During the usability test sessions, the data logger records incidents or problems. This will be used for the later problem identification and categorization. After the think-aloud sessions, the test monitor and the data logger conduct a one hour brainstorming and analysis session. The purpose of this session is to produce a list of usability problems as experienced by the 4 to 6 test subjects.

The roles of the test monitor and data logger during the data analysis are to articulate and discuss the most critical usability problems that they have identified during the think-aloud sessions. Also, they should rate the severity of each problem stating if it is, for example, critical, serious or cosmetic [2]. Assisting the brainstorming and analysis process, the test monitor and data logger may use printed screenshots of the system and notes taken by the data logger during the think-aloud sessions. The aim of the process is not to identify as many usability problems as possible, but to identify the most critical ones.

The analysis session is assisted by a facilitator. The role of the facilitator is to manage the brainstorming and analysis session, asking questions for clarification and writing all identified usability problems on a whiteboard/flip-over as they are presented by the test monitor and data logger. The facilitator should also make sure to keep an overview of the identified problems as the session progresses, categorizing them in themes, avoiding redundancy etc.

After the one hour brainstorm and analysis, the facilitator spends 1-1½ hour on his own writing up the contents of the whiteboard/flip-over into a ranked list of usability problems with short descriptions and clear references the system. Finally, the test monitor, data logger and facilitator run through the problem list together to ensure consensus.

COMPARING IDA TO VIDEO DATA ANALYSIS

We have evaluated the use of the proposed technique for Instant Data Analysis through a usability evaluation of a resource booking system for the healthcare domain.

Participants: The study included five test subjects of between 25 and 64 years of age. They were all staff at Sygehus Vendsyssel with practical experience ranging from 1 year to 37 years. The test subjects had all received training in the booking system. In addition, four trained usability researchers participated in different roles on evaluating the use of the IDA technique. All evaluators had significant previous experience usability evaluations. One researcher acted as test monitor during the test sessions with the five test subjects. A second researcher acted as data logger during the sessions writing down as much as possible during the tests. A third researcher observed the sessions and also logged data for supporting a later video analysis. Finally, a researcher observed the sessions and acted as facilitator in the IDA session.

Settings: The usability evaluation was conducted at the usability laboratory at Aalborg University. From the control room, the data logger could survey the subject room through one-way mirrors and by means of the motorized cameras. During the evaluation, the data logger took notes and created a preliminary log file. From the observation room, two researchers could observe the evaluation through one-way mirrors and on monitors relaying the screen image from the test PC and the cameras.

Procedure: The evaluation was conducted in one day (five hours). The individual sessions were structured by three tasks assignments given to the test subjects one at a time by the test monitor. During the evaluation, the test-subjects were thinking-aloud, explaining their interaction with the system and articulating their comprehension of the design.

Data analysis: The data from the usability evaluation sessions was analyzed independently by two teams of researchers applying a traditional video data analysis technique and the Instant Data Analysis technique respectively.

The Instant Data Analysis (figure 2) produced a list of usability problems ranked as critical, severe or cosmetic with approximately 2 lines of explanation. The total time spent using the traditional Instant Data Analysis technique amounted to 4 man-hours

Figure 2. Instant Data Analysis.

The analysis of the video data followed a standard approach to identifying usability problems. First, the preliminary log-files for each of the five test subjects created during the evaluation sessions were completed by looking through all videos. Following this, the video tapes were then examined thoroughly for identification of usability problems assisted by the log file and each usability problem was described in detail and ranked in relation to its severity.

The Video Data Analysis produced a detailed log file of the five evaluation sessions and a list of usability problems ranked as critical, severe or cosmetic with approximately 5-7 lines of explanation. The total time spent using the traditional Video Data Analysis technique amounted to approximately 40 man-hours.

Following the Instant Data Analysis and the video data analysis, the two lists of usability problems were merged in a collaborative effort. As a part of this, small variations in severity ratings were discussed until consensus had been reached.

FINDINGS

Comparing the IDA results with the results of the video data analysis approach, we found that the latter identified a total of 46 different usability problems where 12 were critical, 15 were serious, and 19 were cosmetic. In total, the two techniques identified a list of 62 different usability problems where 13 were critical problems, 22 were serious problems, and 27 were cosmetic problems.

Considering the identified problems, we found that both approaches assisted in identifying nearly all critical problems, where IDA identified 11 of the 13 (85%) critical usability problems whereas video data analysis identified 12 of the 13 (92%) critical usability problems.

The serious and cosmetic usability problems exhibited a different distribution between the two analysis techniques. Where the IDA technique identified 15 serious problems, a total of 22 serious problems were identified by the two approaches together. Thus, the IDA approach identified 68% of the serious problems found in total. On the other hand, the video data analysis also identified 15 serious problems (68%) meaning that eight serious problems were identified by both approaches. Considering the cosmetic problems, we found that the IDA technique identified 15 of the total 27 problems (56%). The 12 remaining cosmetic problems unidentified by IDA related primarily to specific interaction problems for the subjects typically only experienced by one of the five subjects. A total of 7 out of the 27 cosmetic problems (26%) were identified by both analysis approaches.

A high number of the usability problems identified in the video data analysis approach were experienced by only one subject test subject (26 problems of the total 46). It can be discussed whether these are really problems at all, or if they are noise added to the picture by non-generalizable subjective experiences of interaction with the system. Information about how many test subjects experienced the different usability problems was not included in the problem list generated from the IDA technique. But some of these 26 problems were also identified by the IDA approach. However, the majority of problems experienced only by one single test subject (16 of the 26) were *only* identified in the video data analysis and not in the instant data analysis. Thus, the use of the IDA approach allowed for the omission of a significant part of this noise.

CONCLUSIONS

This paper has reported from two usability studies of IT in the healthcare domain: a longitudinal usability study comparing the usability of an interactive system as experienced by novices and experts and a comparative study of two techniques for analyzing usability data.

In the first study, we observed that novices experienced more usability problems than the expert users of an Electronic Patient Record system. Yet a remarkably high number of problems were experienced both by novices and expert users. These problems were significantly more severe for the novices.

In the second study, we observed that using only 10% of the time required to do video data analysis, Instant Data Analysis helped identify 85% of the critical usability problems in the system being evaluated. At the same time, the noise of unique usability problems was significantly reduced.

REFERENCES

[1] Jacobsen, N.E., Hertzum M. and John, B.E. The Evaluator Effect in Usability Tests. Proc. CHI'98, ACM Press (1998).

[2] Molich, R. *Usable Web Design* (In Danish). Ingeniøren|bøger, (2000).

[3] Nielsen, J. *Usability Engineering.* Morgan Kaufmann, San Diego (1993).

[4] Preece J., Y. Rogers H., Sharp D., Benyon D., Holland S. and Carey T. *Human-Computer Interaction.* Workingham, Addison-Wesley, 1994.

[5] Rubin, J. *Handbook of Usability Testing: How to plan, design and conduct effective tests.* John Wiley & Sons, Inc., New York (1994)

eBag – the digital school bag

Christina Brodersen, Bent G. Christensen, Kaj Grønbæk
Department of Computer Science,
University of Aarhus
Åbogade 34
8200 Århus N, Denmark
+45 8942 5659
sorsha, bentor, kgronbak@daimi.au.dk

Christian Dindler, Ole Sejer Iversen
Department of Information- and Media Sciences,
University of Aarhus
Åbogade 34
8200 Århus N, Denmark
+45 8942 9257
dindler, sejer@imv.au.dk

ABSTRACT
In this paper, we describe the eBag, the individual pupil's repository for digital material as a supplement to the physical school bag. The eBag is a flexible software infrastructure in which children can access their digital material with a seamless login.

Categories and Subject Descriptors
H.5.m, H.3.2, H.3.4, H.3.5 [**Information Systems**]: Information Interfaces and Presentation, Information Storage, Systems and Software, Online Information Services, K.3.1 [**Computing Milieux**]:Computer Uses in Education

General Terms
Design, Experimentation.

Keywords
Mobility, personal, context dependent, group work, project work.

1. INTRODUCTION
8.a on Vestergaardsskolen is busy preparing for the projects they are starting the following week. The general theme of the project week is "Our town", and the pupils have chosen different topics that all deal with how different parts of Aarhus has changed in the past 200 years. Each group is now busy collecting background material for their specific topic and are planning where to go and who to talk to get information for their project presentation. They collect the digital material in a project folder in their electronic school bag, the eBag. They have grouped their individual eBags in a project group and whenever they drag and drop pictures, links or documents into the project folder, it is automatically shared between them. The following week, the pupils visit different parts of town, interview museum custodians, business vendors and wharf managers, take pictures of the streets, the harbour and buildings as they look today and can compare them with photos they found in advance in books at the library and scanned. Photos taken with the small cameras on their mobile phones can be saved directly from the phone into their eBag; so can audio and video recordings made with the phone. Later in the week, they assemble around a large, digital display and start putting the presentation together; combining audio recordings from the interviews with small video clips and photos and search the internet for additional information about the buildings they have seen and the people who live there now to compare with the information they have gathered about life in the city 200 years ago. They decide to visit one more place before they finish their presentation. At the end of the week, they have collected enough material to make a great presentation of the town as it is today compared to 200 years ago.

With the current advances in wireless and mobile technology as well as large displays, this type of scenario is no longer science fiction. With this paper, we introduce the eBag: a digital counterpart to each pupil's physical school bag. The eBag serves as a link between different types of displays, through which its contents can be accessed, and allows the pupils to collect, carry, access, and share digital information very easily.

2. eBag
The eBag is a personal, digital repository in which you can place pictures, video, music, text documents and other digital material for use in and outside of school. In the following a description of the eBag architecture is provided.

2.1 Architecture
The eBag application is developed on top of the HyCon framework and architecture [3]. The HyCon framework and architecture for context-aware mobile hypermedia was developed to provide a general platform suited for experiments with hypermedia mechanisms in a context-aware and mobile environment as described in detail in [3]. The logical layers and infrastructure of the architecture are seen in Figure 1, with four layers divided into Storage, Server, Terminal, and Sensor. The bottom layer, the Storage layer, handles persistent storage and retrieval of hypermedia structures produced in the system. The Server layer includes components handling annotation, link, location, and Search functionality. The functionalities of these components are offered through services implementing the Service Interface, which are realised as Web Services and Java servlets. Through these interfaces applications in the Terminal layer communicate with the services in the Server layer. Applications in the Terminal layer are not limited to a specific platform, but may be running on a variety of hardware platforms and software environments (phones, tablets, laptops, Web browsers). The key property of the HyCon framework is the last layer, the Sensor layer. This layer is introduced to logically group all sensors deployed to obtain contextual information.

The eBag service is realised by components in all four layers of the HyCon framework and architecture. Starting from the bottom, the data model of the storage layer is extended by an eBag data type for profile information of an eBag. For the actual content of an eBag such as digital pictures, video, music, text documents, a webdav server is used, providing versioning and locking mechanisms for the digital media stored. In the Server layer, an eBag component encapsulates all functionality for creating new eBags or modifying existing ones in the persistent storage. This functionality can be used from applications in the Terminal layer. These applications can be running on various technical platforms, affording different kinds of interaction with the eBag application. The platforms and display types that we have currently used range from tabletPC, laptop, and desktop computers with screen display, SmartBoard display, and floor projection. The eBag service is depending heavily on the Bluetooth sensor component in the sensor layer. This component continuously makes inquiries for nearby Bluetooth units and offer this information to the eBag application to react upon.

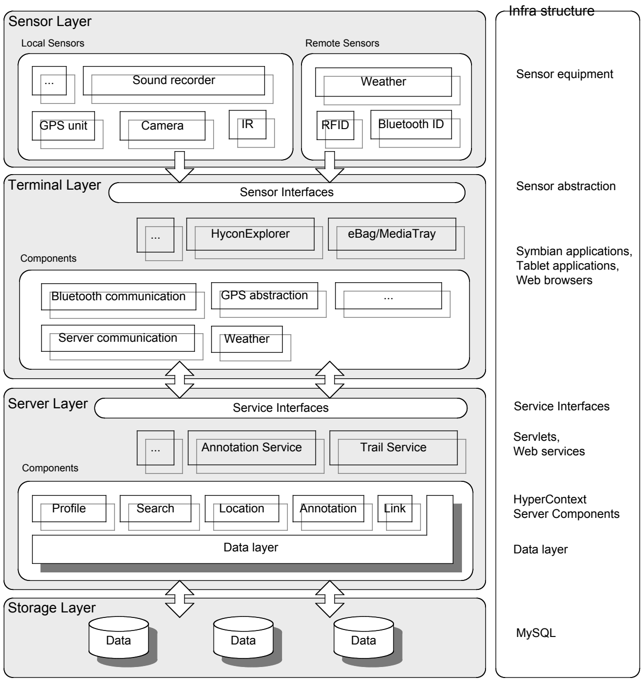


Figure 1: The HyCon service framework architecture. The framework is divided into four layers: the Storage layer, the Server layer, the Terminal layer, and at the top the Sensor layer providing sensed contextutal information for the terminal layer applications

2.2 Proximity based login/logout

A key feature of our sensor hardware configurations is that the Bluetooth sensor equipment (in our case: BlipNodes[1]) allow the user to control the signal power for the inquiry. Thus, the signal power can be regulated to reach a certain distance, e.g. two meter, and the Bluetooth units will only be discovered if they are within this proximity.

[1] http://www.blipsystems.com

We use this in the eBag service to provide a login (and logout) mechanism for the eBag. Each pupil is given a Bluetooth cellular phone, and the Bluetooth id of the phone is registered with the pupil's eBag profile. When a registered phone of a pupil is discovered by one of the Bluetooth sensors, the corresponding eBag icon is presented on a display nearby of a computer running the eBag service. Proximity-based login has many advantages, particularly in highly mobile situations [1] and it makes interaction with a changing range of computers more natural and direct and much less cumbersome. An eBag icon presents the name of the eBag profile and a picture of the owner. However, these two pieces of information can be changed by the pupils. Examples of two eBag icons is depicted in Figure 2.

Figure 2: Two eBag icons as presented when a phone is discovered by the Bluetooth sensors. At the top of the icon the eBag profile name is shown, and below a picture of the owner is presented.

When the pupil activates the icon by double clicking it, the eBag opens and the pupil will have access to her material. Figure 3 shows an open eBag.

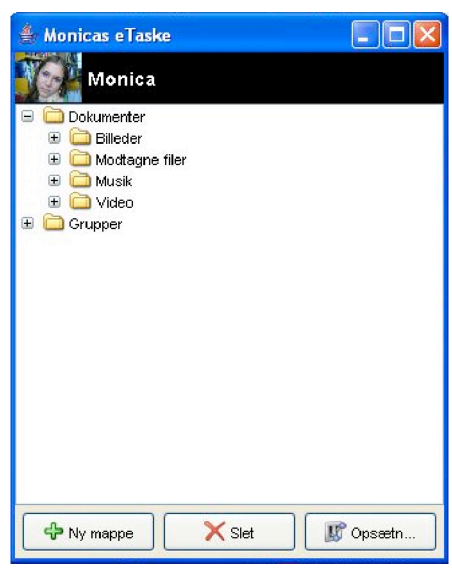


Figure 3: An opened eBag

The eBag provides a file hierarchy to organise the pupils' material. Files such as pictures, video, music, and text documents can be dragged freely back and forth from the eBag and the computer where the eBag is opened. This allows the pupil to e.g. write a document for an assignment and to put the document into her eBag when finished. This way the pupil have safely stored the produced material and may easily fetch the documents from her eBag later on, possibly on a different location.

2.3 Collaboration support

The eBag supports project work by allowing eBags to form groups in which the distribution and sharing of material becomes very easy. By dragging two or more eBag icons close together, a group is formed. When doing this, a group name is propted for and a shared group folder is created, which all the members of the group have access to from their eBags. The eBag icons are now assembled into a group icon, where each eBag icon is included. An example of this can be seen in Figure 4.

Figure 4: The group foo2 is created and consists of netkids4 and Netkids1

Besides dragging files the to opened eBag, the eBag application also allows the pupils to drag their files to the eBag icons directly. This will copy the files to an "incoming files" folder in the eBag. The same is possible for the group icon, however when files are dragged over to the group icon they will be copied to the shared group folder instead.

Figure 5: Pupils at Vestergårdsskolen, using the eBag application on a large wall-display

3. RELATED WORK

In the following, we take a look at other, similar eBag concepts.

The *school*Bag is Uni-C's proposal for an electronic school bag. It is a web-based service that provides each pupil with space on a web server (webdisc) for the files he or she wishes to save along with notes, links and addresses. While the *school*Bag is distributed because it is web-based, it is not mobile unlike the eBag that allows the pupils to bring their digital context with them as they move around on the school or go on field trips. Uni-C's vision for an electronic school bag does include aspects of mobility: however, it is largely based on providing every pupil with a notebook computer that they can carry everywhere and access the digital material: "In Jens' class every pupil has received a laptop computer. Where the school bag used to be a heavy thing dangling from the back when Jens hurried to the train to Århus, he now only needs the laptop, and, of course, his lunch!" [6] It is, unfortunately, a very popular view in the Danish school system that if you equip each student with a laptop computer, all will be well in the world; pupils will be more productive because they can work everywhere and they will do so happily. As the writers prophetically announce: "The school bag as we know it is long gone. It has been replaced with a virtual version." [6] We find this vision highly questionable because it is so PC-centric that it fails to even consider the range of mobile technologies that are available, e.g. mobile phones that become increasingly powerful and are already an integrated part of most pupils' social lives.

Another example of a digital school bag is provided by the Norwegian system, Fronter which is used in several Danish high schools[2]. Fronter is a web-based conference system not unlike First-Class (which also forms the basis for Uni-C's school portal, SkoleKom). With Fronter, the teacher can create a virtual building which is separated into rooms in which pupils and teachers can collaborate with different tools on different subjects. Fonter is basically an administrative system that allows the teacher to e.g. open and lock rooms or folders in which pupils hand in their assignments. If the paper has not been uploaded on time, the teacher will know. However, Fronter does not support differentiated teaching and it cannot support the social structures that are essential to the pupils in order to function well as a virtual community. It may use the metaphor of building a house, but it is a space rather than a place [5].

An electronic schoolbag is also an important component in building ad hoc classrooms in [4]. Their eSchoolbag consists of a number of components, e.g. the electronic book, parents' contact book, pencil case and can be accessed through a notebook computer or PDA. In their example, they present a vision for school work where everything is done through the PDA, from reading and doing exercise to communicating with the pupil's parents. While we support the idea of making digital information accessible outside of the classroom and on the move, we want to emphasise that supplying the pupils and teachers with a range technologies to support school work rather than replacing one technology for another is a crucial point if we are to succeed in supporting education and learning. Furthermore, we find the assumption that the eSchoolbag can and should be shared between the pupil and his or her parents (parent check the e-contact book on the PDA) to be a violation of the child's privacy.

Finally, with CoolTown [1] Hewlett-Packard presents their vision for pervasive computing where homes, schools, workplaces and life in general is digitally enhanced and connected through the world wide web. Small, mobile devices provide fast access to e.g.

[2] See http://fronter.dk

presentation material (the CoolTown video[3] shows a business woman "beaming" her presentation material up on a large, digital wall display) and makes studying fun (represented in the video by a little girl who is learning Spanish through a voice-activated program she has downloaded to her wrist watch). However, the CoolTown vision also assumes a lack of privacy that is quite disturbing, e.g. exemplified by the same little girl who uses a public display on the refrigerator to receive a message from her teacher and consequently gets her test scores broadcasted on it, or Bob who is loudly and repeatedly congratulated by the technology he comes into contact with because he has received a reward. The vision also lacks a fundamental technical soundness that makes the underlying infrastructure for the "wired world" they describe believable. However, we agree with the basic idea of being able to move and access digital material through a number of heterogeneous technologies that allow users to choose the tool they find most useful in the given context; a vision quite similar to Mark Weiser's concept of ubiquitous computing [7].

4. FUTURE WORK

In its first implementation, the eBag is "tied" to a mobile phone whose Bluetooth unit serves as a unique identifier for each pupil's eBag. However, it can easily be associated with other types of Bluetooth carriers (e.g. BlueTags) and shaped like key-chains, buttons, toys, or whatever form is most appropriate for the children in question. There are, however, great advantages of using the mobile phone when working with older children. First, most children in the 7th to 10th grade already carry a mobile phone and are often very attached to it to support their social networks. Second, a variety of mobile phones offer cameras, audio recorder and video recorder functionality that provide valuable means of capturing data e.g. during project work.

We are currently in the process of dramatically re-designing the eBag, particularly with respect to the user interface, information structure and navigation, but also enhancing the functionality. From our preliminary tests with pupils from Vestergaardskolen, we see a need to provide the pupils with the possibility to personalise their eBag beyond changing name and picture, e.g. supporting the social structures by displaying their friend's pictures and eBags, and making it possible to showcase themselves to others through a gallery and a blog-type message board.

We are also developing the context dependencies with respect to groups to include class/teacher- and class/school relations. Thus when the physics teacher wants to distribute a new project folder about Ohm's law to the entire class an Ohm's Law sub-folder is created in the individual pupil's physics folder. The teacher can also add additional material for the weaker students and more challenging assignments for the stronger students and thus support a differentiated teaching strategy. Naturally, the teacher has no access rights to the eBag in general but it is vital that she can add material that is placed in the right context in order to avoid material getting lost in a general "inbox". The pupil can grant access to the teacher and other students for different folders in different contexts.

Finally, we are supporting a user-controlled awareness service on the eBag that allows pupils to change their status from the mobile phone. A pupil can e.g. add a status message saying "Taking the bus back to the school at 11 o'clock" or "Visiting The Old Town – going to the town hall at 10". This is particularly useful on field trips as a means of asynchronous communication with other group members or teachers at home.

5. ACKNOWLEDGMENTS

This work has been supported by Center for Interactive Spaces, ISIS Katrinebjerg (Project #110). We are furthermore indebted to the pupils and teachers at Vestergårdsskolen in Århus for their valuable evaluation and design ideas for the first version of the eBag.

6. REFERENCES

[1] Bardram. J. E. (2003*). The Trouble with Login – On usability and Computer Security in Pervasive Computing. Technical Report CfPC 2003–PB–50*, Centre for Pervasive Computing, Aarhus, Denmark, 2003. Available from http://www.pervasive.dk/publications.

[2] Barton, J. & Kindberg, T. (2001). *The Cooltown User Experience. Tech Report: HPL-2001-22*. HP Labs. http://www.hpl.hp.com/techreports/2001/HPL-2001-22.html

[3] N. O. Bouvin, B. G. Christensen, K. Grønbæk, and F. A. Hansen (2003) HyCon: A framework for context-aware mobile hypermedia. In *The New Review of Hypermedia and Multimedia*, 9, 2003

[4] Chang, C, & Sheu, J. (2002). Design and Implementation of Ad Hoc Classroom and eSchoolbag Systems for Ubiquitous Learning. In *Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'02)*, 2002, pp. 8 - 14

[5] Harrison, S. and Dourish, P. Re-place-ing space: the roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work (CSCW'96)*, ACM Press, New York, NY, pp. 67 - 76

[6] Holmboe, P & Jessen, P. K. (2003). Skoleveje : inspirationsmateriale til SkoleKom., *Uni-C teaching material #37.134*, http://3-klasse.dk/sider_uden_menu/materialeliste.htm, 2003, s. 114 – 126

[7] Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 1991, pp. 933–940.

[3] See: http://www.cooltown.com/cooltown/cooltown-video.asp

The Automatic Pool Trainer - a Platform for Experiments with Multi Modal User Interaction

Lars Bo Larsen

SMC - Speech and MultiMedia Communication
Dept. of Communication Technology
Aalborg University, Denmark
Phone: +45 9635 8635
Email: lbl@kom.aau.dk

ABSTRACT

This paper presents The Automatic Pool Trainer (APT). The APT serves a purpose as a platform for research of multi modal user interaction using computer vision and speech recognition, agent technologies, etc. However, it is also used for experiments and student projects at the Intelligent Multimedia (IMM) Programme. The paper documents the development of the platform through several iterations. The technological problems involved are briefly discussed.

Keywords

Human Factors. ACM categories: H.5 Information Interfaces and Presentation, in particular H.5.2 User Interfaces.

Multi modal user interaction, speech and vision processing, integration of modalities, computer aided learning. HCI, evaluation.

1. INTRODUCTION

The Automatic Pool Trainer (APT) has been developed over the last five years by the SMC group in close collaboration with a number of students at the IMM (Intelligent MultiMedia) master's programme. Its aim is to research and demonstrate multi modal user interaction as well as serve as a pedagogical platform for students. This has to a large degree been successful.

The pool training system is based on a platform developed within a project carried out at the Center for Person-Kommunikation at Aalborg University. In this platform a number of hardware and software modules were integrated into an open architecture to provide the "IntelliMedia WorkBench" [1]. The intention of the workbench was to provide a basic set-up, in which new interaction paradigms can be investigated. The Automatic Pool Training system can be viewed as such an application. The Automatic Pool Training (APT) system is based on Target Pool [3], a widely used scheme for pool training, developed by the professional pool player Kim Davenport. The purpose of Target Pool is to enable a trainee to follow a self-study course consisting of a number of exercises and evaluating his progress over time.

The APT has existed in a number of different versions ([2],[4],[5],[6],[8],[9]). The following sections present the original Target Pool training scheme and an overall description of the APT. This is followed by a closer account of the system architecture and the individual components in greater detail. The design of the user interaction is presented and discussed, and results of user tests performed on the APT are given.

The paper presents the current version of the APT dubbed Intellipool$_3$ [10] (as of autumn 2004) and gives a brief history of the development leading up to this. A more detailed description of the Pool Trainer together with background, images, video clips and a bibliography can be found at: http://cpk.auc.dk/SMC/pooltrainer/.

1.1. Target Pool

The basic idea of Target Pool is to present a trainee with a number of exercises (accompanied with detailed instructions). The trainee records and evaluates his or hers performance after each shot on a score sheet. Based on the performance a new exercise (or the same again) is recommended. The only equipment needed apart from the pool table and cues, etc., are a booklet describing the exercises, a score-board, and a thin cloth with a printed target, to be placed on the pool table. Usually only the cue ball and one other ball (the object ball) are used. The training programme consists of a number of exercises. The objectives of all the exercises are to shoot the target ball into a specified pocket and position the cue ball in a specific position (denoted the target). The exercises differ in the topic they address and the level of skill required. In total more than 130 exercises are grouped into 10 courses (only a few has been implemented into the APT). Figure 1 shows an example of an exercise description for Target Pool.

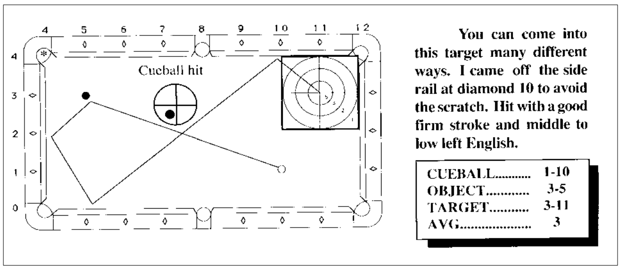


Figure 1. Target Pool exercise [3]

1.2. The Automatic Pool Trainer

In most configurations, the Pool Trainer features spoken in- and output, as well as visual detection of the pool balls. The setup consists of a pool table, cues and balls, etc. In the ceiling above the pool table a projector (a laser in earlier versions) and a camera are mounted. The projector displays shapes (e.g. lines and targets, menus etc.) on the surface of the table, and the camera captures the positions and movements of the balls, as well as "virtual button-presses". A large projector screen is mounted on

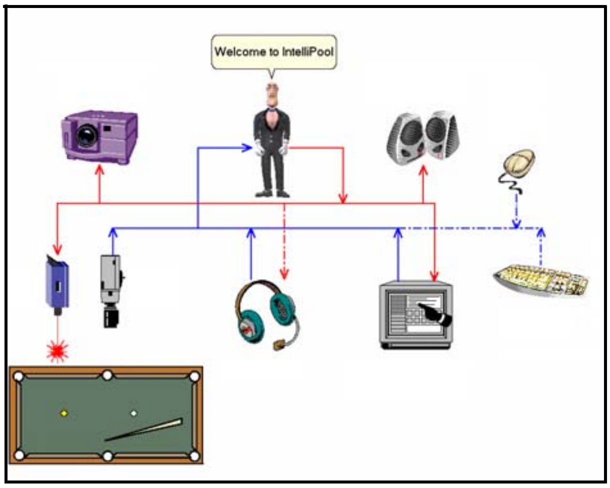


Figure 2. Overview of input and output devices in Intellipool$_2$ [6]

one wall of the room and has been used to show instructions to the users, replays of the users shot, evaluations of the performance, etc. Apart from the users actions directly at the pool table, the system is addressed either by voice or via a touch screen. In some versions (Intellipool$_2$ see [4],[6],[8]), the user communicates by voice with an animated interface agent "James", see Figure 4. James accepts spoken commands and uses synthetic speech to give instructions. He is capable of moving around on the screen and e.g. point to various objects.

Figure 2 above shows a rich picture of the interaction devices. The headset symbolises speech input. Keyboard and mouse are only used in special cases (apart from development and debugging), such as logging in and setting up a user account.

2. The Development of the APT

This section gives a brief account of the development of the Automatic Pool Trainer and discusses the technologies used in the user interface.

- The first version of the pool trainer used computer vision and speech inputs and laser and speech for output. Although it envisioned many of the current features, the functionality was focused on drawing guidelines of a potential shot based on the direction of the cue and location of the balls. This notion was later dropped, because there was no way to anticipate the spin, draw, etc. that the user might apply to the cue-ball.
- Instead a new version based on a number of predefined exercises (Target Pool) was developed [4]. The basic ideas of this design is still the predominant ones. This version used a blackboard architecture for module communication and an animated agent "Q" for spoken user interaction. See Figure 3 below. It also introduced the

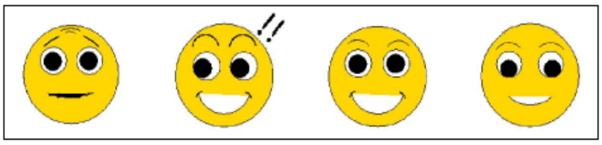


Figure 3. The first interface agent "Q" was capable of expressing a range of moods, depending on the users' performance. Q's eyegaze was used to turn the users attention towards various objects on the screen ([4]).

Target Pool concept. This version was based on two Linux-based Pcs and the speech recogniser was implemented on a SUN workstation.
- Later, the software was rewritten into Java (except for the image processing, which were implemented in C) and more tightly integrated and the agent and blackboard was removed. Functionality for selecting exercises using a dynamically changing "graph" was designed, together with elaborate help and user management system [5]. The IBM ViaVoice speech recogniser for spoken interaction, replaced the SUN based one, and all modules (except for the laser controller) were integrated on a single PC.
- Based on user studies, the speech recognition and -synthesis and the graphical user interface were revised. Again, an animated interface agent ("James") were integrated into the user interface. The agent James is provided by Cantoche [7] and based on the Microsoft Agents technology [6],[8].

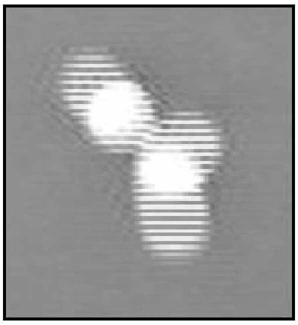
Figure 4. The interface agent "James" [7]

- In order to identify the users' errors and provide valuable feedback, a Causal Probabilistic Network (CPN), known mostly from medical decision support systems, has been designed and tested [9]. The CPN is currently able to identify errors in aim, strength and spin of the stroke by analysing the movements of the balls. Feedback is provided through either a graphical representation of the probabilities that a certain type of error has occurred (or indeed not occurred) or as a verbal evaluation

2.1. Elements of the User Interface

The APT has been through a number of iterations, where experiments with various user interface configurations have been tried out. However, the basic idea in all the configurations has been to allow the user as much freedom of movement as possible. Since pool is a "hands busy - eyes busy" activity, and the user must be able to move freely around the table, it poses some interesting challenges to the design. An additional requirement is that any ordinary pool table, cues, pool balls, etc. must be used. This has ruled out the use of the use of e.g. specialised localisation sensors in the table, pool balls, etc. Instead, we have used a computer-vision based solution, with a camera mounted above the table in all versions.

The camera is used to identify various events on the table, such as the start- and ending times of a shot, whether the user has placed the balls in the correct positions, tracking the balls during a shot and, in the current version, the location of the users hand, as s/he navigates a menu using "virtual buttons" on the table. Some examples of the images captured by the camera are shown Figure 5 below.



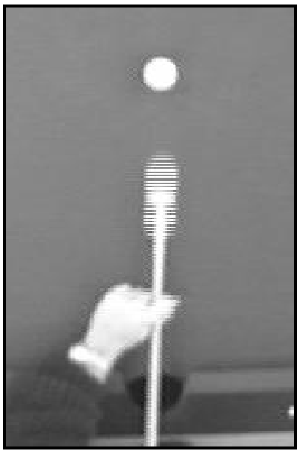


Figure 5. Illustration of some of the problems with separating objects on the table. The "striped" patterns is an effect of interlacing [9]

In general, it is possible to determine the position of the stationary pool balls with an accuracy of 1-2 cm. This is quite sufficient for the current application and is not a challenge for a standard PC at a frame rate of 25 fps. However, determining the position of moving balls accurately poses some additional problems [9]. Likewise, the use of computer vision techniques has made the APT very sensitive to light conditions and we have had difficulties when setting it up outside the lab.

Because the user has the attention focused on the pool table, the natural choice of graphical output would be to directly display the information here. We have employed a laser device for this, as illustrated in Figure 6.

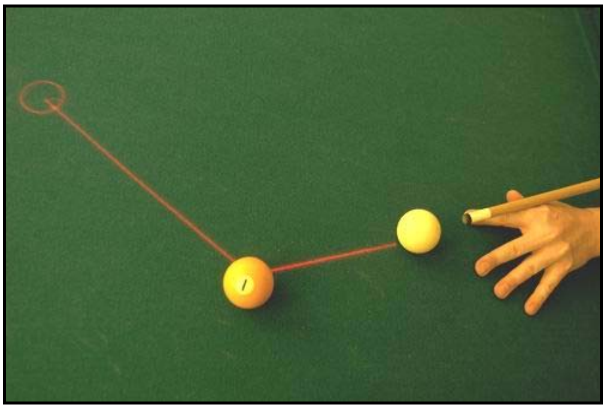


Figure 6. The laser beam is used to "draw" lines and mark points directly on the surface of the pool table [6].

This has shown to be very effective, as users intuitively understand that they are expected to place the balls in various marked positions and aim following the lines drawn on the table. The laser is capable of drawing 30.000 points per second [1], and the figures drawn on the table are stable and pleasant to look at. One drawback has been that the laser is too weak for very long routes. Figure 7 shows the two small mirrors controlling the laser beam (left) and the target drawn by the laser.

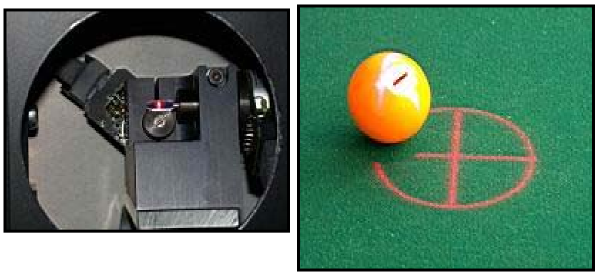


Figure 7 The picture on the left shows the two small mirrors controlling the laser beam. On the right a target has been drawn with the laser

In addition to the laser, graphical output has been provided by a projector and a wall-mounted screen. Information, such as instructions of how to set up an exercise, where to hit the cue ball, navigation of menus, playback of shots (recorded by the camera above the table), keeping track of exercises and user feedback is presented on the screen. In many cases, this has been done via the interface agent described above.

Figure 8. Player wearing a wireless microphone with the wall-screen in the background.

However, to allow the user sufficient freedom of movement, using mouse and keyboard for input is considered impractical. Instead, the user wears a wireless microphone and issues commands verbally. We have tried this in a number of setups and user experiments and our conclusions are that the commands must be very short and precise, to avoid that the speech recogniser accepts "false" commands (e.g. speech addressed to other players). We have used the interface agent as a "personification" of the pool system, as most finds it unnatural to speak to a pool table or screen. Figure 8 shows a player with a microphone and the screen in the background. A consequence of using speech as input is that the user does not have to look at the screen when issuing commands. This means that s/he can keep the attention on the table, since the agent will also respond verbally, with synthetic speech.

3. The current version of the APT

The versions described in the previous section were mostly developed with the aim to research multi modal user interaction, and a number of technologies have been introduced and subjected to user evaluation. However, based on our experiences with this, we found that an interesting problem would be to aim at a streamlined, “minimalistic” user interface, with as few devices and modalities as possible.

A clear conclusion from the previous results was that the use of computer vision to capture the status of the pool table, and the capability of drawing directly on the surface of the pool table had been highly successful. However, the laser suffers from a number of limitations. It requires a dedicated controller PC, the interface to the remaining system is slow, but most importantly, it is only capable of drawing simple figures, such as lines and circles. Based on these observations, it was decided to concentrate the user interaction solely on the pool table, and consequently remove the screen and speech interaction. To do this, the laser was replaced by a projector, mounted in the ceiling next to the camera and pointing downwards [10]. This greatly improved the graphical display on the table, as shown on Figure 9

Figure 9. The pool table with a projected target and text

However, there still exists the problem of e.g. switching between exercises and otherwise control the system. This has been solved by projecting a menu on the pool table, and introduce four “virtual buttons”, which the user “presses” by placing her hand above it for a few seconds (Figure 10).

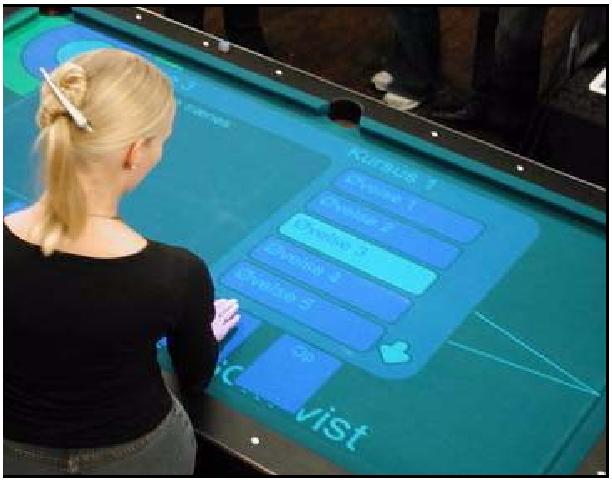


Figure 10. The menu and virtual buttons

While quite effective and robust, this takes a little time to get used to for most users. The menu is launched by covering two of the diamonds for five seconds [10].

4. Conclusions

The APT has been subjected to a number of user tests, with novice users, as well as experienced players and instructors, who all been favourable in their evaluation. In September of 2002 and 2004, the APT was a part of Aalborg University’s exhibition for the Danish Science Festival. Our experience is that nearly all users immediately capture the idea of the pool trainer and there is only a need for minimal instruction (especially for kids). With the introduction of the projector, the APT is now built solely from consumer “off-the-shelf” components, making it commercially attractive, either as a product or a an Open Source project.

The pool trainer has shown itself to be the source of a number of challenging problems, both in terms of the technological issues involved, but it has also proved to be very interesting from the point of view of multi modal user interface design. As such it has served its purpose. To this day more than 25 master students at the IMM have worked on it, and numerous reports, theses and papers have resulted from the project.

Our plans for the future include a reintroduction of the interface agent as an alternative or supplement to the menu. Given time, the agent might be displayed in 3D moving about on the pool table.

5. ACKNOWLEDGEMENTS

The author wishes to thank all the students, who have enthusiastically contributed to the Pool Trainer. We have gratefully received financial support from The Faculty of Engineering and Science, Aalborg University

6. REFERENCES

[1] Brøndsted, T., Dalsgaard, P., Larsen, L.B., Manthey, M., Mc Kevitt, P., Moeslund, T., Olesen, K. "A platform for developing Intelligent MultiMedia Applications", Technical Report R-98 1004,May, 1998, CPK, Aalborg University

[2] Buch, J. et al: "Intelligent Multimedia Based Pool Trainer”. Report, IMM June 1998

[3] Davenport, K.: “Target Pool”, Target Pool Productions, P.O Box 219, Marysville, Michigan, 48040, 1992

[4] Bondesen, P., Poulsen, P., Lykkegaard M., “SmartPool – A multi modal pool training system”, Aalborg University, June 1999.

[5] Larsen, L. B., P. M. Jensen, K. Kammersgaard, L. Kromann,"The Automated Pool Trainer - A multi modal system for learning the game of Pool", Proceedings of ICI-MADE'01 - Intl. Conf. on Intelligent Multimedia and Distance Education, June, 2001

[6] Jensen, M. and W. K. Vodzi: "Adding speech and a software agent to the Intelligent Pool Trainer" Thesis Report, Aalborg University, Aalborg, June 2002.

[7] Cantoche, 1999. http://www.cantoche.com.

[8] Larsen, L. B., M. D. Jensen, W. K. Vodzi, "Multi Modal User Interaction in an Automatic Pool Trainer", Fourth IEEE International Conference on Multimodal Interfaces (ICMI 2002), October, 2002 Pittsburgh pp 361-366.

[9] Chaib,D.,Y. Pannet,A. Tseveen-Ochir: "User error diagnosis in the Automatic Pool-trainer system", Report, IMM January 2004, Aalborg University

[10] K. L. Jensen R. B. Jensen, I. Kun, S. Larsen: “Intellipool$_3$” Report, IMM June 2004, Aalborg University

Remembering Multiple Passwords by Way of Minimal-Feedback Hints: Replication and Further Analysis

Morten Hertzum
Computer Science
Roskilde University
Roskilde, Denmark
+45 4677 3077
mhz@ruc.dk

ABSTRACT

Passwords are a prominent mechanism for user authentication but entail a conflict between ease of use and security in that passwords must be both easy to remember for the password holder and difficult to guess for everybody else. To support users in remembering their passwords minimal-feedback hints for remote authentication (MiFA) provide users with a couple of the password characters when users are prompted for their password. In this study MiFA hints, originally devised by Lu and Twidale (2003), were evaluated by having 14 participants create five passwords each and prompting them for these passwords after one week and after four weeks. With the aid of MiFA hints participants remembered significantly more passwords and were significantly more confident in the correctness of their memory of their passwords than without hints. However, many of the passwords created by the participants were weak, for example a word followed by one or more digits, and vulnerable to dictionary attacks.

Categories and Subject Descriptors

K6.5 [**Management of computing and information systems**]: security and protection – *authentication*

General Terms

Experimentation, Security, Human Factors.

Keywords

Security, ease of use, user authentication, passwords, MiFA.

1. INTRODUCTION

Passwords are a widely used mechanism for user authentication and thus critical to the security of many systems [1, 11, 12]. To provide effective security, passwords should be known to the password holder but remain unknown to everybody else. While passwords such as personal information and real words are relatively easy for a user to remember they are weak from a security point of view because they are vulnerable to informed guessing and dictionary attacks. Strong passwords (e.g., b5j#Kv!8N) are less vulnerable to attack but at the same time more difficult to remember. However, the sheer number of passwords people must have to accomplish their day-to-day activities exceeds most humans' capacity for remembering meaningless strings of characters [1].

Lu and Twidale [8] suggested a technique, minimal-feedback hints for remote authentication (MiFA), for supporting users in remembering their passwords. The distinguishing characteristic of MiFA is to aid users' memory by providing them with a couple of the password characters when prompted for the password. This study replicates the study by Lu and Twidale [8] to assess whether MiFA hints (1) aid users' memory of their passwords, (2) make users more confident in the correctness of their memory of their passwords, and (3) coincide with password-creation strategies that yield strong passwords.

2. MINIMAL-FEEDBACK HINTS

Most user-authentication mechanisms present users with a blank password field and leave it entirely to users to be able to remember their passwords. MiFA hints introduce minimal feedback with the thinking that "a few carefully revealed hints will jog an authorized user's memory, but will be of insufficient help to an unauthorized user who does not know the password in the first place" [8].

At the time of password creation, users will select which characters from their newly created password should be provided as hint. Then, the password with the hint characters revealed and all other characters replaced by, for example, asterisks is converted to an image and slightly distorted. This conversion and distortion is done to provide additional security against password-cracking software.

At login, the image is presented to users who will be able to determine which of their passwords must be the right one for this particular account, or to narrow down the set of likely passwords based on the hint. Lu and Twidale's [8] exploratory study indicates that with MiFA hints users remember 75% of their passwords correctly in their first attempt. Their study involved five users who were prompted for their passwords ten days after creating them.

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

3. METHOD OF REPLICATION STUDY

The participants in this study were 16 graduate students in Computer Science, seven female and nine male. Their age ranged from 21 to 42 years with a median of 24 years. All participants had years of experience creating and remembering their personal passwords.

During an initial session the participants were asked to create passwords for five imagined web services: a bank, a book store, a university portal, an email account, and a travel site. Participants were provided with a form for writing down their passwords. This form also summarized the instructions given to participants before the form was handed out. Participants were instructed that the passwords had to be at least eight characters long and include both letters and digits, as is common practice in many systems, and that the five passwords had to be different but were allowed to be related. Further, participants were instructed not to use any of their real passwords but asked to use their normal strategies for creating passwords. Participants were also told that each password could have one, two, or three hint characters, which would be revealed to them when they were subsequently prompted for the password. Then, participants were asked to choose hint characters for each of their passwords and indicate them by underlining. Finally, participants were asked to write down any strategies they had applied in trying to create passwords that were memorizable.

On two subsequent occasions participants were prompted for their passwords. These password-prompting sessions took place one week and four weeks after the participants had created their passwords, and they proceeded in the same way. First, participants were asked to provide their passwords without the support of hints. For each password participants were also asked to indicate on a five-point scale from very uncertain to very certain how confident they were that they remembered the password correctly. Second, participants received their MiFA hints and were again asked to provide their passwords and indicate their confidence in the correctness of the provided passwords. The MiFA hints were provided to participants in a format that gave the hint characters and their approximate position in the passwords. Examples:

___ b ________ 8

__ a s _________

In the first example the hint characters are 'b' and '8'. The 'b' is preceded by one or more characters but is closer to the beginning than to the end of the password. The '8' is the final character of the password. In the second example the hint characters are 'a' and 's', and they appear next to each other near the beginning of the password.

Participants spent about 20 minutes creating their passwords and about 10 minutes on each of the two sessions where they were prompted for their passwords. Two participants failed to create passwords that were at least eight characters long and contained both letters and digits. These participants were excluded from the data analysis, leaving 14 participants.

4. FINDINGS

Table 1 shows the participants' ability to remember the five passwords one week after creating them and four weeks after creating them. After one week, participants remembered an average of 4.10 passwords with the aid of hints. This is 1.70 passwords more than participants remembered without hints, a significant improvement (T-test, p=0.022). After four weeks,

Participant	Remembered after 1 week		Remembered after 4 weeks	
	No hints	Hints	No hints	Hints
P1	5	5	5	5
P2	2	5	0	5
P3	5	5	0	5
P4	-	-	0	5
P5	0	4	4	4
P6	-	-	4	4
P7	-	-	2	3
P8	-	-	0	3
P9	1	4	2	2
P10	2	4	2	1
P11	0	0	0	0
P12	5	5	-	-
P13	0	5	-	-
P14	4	4	-	-
Average	2.40	4.10	1.73	3.36

Table 1. Number of passwords remembered by the participants ('-' indicates a missing value). Each participant was prompted for five passwords one week after creating them and four weeks after creating them.

Participant	Confidence after 1 week		Confidence after 4 weeks	
	No hints	Hints	No hints	Hints
P1	4.8	5	5	5
P2	1	3.6	1	3.4
P3	5	5	1	5
P4	-	-	1.6	4.4
P5	4	4.2	4.2	4.2
P6	-	-	3	4.2
P7	-	-	3	3.6
P8	-	-	1	2.8
P9	2	5	4	4
P10	4.8	4.8	5	5
P11	1.4	2.2	1	2.8
P12	5	5	-	-
P13	2	4.2	-	-
P14	-	-	-	-
Average	3.33	4.33	2.71	4.04

Table 2. Participants' confidence in their memory of their passwords ('-' indicates a missing value). Each number is the average of five password-confidence scores on a scale from 1 (very uncertain) to 5 (very certain).

participants remembered an average of 1.63 passwords more with hints than without hints, again a significant improvement (T-test, p=0.046). Between the first and the fourth week participants' ability to remember their passwords decayed by 0.67 for recall without hints and by 0.74 for recall with hints. This decay is, however, not significant (T-test, no hints: p=0.793, hints: p=0.182).

Table 2 shows the participants' confidence in the correctness of their memory of their passwords. After one week, participants' confidence averaged 4.33 when they recalled passwords with the aid of hints. This is 1.00 more than their confidence without hints, a significant difference (Wilcoxon test, p=0.028). After four weeks, participants were an average of 1.33 more confident in the correctness of passwords recalled with hints than without hints, again a significant difference (Wilcoxon test, p=0.018). Between the first and the fourth week participants' confidence dropped slightly but not significantly (Wilcoxon test, no hints: p=0.917, hints: p=0.855).

Password-creation strategies were provided by 12 of the 14 participants. The contents of these strategies included:

- Eight participants' passwords consisted of a sequence of letters followed by one or more digits. For six of the participants this property was an explicit part of their password-creation strategies. On average, 3.07 of the five passwords created by each participant had this property.
- Six participants created only passwords that were the minimum of eight characters long, and four participants explicitly mentioned this as one of their password-creation strategies. On average, 3.29 of the passwords created by each participant were eight characters long.
- Five participants created their passwords around words related to the topic of the service to which the password provided access, for example 'money' for the e-bank. On average, 1.71 of each participant's passwords contained such a topical word.
- Four participants had password-creation strategies involving the concatenation of two meaningful words, typically their name (see below) and a topical word. On average, 1.21 of each participant's passwords contained two concatenated words. An additional 2.21 of each participant's passwords contained one meaningful word.
- Three participants incorporated their own name (i.e., first name, middle name, last name, or userid) in their passwords. On average, 1.50 of each participant's passwords contained the participant's name. In addition, one participant used the names of friends' children and pets in his passwords.
- Two participants created passwords that corresponded to memorizable patterns on the keyboard (e.g., the two leftmost columns of keys). One of these participants consistently included special characters (i.e., neither letters nor digits) in his passwords, but apart from the five passwords created by this participant only one password contained a special character.

For each password the participants selected the one, two, or three characters they wanted as their MiFA hint. The distribution across one-, two-, and three-character hints was 29%, 66%, and 6%, respectively. Participants used the hints to amplify their password-creation strategies. This was mainly done by having the hints signal the start of the chunks of which the passwords were constructed:

- The initial letter of every meaningful word contained in a password was often included in hints. On average, each participant created 3.42 passwords containing meaningful words and for 2.29 of these passwords the hints included the initial letter of each of these meaningful words. Five participants used this method for all their passwords.
- Often passwords ended in a number external to the main password-creation strategy, and the initial digit of this number was included in the hint. On average, each participant created 4.21 passwords ending in a number and for 2.29 of these passwords the first digit of the number was part of the hint. Four participants used this method for all their passwords.

5. DISCUSSION

Four weeks after creating their passwords the participants remembered 67% of them correctly when aided by MiFA hints. This is significantly better than the 35% they remembered without hints and comparable to the 75% remembered after ten days in the study by Lu and Twidale [8]. Both their study and this study concern users' ability to remember their passwords in their first attempt. Given more than one attempt users will also succeed if the hints enable them to restrict the set of candidate passwords to a small number of alternatives.

Participants were also significantly more confident in the correctness of passwords recalled with the aid of MiFA hints than without such hints. Thus, participants' perception of the MiFA hints is consistent with their improved performance. The participants' high level of confidence suggests that they consider MiFA hints useful and may adopt them if introduced in operational systems. The decay in the participants' memory from one week after creating their passwords to four weeks after creating them was not significant but calls for testing MiFA hints longitudinally, if possible as an element of their introduction in an operational setting.

Without hints it is unrealistic to require that users always choose strong passwords, change them frequently, and never write them down. This entails a conflict between security and ease of use [e.g., 4, 5, 6, 9, 10, 12]. Passwords may be attacked by outsiders that aspire to gain access to systems. Such attacks can be broken into four types:

- *Informed guessing*: cracking a person's password by combining knowledge about the person with knowledge about frequently used password-creation strategies.
- *Social engineering*: persuading a person to reveal passwords by exploiting that humans are, in general, unsuspecting and want to help out if they can.
- *Dictionary attacks*: cracking passwords by trying a large number of candidate passwords in a brute-force manner.
- *Interception*: capturing passwords when they are entered by or echoed to legitimate users, for example by wiretapping data lines.

This simple typology serves to illustrate that unless users understand the different types of attacks they are likely to behave in ways that counter some types of attacks but remain vulnerable toward others. Further, the vulnerability of passwords toward dictionary attacks is increasing as still more powerful computers make it feasible to test passwords against still larger dictionaries [6, 11]. While humans' capacity for memorizing passwords is not going to change appreciably over the next decades, still longer passwords will be needed to prevent password-cracking algorithms from brute-force testing all possible character strings.

A frequent characteristic of the passwords created by the participants was that they had one or more digits appended at the end. This is known as salting and incorporated in many password-cracking algorithms. The participants' frequent use of names and topical words make the passwords more vulnerable to informed guessing. Further, their use of real words, whether topical or not, increase the vulnerability of their passwords to dictionary attacks. This vulnerability is, however, partly mitigated when passwords are created by concatenating two words. The near absence of special characters in the participants' passwords is a serious weakness because it reduces the password space dramatically. Finally, the participants' marked preference for passwords consisting of exactly eight characters is a regularity that can be exploited by password-cracking algorithms.

While the participants have, at least to some extent, succeeded in using the MiFA hints to amplify their password-creation strategies it appears that this amplification has been used primarily to make the passwords more memorizable and less to make them stronger. This suggests a need for supplementary user-authentication mechanisms that focus primarily on strength. Such supplementary mechanisms may include graphic challenges [2, 10], which exploit that humans read distorted text much better than computers, and dynamic identity verification by means of keystroke characteristics [7] or pointing characteristics [3].

6. CONCLUSION

This study replicates and confirms work by Lu and Twidale on minimal-feedback hints for user authentication. With the aid of hints the 14 participants in this study remembered significantly more passwords and were significantly more confident in the correctness of their memory of their passwords than without hints. Hints were frequently used to amplify password-creation strategies by revealing the initial character of the chunks of which passwords were constructed. However, many of the passwords created by the participants were weak in spite of the improved support for remembering passwords.

7. ACKNOWLEDGEMENTS

This work was funded in part by the IT University of Copenhagen. Special thanks are due to the students who participated in the study.

8. REFERENCES

[1] Adams, A., and Sasse, M.A. Users are not the enemy. *Communications of the ACM*, 42, 12 (1999), 40-46.

[2] Ahn, L. von, Blum, M., and Langford, J. Telling humans and computers apart automatically. *Communications of the ACM,* 47, 2 (2004), 57-60.

[3] Barrelle, K., Laverty, W., Henderson, R., Gough, J., Wagner, M., and Hiron, M. User verification through pointing characteristics: An exploration examination. *International Journal of Human-Computer Studies*, 45, 1 (1996), 47-57.

[4] Dourish, P., and Redmiles, D. An approach to usable security based on event monitoring and visualization. In *Proceedings of the 2002 Workshop on New Security Paradigms*. ACM Press, New York, 2002, 75-81.

[5] Hertzum, M., Juul, N.C., Jørgensen, N., and Nørgaard, M. (forthcoming). Usable security and e-banking: Ease of use vis-à-vis security. To appear in *Proceedings of OZCHI 2004*.

[6] Klein, D.V. "Foiling the cracker": A survey of, and improvements to, password security. In *Proceedings of the Second USENIX Security Workshop*. USENIX, Berkeley, CA, 1990, 5-14.

[7] Leggett, J., Williams, G., and Usnick, M. Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, 35, 6 (1991), 859-870.

[8] Lu, B., and Twidale, M.B. Managing multiple passwords and multiple logins: MiFA – minimal-feedback hints for remote authentication. In *Proceedings of the IFIP INTERACT 2003 Conference*. IOS Press, Amsterdam, 2003, 821-824.

[9] Morris, R., and Thompson, K. Password security: A case history. *Communications of the ACM*, 22, 11 (1979), 594-597.

[10] Pinkas, B., and Sander, T. Securing passwords against dictionary attacks. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*. ACM Press, New York, 2002, 161-170.

[11] Schneier, B. *Secrets and Lies: Digital Security in a Networked World*. Wiley, New York, 2000.

[12] Schultz, E.E., Proctor, R.W., Lien, M.-C., and Salvendy, G. Usability and security: An appraisal of usability issues in information security methods. *Computers & Security*, 20, 7 (2001), 620-634.

Studying cross cultural think-aloud usability testing – some suggestions for an experimental paradigm

Torkil Clemmensen
Department of Informatics
Copenhagen Business School
Denmark
tc.inf@cbs.dk

Shivam Goyal
Department of Design
Indian Institute of Technology
Guwahati, India
shivam@iitg.ernet.in

1. INTRODUCTION

Having test users to think aloud during usability testing is generally thought to be an effective and successful technique [7]. According to Nielsen (as cited in [2]), thinking aloud usability testing is the single most valuable usability engineering method for evaluating the usability of user interfaces. However, the descriptions in the Usability Literature of how this method is followed in the industry do not conform to the theoretical basis of this method [2].

The theoretical basis of the method is described in 'Protocol Analysis: Verbal Report as Data' [2, 3]. A certain relaxation from theoretical rigor when we apply a scientific method to practical purposes is what we may expect.

What is less understood and accepted, however, are the significant differences in how people from different cultures respond to directions and test methodologies. With the advent of globalisation and IT revolution, we can no longer overlook the aspect of culture in the design of user interfaces and products [6]. Taking cultural issues into account has now become one of the key factors for the success or failure of a global product. Despite this fact, however, we don't have any kind of formal method which guides us to evaluate a product to a certain standard while keeping in the sensitivity to cultural issues in different countries. International Usability Testing just involves a usability expert from one country and a local facilitator in the target country [4].

Previous studies on cross cultural usability evaluation have shown that culture broadly affects the usability evaluation assessment processes. In one sense, we try with the ongoing research presented in this paper to look into some of the broad issues raised in current international usability research [8]: "How do we avoid cultural bias in requirements elicitation and usability data collection?" and "What user based evaluation methods address cultural diversity in both the moderator and user?"

Specifically, in this paper we are interested in where to start an investigation of the assumption that the usability evaluator almost needs to belong to the target culture to (a) completely understand how people will respond to the think aloud directions and test methodology and (b) understand what is the effective way to obtain test users' usability feedback, without actually disguising the usability problems.

In the rest of the paper we present our planned experimental setup for studying think aloud cross cultural usability test.

2. SUGGESTED EXPERIMENTAL PARADIGME

We have suggestions for an experimental paradigm in three areas: the selection of participants, the selection of a test application and the data collection and analysis approach.

2.1 The participants

The Test users and the evaluators we will chose from Europe (Denmark) and India. As we can't take east and west as areas of study because culture is not at all uniform in whole east or in whole west [5], we choose one country from east and compare the result with one country from west, i.e. we chose national culture as a way to operationalize the selection of participants with differences in cultural practices.

Furthermore, despite the risk of selecting culturally too homogenous participants, we want participants who

- Can do the experiment in English. We will ask them about their ability in English and their experience with and knowledge of usability testing and thinking aloud. They should be able not only to talk English, but also to read English and to work with computers in English. They should be able to think aloud and even to do this in a foreign language like English.

- Can use the application and could perform the test-task. Therefore, the participants should be regular users and proficient in the 'host application' for the specific application to be tested, i.e. Microsoft word, and preferably have experience with the clipart features, and they should have experience with the test user task, i.e. making a birthday invitation by using a word processor and clipart.

To check our immediate selection of participants, we will screen them by having them to give us the above presented information about themselves in a questionnaire.

As a substitute for established acculturation indexes (see e.g. [9]), we will try to get a rough measure of test users' familiarity with local national culture. By asking the participants about the time of their stay in the local national culture (Denmark), which should be either from birth (local test user) or between two months and six months (distant test user; at least two months stay to avoid test users with cultural shock and rejection of local culture known to occur as a response immediately after the 'honeymoon experience of being in another country). Local national culture may thus expected to be either very familiar or to be very unfamiliar (but interesting) to the test users. In our first study, we will not consider status or age and sex relations, but only focus on conducting all the possible combinations of test user and evaluator pairs.

2.2 The test application

The interface that we choose has to be culturally localized. We suggest the use of a ‘cultural clipart application’, which is collection of culturally specific images and icons of various cultures. Thus an application will help a user to make invitations, documents containing graphics, web pages.

The need for localizing clipart collections. At present people use the Clip Organiser feature provided by the Microsoft Office in many different cultures. But the problem that we found with the Clip Organiser is that it’s not currently localized, meaning the images and graphics that it contains now don’t have the cultural sensitivity for each culture in particular. Our intention is to make a ‘Cultural Clipart’ application that is more sensitive to culture.

The prototype cultural clipart collection. Since our application is on the ideation phase we must combine it with the Microsoft Word ® Clipart application. We have incorporated our Cultural Clipart in this clip Organiser by adding a sub folder named Cultural Clipart in My Collections. The Cultural Clipart folder contains the cultural specific sub-folders named Denmark, Germany, Sweden, etc. From any of these folders, the user can choose images and graphics to add to their document.

The creation of a cultural clipart collection. But how did we get all the clip arts as the experimental material? The clipart material that we propose to use is a result of (a) a two-week long camera-ethnography on the streets of Copenhagen, (b) discussions about Danish culture with an anthropologist specialized in the concept of nationality, (c) search for clipart on the internet and (d) our informal analysis of examples of Danish birthday invitations. In the end, we have decided to go for 'tourist' clipart to symbolize Denmark, realizing that the task we will ask the Test user to perform require such distinctions.

Inclusion of potential usability problems. In order to increase the chances of culturally specific interaction between Test user and evaluator we will introduce some errors in the clipart whom only a user from Danish possibly can identify. Therefore, we embed the following usability problems in our ‘Denmark’ folder:

- A Norwegian flag in the collection
- An image of Norwegian parliament in the Denmark folder
- Image of Heineken beer (a Dutch beer)
- A Reindeer (which does not live in Denmark)
- A Norwegian skier
- Giving the blind-fold game (not Danish) the birthday keyword
- Images of Birthday certificate, eagle and scenery of which none was Danish
- Wrong keywords to images of Danish flags
- Giving the keyword Birthday to an amusement park of Denmark
- Giving the keyword of birthday cake to various cakes that were neither birthday cakes nor Danish cakes

These problems are supposedly culture specific.

2.3 The test scenario and tasks.

In the test scenario, a company that develops a new application has hired an evaluator from a usability consultant company to perform think aloud usability tests of the product and has also provided test users. The test users and the evaluators will be informed about the scenario and asked to provide background information about them in a questionnaire sent to them before the experiment take place.

The experiment will begin with the evaluator’s arrival at the company’s office, i.e. the research lab, followed by arrival of test user. Then the test session is performed, and after this, the evaluator does the post-test interviews with the test user, and, finally, the researcher interviews both participants before they leave the lab.

Evaluator’s task. Evaluators will be asked to conduct the test in the way they are comfortable with and they will have to build their own test plan and interview guidelines. We provide them with list of tasks that they need to ask user to perform. They will also be provided with objectives of the test, which is – for the evaluator – to identify usability problems in a broad sense, both regarding content and navigation of our cultural clipart collection.

Each test session begins with the evaluator greeting the test user. The evaluator then explains the test set up, about the objectives of the test, about the interface, about video recording and camera. Then the evaluator asks the test user to read all the sub tasks very carefully and then perform them. After the completion of all the tasks the evaluator interviews the user about the interface and the task he/ she performed. The questions of the interview are solely on Evaluator’s choice and according to his test plan keeping in mind the objectives of the test.

Test user’s task. The test user will be provided with the task of making a birthday invitation for his son. It has the following sub tasks:

1. Please write the text that you want to appear on the Invitation.

2. Please choose the appropriate font(s) for the text.

3. Please choose the appropriate style(s) for the text.

4. Please choose the colour(s) for the text.

5. You are free to choose any kind of formatting and layout that you require for this text.

6. Now using the Cultural Clipart sub-folder in My Collections folder in Microsoft Clip Organiser add some images and graphics so that its looks like Birthday invitation.

7. Please make this invitation look happy, colourful, and joyful as this is for birthday.

8. Since primarily all your guests are from Denmark and are Danes, make this invitation look Danish.

The Test user is asked to read all the sub tasks very carefully and then perform them. The test users shall perform the tasks using Microsoft Word (Microsoft Office XP Professional) on a Windows 2000 computer or similar.

2.4 Data collection

The experiment will be performed as a laboratory experiment and take place in our research lab, which basically is a standard office space that can be found in any major company.

Both the test user and evaluator will asked to fill in a questionnaire which judge their knowledge of Usability testing and Microsoft Word. The questionnaire will also give insight into the participants' knowledge of the two cultures.

We record all instances of the participants' behavior during the thinking aloud period from three different angles. We will have a digital video camera directed at the test user's chair allowing capture of the test user's facial expression, and another digital video camera placed at 2 m distance allowing a capture of evaluator and test user interaction, as well as their interaction with the experimenter during the post test interview. Furthermore, we will use the Camtasia ® screen recorder software to capture the screen events. Among these three sources of data, the primary source is expected to be the wide-angle camera that allow us to see both participants as their relationship unfolded during the experiment.

The data collection method will thus be continuous, i.e. test participants are observed for the whole period of the thinking aloud usability test, including evaluators introduction and post-test interview with the test user, but excluding the researchers' post-session interview with the evaluator and test user.

The research interviews. The think-aloud session will be followed by three research interviews. The researchers interview first the evaluator and test user at the same time, and, secondly, the researchers interview the evaluator alone and, thirdly, the researchers interview the test user alone. These interviews are conducted to explore the relationship between the evaluator and test user during the test, e.g. the interviewer probe into the interaction between the test user and evaluator; and how the user feel during the test and what else he/she wants from the evaluator and do the nature of the evaluator's reminders affect his or her performance, and, furthermore, does the evaluator feel that he or she understood the test user's thoughts. The test user and the evaluator are asked to relate to the other's statements in order to get a dialogue about their relationship and its effect on the thinking aloud during the usability test and the number of usability problems that were found.

2.5 Data analysis

We wanted to produce a complete behavioural record of the participants' relationship as it unfolded during the experiment, including the time at which each instance of a behavioural relation occurred (events) or began and ended (states). Our analysis of the videos therefore focused on the wide-angle camera that allowed us to see both participants.

The coding scheme that we propose has been developed during a small pilot study with the above presented methodology. It consist of a number of behavioral classes that each contains different subtype behaviors:

2.5.1 Reminders

Neutral (Ericsson & Simon type reminder), Affirmative (active listening type reminder), Interrogative.

1. **Interrogative** Reminders are those in which Evaluator asked a question to make the Test user think aloud. E.g. what are you trying to do? What are you looking for? Etc.
2. **Affirmative** Reminders are those in which Evaluator conveys a message to the Test user that he/she (the Evaluator) is an active listener. These reminders are similar to those identified by Boren and Ramey[2] as 'Acknowledgement tokens', e.g. Hmmm, Mm Hmm, Yeah, Ok.
3. **Neutral** Reminders are those which are given to ask test user to continuing thinking aloud. E.g. Keep Talking, Please Keep Thinking Aloud, What are you thinking? (In a non-interrogative tone). These reminders are similar to those proposed by Ericsson and Simon [3]in the appendix to their book.

2.5.2 Task fulfillment evaluator behavior

These behaviors are of different types: Observe Silently, Comment, Answer Questions, Help Out.

1. **Observe Silently** means the Evaluator act as a passive listener and don't say anything while user thinks aloud, or ask question or obviously have difficulties.
2. **Comment**s are those in which evaluator pass some comments; e.g. the Evaluator interrupts and says "You have to read the task first", "You can only use Microsoft Word to fulfil this Task"
3. **Answer Questions** are those evaluator behaviours in which he/she has to answer to the user's doubt about either the task or the application.
4. **Help Out**s are those behaviours in which the Evaluator actually helps the user to complete the task. These behaviours are directly against the suggestions by Ericsson and Simon [3] and also against newer Usability Testing theory [1].

2.5.3 Usability problems

These are related to understanding image content, finding clip organizer, modifying images, choosing images and other problems, including general use of the word processor.

1. **Images:** These are the number of responses which the Test user gives when he/she is not happy with image collection, quality, and its cultural significance.
2. **Clip Organiser:** These are the number of responses which the user gives when she/he has problems using the Clip Organiser. E.g. the name of the image can't be found, the keywords were not right, it doesn't show the actual size of the image etc.

3. **Choosing:** These are responses given when he/she has trouble transferring the file from the clip organiser to the Word Document.

4. **Word:** These are the problems related to the Microsoft word.

2.5.4 User behavior

These are of different types: Thinking Aloud, Silence, Explanation, Positive Comment, Negative Comment, Cultural comment, Question, Suggestions, Other responses.

1. **Suggestions:** These are the responses given in order to improve the usability of the application.

2. **Cultural Comments:** These are the responses and references given by the user on the users' native cultural or on the localised culture of the application.

3. **Positive Comments:** These are the user response which says something about the positiveness of the design and application.

4. **Negative Comments:** These are the user responses which comments upon the disapproval of design or the application.

These codes shall be applied on the recorded videos by at least two researchers or by one researcher after thorough discussion of each class and subtype of behavior.

3. CONCLUSION

The presented paradigm for studying cross cultural think-aloud usability testing has advantages: Elementary statistics can be performed on the codes, and presented and discussed as results. The analysis of the interviews can be used to interpret and backup our understanding of the results of the coding.

One major limitation with this approach is our focus on events, i.e. how many times a test user began to think aloud. Instead, it may be relevant for comparison between subjects to get a measure of states, i.e. how long the evaluator and test user is in a 'think aloud mode', as the important thing in think aloud sessions during usability testing is to get the user to think aloud much of the time and not just many times.

4. REFERENCES

1. Barnum, C.M. Usability testing and research. Longman, 2002.

2. Boren, M.T. and Ramey, J. Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication, 43 (3). 261-278.

3. Ericsson, K.A. and Simon, H.A. Protocol Analysis. Verbal reports as data. Cambridge Massachusetts, 1993.

4. Nielsen, J., Designing for International Use. in CHI 1990, (1990).

5. Nisbett, R.E. The Geography of Thought. Nicholas Brealey Publishing, London, 2003.

6. Russo, P. and Boor, S., How Fluent is Your Interface? Designing for International Users. in INTERCHI 1993, (1993).

7. Schneiderman, B. and Plaisant, C. Designing the User Interface. Pearson Addison Wesley, 2004.

8. Smith, A. and Yetim, F. Global human-computer systems: Cultural determinants of usability. Editorial. Interacting with Computers, 16.

9. Zheng, X., Sang, D. and Wang, L. Acculturation and subjective well-being of chinese students in Australia. Journal of Happiness Studies, 5. 57–72.

Usability Evaluation with Children

Benedikte S. Als[1], Janne J. Jensen[2], Mikael B. Skov[1]

[1]Department of Computer Science
Aalborg University, Denmark
+45 96 35 80 80
{als, dubois}@cs.aau.dk

[2]Department of Communication Technology
Aalborg University, Denmark
+45 96 35 80 80
jjje@kom.aau.dk

ABSTRACT

Emerging technologies for children often require the involvement of children as test subjects in software development projects. Previous research studies have indicated that children behave differently than adults when involved in usability test sessions. Thus, children impose new opportunities and limitations in the evaluation and we still need to investigate proper and fruitful ways of involving children.

We present two studies on usability evaluation with children. The first study involved eight children in a development project on a mobile educational device. The children evaluated a number of different prototypes. The second study involved 60 children who participated in the evaluation of an existing mobile technology where the children applied either the think-aloud protocol or constructive interaction.

The results from the first study showed revealed that evaluation with a high-tech prototype does not necessarily provide more useful information, compared with an evaluation of a low-tech prototype. Our results from the second study revealed that constructive interaction provided the identification of more usability problems compared to think-aloud when the pair composition in constructive interaction is acquainted dyads.

Keywords

Children, usability testing and evaluation, think-aloud, constructive interaction informants

1. INTRODUCTION

The design and evaluation of children's technologies have received increased attention during the last several years [9]. Children should be considered individuals with strong opinions, needs, likes, and dislikes, and they should be treated as such [7]. Druin [6] provides a classification of involvement where children play the roles of users, testers, informants, or design partners. The four roles encompass different levels of engagement and impose different opportunities and limitations. All roles apply usability tests where children participate as subjects.

Usability testing has been studied extensively and is generally acknowledged to identify some of the key interaction problems in user interfaces [12]. Emerging technologies for children have produced the need for involving children as test subjects in software development projects. However, usability testing may be challenging with children situation since they are typically less organized [5]. Hanna et al. [10] propose adjusted guidelines for usability testing with children such as reflection on common target age ranges and how different age groups can verbalize their thoughts and feelings during a test.

Going through CHI Proceedings for the last ten years, we found several studies in which children participated as test subjects in usability tests applying e.g. think-aloud [1, 4, 6], constructive interaction [11] or both approaches [12]. Typically, the children are involved in the testing of existing technologies or high-tech prototypes [2, 3, 4, 8, 11, 12] whereas they are less involved in the testing or evaluation of low-tech or paper prototypes [12].

In this paper, we report from two different studies on usability evaluation with children. The first study involved eight children in a development project on a mobile educational device. These children evaluated a number of different prototypes from the development project. The second study involved 60 children that participated in the evaluation of an existing mobile technology where the children applied either the think-aloud protocol or constructive interaction.

2. STUDY A

This study used a revised form of a four-phased method used in the [12]. The revised form consisted of a method were all expert adult influence were removed from the first three phases. In the fourth phase we excluded all but the adult experts, who should comment on the correctness of the information provided by the children. The aim of the study was: How much information can we derive from children alone? And how valid is the information?

2.1 Participants

8 children (3 girls and 5 boys) at the age of 10 to 12 years old (*mean*=11.5, *SD*=0.76) participated. Five of these children had been living with diabetes for an average of 2,87 years. The adult participants had an average of 30 years working as a nurses; one had worked with diabetic children for 21 years the other for 1½ year. The children were not aware of that they would receive a small gift for their involvement in the study.

2.2 Procedure

The usability test sessions were conducted in a specialized usability laboratory. The laboratory integrated two rooms; an observation room in which the evaluations took place and a control room where one of the researchers would handle electronic equipment for recording the sessions. The two rooms were separated with a one-way mirror allowing people in the control room to see what was going on in the observation room.

The first usability evaluation was conducted with a low-tech prototype; it was made out of colored paper and plastic slides fitted to the same size as the screen of the PDA, so the children could get an idea of the size of the screen. For the second evaluation we used a high-tech prototype running on the PAD, it was developed in eMbedded Visual Basic 3.0.

2.3 Tasks

The children in the first session were given two tasks, tell us about diabetes, and tell us about your likes and dislikes regarding computer games and mobile phones. During the second session, with the same children, the children were presented with a paper prototype of the system, they commented on the idea and the design. We also presented them with 26 questions that could be implemented in a system, they were asked to answer the questions (pick the right answer from the four options) and comment on the questions.

In the fourth session the children were asked to solve the same tasks as the children in the second session. The only difference were that the prototype they were presented for were a running prototype. Afterwards the two adult experts solved the same tasks, thereby giving the team knowledge about the correctness of the information given by the children.

2.4 System

The system used in the experiment was a an edutainment system, the users were presented with a paper prototype of the system and a prototype which ran on a Compaq PocketPC

The target group of the system was children who had been diagnosed with diabetes and their friends. The system should teach the children specific information about diabetes, as well as entertaining the children, the reason for this being that it is easier to capture children's attention through edutainment systems, than it is through purely educational systems. The inspiration for the system came from the game "who want's to be a millionaire", which has one right and three wrong answers for each question. Thereby avoiding that the children should write an answer for each question.

3. STUDY B

Our experiment utilized a setup for comparison of think-aloud and constructive interaction for usability testing with children. In particular, we wanted to

1) Measure think-aloud and constructive interaction on identification of usability problems

2) Explore the impact of different compositions of pairs in constructive interaction

3) Analyze children's perception of the testing situations using think-aloud and constructive interaction.

We designed the experiment as a 3x2 matrix consisting of three types of sessions: individual testers using think-aloud, acquainted dyads (pairs) using constructive interaction, and non-acquainted dyads using constructive interaction. Furthermore, we configured the usability test sessions with same-sex dyads having sessions with girls and boys for each of the three setups.

3.1 Participants

60 children (30 girls and 30 boys) at the age of 13 and 14 years old (*mean*=13.35, *SD*=0.48) participated as test subjects in the experiment. The children were all 7th grade pupils from five different elementary schools in the greater Aalborg area. The children did not receive compensation for their involvement in the experiment.

3.2 Procedure

The purpose of the evaluation was explained in detail to the children and they were shown the facilities of the usability lab. Test subjects intended for roles as non-acquainted dyads were kept separate before the test sessions. The children received questionnaires on which they had to provide answers to such as age, name, school, and mobile phone experience.

The usability test sessions were conducted in the same usability laboratory as Study. All sessions were recorded on videotapes for later analyses including perspectives of the children and of their interactions with the mobile phone.

3.3 Tasks

The children were asked to solve twelve tasks one at a time addressing standard and advanced functionalities in the innostream mobile phone. This included making a phone call, sending a short text message, adjusting the volume of ring tones, and editing entries in the address book. We did not specify any time limits for the tasks, but required the participants to try to solve all tasks. All children were able to solve all specified tasks. On average, the children spent 26:45 minutes (*SD*=06:39) on the twelve tasks.

3.4 System

The selected system for our experiment was an inno-100 mobile phone by innostream. This particular mobile phone was selected since it had not been released on the European market at the time of our experiment. Thus, all children would have to learn to use the mobile phone.

The inno-100 integrates a range of standard mobile phone features, such as making and receiving phone calls and short text messages, and more advanced features, including speed dial functions and options for creating personalized ring tones. The inno-100 has two separate screens with a main 128x144 pixel 16 bit color screen and 64x80 pixel sub screen on the cover. The navigation is primarily based on icons in the two upper menu levels. The lower levels are textual based including choice menus for setting values. Furthermore, the inno-100 integrates a number of games.

4. RESULTS

4.1 Study A

Our results indicate that the overall comments from the children testing a low-tech prototype are almost identical to the comments we provided by the children testing the high-tech prototype. The only difference we could find were that the second evaluation provided information about functionality errors.

From the usability evaluation of the low-tech prototype, we found that the children were capable of imagining a real system while looking at a paper prototype. Both the boys and the girls commented on what they thought could be funny features. We got suggestions as animation of face, changing the needles gender, music, reading the text out loud, giving the needle arms etc. The children understood most of the functionality that would be incorporated in the buttons, the girls had some trouble understanding and "quit" they suggested that quit should be replaced with a Danish word. The girls had a hard time imagining what could be search for in the game, and therefore they didn't understand the "search" button, after seeing the search page, they understood the function. When presented with the search page none of the children thought that it would be a good idea if the player should write the word themselves, since a person without diabetes wouldn't know the words. Furthermore one of the boys suggested that a historical anecdote could be added to each topic, and one of the girls suggested the addition of pictures or small clips of film. As for the game, all the children liked that they had four possible answers to choose between, since it would be easier than of they had to write it themselves. The boys liked the idea of sticking the needle in the right answer; they suggested that the background had skin colored texture. All the children suggested that the game could be played as a multiplayer game, and the boys suggested that it should be possible to race the clock. During the talk over amount of questions one of the girls noted that if the computer would pick questions randomly, some of the most important information regarding diabetes might be missed.

The information provided by the children in the second evaluation was almost identically as the results from the first evaluation. The children however stressed that it was important that the voice of the needle was the voice of a child, since they didn't want it to be an adult. All the children in this session had doubts on whether the "help" function would give them help to the quiz or how to play the game. The boys suggested that it should be possible to click on the needle during the quiz, to get some help if needed. The two children with diabetes didn't like to stick the needle into the right answer, whereas the boys who didn't have diabetes thought that the needle could be helpful to children who were afraid of needles since he looked so nice. The two boys without diabetes liked the search function, which was found a bit boring, by the two children with diabetes. During the quiz the girl accidentally hit the "next" button twice, and thereby answering the next question with the second tab. Additionally one of the children suggested that it should be possible to race each other with two linked PDA's online. The two boys also suggested an idea for a game where they could control the needle and maybe shoot unhealthy food.

4.2 Study B

Our results indicated that constructive interaction provided the identification of a higher number of usability problems compared to think-aloud, but the differences were mostly not significant. However, we found significant influence of the pair composition in constructive interaction as the non-acquainted dyads identified significantly less problems than the acquainted dyads. The acquainted dyads identified more total numbers of problems and serious problems. However, this did not seem to increase level of frustration for the acquainted dyads. We further found that the girls identified more problems in constructive interaction as acquainted dyads compared both girls applying think-aloud and non-acquainted girls. No similar differences were found for the boys.

Specifically, our study resulted in the identification of 81 different usability problems. Based on a classification scheme, we classified 32 of these 81 usability problems as critical problems, 13 as serious problems, and 36 as cosmetic problems.

Our results showed that the sessions with acquainted dyads identified the highest number of usability problems of the three setups. The 12 acquainted dyads identified a total of 65 of the 81 usability problems whereas the non-acquainted dyads identified only 51 of the 81 usability problems and this difference was significant according to a two-tailed Chi-square test (χ^2=5.131, df=1, p=0.0235). The individual testers identified 56 of the 81 usability, but the difference between the individual testers and acquainted dyads was not significant (χ^2=2.090, df=1, p=0.1483) nor is the difference between the individual testers and acquainted dyads (χ^2=0.440, df=1, p=0.5069).

Looking at problem severity, we found that the acquainted dyad sessions identified nearly all critical problems (28 of the 32 critical problems), but this was not significant compared to the individual testers or the non-acquainted dyads according to a Chi-square test (χ^2=0.439, df=1, p=0.5076) (χ^2=2.286, df=1, p=0.1306). However, we found that the acquainted dyads identified significantly more serious problems than the non-acquainted dyads (χ^2=4.514, df=1, p=0.0336). Alternatively, no significant differences were found between acquainted dyads and individual testers on serious problems (χ^2=1.950, df=1, p=0.1626) nor between individual testers and non-acquainted dyads (χ^2=0.155, df=1, p=0.6940). We found no significant differences for the cosmetic problems.

5. CONCLUSION

This paper has reported from two different studies on usability evaluation with children. The first study involved eight children evaluating a number of different low-tech and high-tech prototypes. The second study involved 60 children in the evaluation of an existing mobile technology where the children applied either the think-aloud protocol or constructive interaction.

The first study revealed that evaluation with a high-tech prototype does not necessarily provide more useful information, compared with an evaluation of a low-tech prototype.

The second study revealed that constructive interaction provided the identification of more usability problems compared to think-aloud when the pair composition in constructive interaction is acquainted dyads.

ACKNOWLEDGMENTS

We would like to thank the participating children and their parents from the two studies..

REFERENCES

1 Benford, S., Bederson, B. B., Åkesson, K-P, Bayon, V., Druin, A., Hansson, P., Hourcade, J. P., Ingram, R., Neale,

H., O'Malley, C., Simsarian, K. T., Stanton, D., Sundblad, Y., and Taxén, G. (2000) Designing storytelling technologies to encouraging collaboration between young children. In Proceedings of the Human Factors and Computing Systems CHI'00, ACM Press, pp. 556 - 563

2 Bers, M. U., Ackermann, E., Cassell, J., Donegan, B., Gonzalez-Heydrich, J., DeMaso, D. R., Strohecker, C., Lualdi, S., Bromley, D., and Karlin, J. (1998) Interactive Storytelling Environments: Coping with Cardiac Illness at Boston's Children's Hospital. In Proceedings of the Human Factors and Computing Systems CHI'98, ACM Press, pp. 603 - 610

3 Bruckman, A. and Edwards, E. (1999) Should We Lev-erage Natural-language Knowledge? An Analysis of User Errors in a Natural-language-style Programming Language. In Proceedings of the Human Factors and Computing Systems CHI'99, ACM, pp. 207 - 214

4 Danesh, A., Inkpen, K. M., Lau, F., Shu, K., Booth, K. S. (2001) Geney: Designing a collaborative activity for the Palm handheld computer. In Proceedings of the Human Factors and Computing Systems CHI'01, ACM Press, pp. 388 - 395

5 Druin, A. and Solomon, C. (1996) Designing Multimedia Environments for Children. Wiley & Sons, New York

6 Druin, A. (1999) The Role of Children in the Design of New Technology. HCIL Technical Report No. 99-23, University of Maryland, USA

7 Druin, A. (1999) The Design of Children's Technology. Morgan Kaufmann Publishers, Inc., San Francisco, CA

8 Ellis, J. B. and Bruckman, A. S. (2001) Designing Palaver Tree Online: Supporting Social Roles in a Community of Oral History. In Proceedings of the Human Factors and Computing Systems CHI'01, ACM Press, pp. 474 - 481

9 Gorriz, C. M. and Medina, C. (2000) Engaging Girls with Computers through Software Games. Communications of the ACM, vol. 43, No. 1, pp. 42 – 49

10 Hanna, L., Risden, K., and Alexander, K. J. (1997) Guidelines for Usability Testing with Children. In interactions, September + October, pp. 9 – 14

11 Montemayor, J., Druin, A., Farber, A., Simms, S., Churaman, W., and D'Amour, A. (2002) Physical Programming: Designing Tools for Children to Create Physical Interactive Environments. In Proceedings of the Human Factors and Computing Systems CHI'02, ACM Press, pp. 299 - 306

12 Nielsen, J. (1993) Usability Engineering. Academic Press

13 Scaife, M., Rogers, Y., Aldrich, F., and Davies, M. (1997) Designing for or Designing with? Informant Design for Interactive Learning Environments. In Proceedings of the Human Factors and Computing Systems CHI'97, ACM Press, pp. 343 - 350

Fitts' law in an audio perspective *

Amalia de Götzen
Medialogy
Aalborg University
Copenhagen, Danemark
adgo@ihk.dk

Stefania Serafin
Medialogy
Aalborg University
Copenhagen, Danemark
sts@cs.aue.auc.dk

Juraj Kojs, Rolf Nordahl
Medialogy
Aalborg University
Copenhagen, Danemark
juko@ihk.dk,
rn@medial.aau.ihk.dk

ABSTRACT
Fitts' law is a very well known instrument in Human Computer Interaction. It allows designers to evaluate interfaces and input devices. Substantial research has been developed in this direction; however little has been done to use Fitts' law to evaluate multimodal performance where sound has to be taken into account. The aim of this paper is to describe some unexplored directions in evaluating new kinds of interaction in interface design.

Categories and Subject Descriptors
H5.5 [**sound and music computing**]

General Terms
Fitts' law, Psychoacoustics, multimodal interaction

1. INTRODUCTION
Sound started to find its place in the broad field of Human Computer Interaction in the last decade of twentieth century; it became very important then to find new auditory interfaces and displays.

This kind of studies involves a lot of different research domains which are finally communicating and working together: human performance, auditory perception and signal processing are all involved in this investigation area. In particular, our focus is on continuous human-machine interaction with multimodal feedback; in this perspective, Fitts' law is a very open field: considerable work has been carried out in HCI on the Fitts' law model but the literature on Fitts' law with sound feedback and expressive gestural control appears to be very scarce.

This investigation is also suggested by the big role played by multi-modality and multi-sensory communication in the design of next generation interfaces: non-speech communication will play an important role inside the information stream established between machines and users.

*(Produces the permission block, copyright information and page numbering). For use with ACM_PROC_ARTICLE-SP.CLS V2.6SP. Supported by ACM.

2. ABOUT FITTS' LAW: ORIGINS AND RECENT DEVELOPMENTS
The origins of the Fitts' performance model, so useful in human-computer interaction, must be kept in mind when considering the Fitt's law. The law takes its name from its author whose innovative idea, in 1954[8], was to apply information theory to human-motor systems. The model is based on time and distance. It enables the prediction of human movement and human motion based on rapid, aimed movement (i.e. not drawing or writing) like in the tapping experiment shown in fig.1.

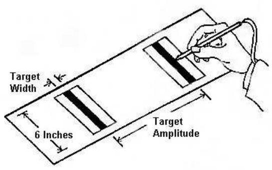


Figure 1: *The reciprocal tapping paradigm, the first experiment conducted by Fitts*

An intuitive idea is that movement time is be affected by distance and by the precision required by the size of the target towards which one is moving.

Fitts discovered that movement time was a logarithmic function of distance when target size was held constant, and that movement time was also a logarithmic function of target size when distance was held constant. Mathematically, Fitts' law is stated as follows:

$$MT = a + blog_2(2A/W), \text{ where:}$$

- MT = movement time
- a, b = regression coefficients
- A = distance of movement from start to target center

- W = width of the target

Here lies the innovative aspect of Fitts' law: a quantitative way to measure the difficulty of a motor task becomes available through it and a "new" way to transmit information is implicitly described through the definition of a human channel. Fitts defined also some other indexes that show the analogy with the Shannon formulation:

1. $ID = log_2(2A/W)$, the index of difficulty
2. $IP = ID/MT$, the index of performance, analogous to channel capacity C.

In [16] there is a detailed analysis of all the variations on Fitts' law, (i.e., Welford, MacKenzie formulation) derived from the need to correct the approximation by Fitts of the Shannon theorem, avoiding:

1. the theoretical problem of a negative rating for task difficulty,
2. the problem of the disproportion of the relative contribution of A and W in the prediction equation: similar changes in target amplitude and target width don't affect similar but inverse changes in movement time, as suggested in the original Fitts formulation

Skipping the detailed analysis of the data done by MacKenzie, we show here the derivation by direct analogy to the Shannon formula, which is also the most frequently used because it fits better with empirical data:

$$MT = a + blog_2(A/W + 1)$$

3. FITTS' LAW IN MULTI-MODAL INTERFACES

Fitts' law has had a strong development driven by HCI input devices. As a consequence, a great deal of investigation has been done, for instance, on the consideration of the target shape [5] and dimension (improving the prediction power of the model by substituting the target width with a measure of subject response variability in two or three dimensions) or taking into account the approach angle to the target itself; many new suggestions have been developed along the years, trying to verify the power and extension of the Fitts' model. Fitts' law is in fact a predictive model[17] widely useable in HCI which allows metrics of human performance to be determined analytically without undertaking time-consuming and resource-intensive experiments. In fact, extensive research work has been done using Fitts' law to compare input device performance characteristics and extending the model to other situations/applications (like, for example, the dragging interaction[18]).

In this context, we envision two possible avenues (among several others) to apply some of this previous research to a broadly intended field of sound:

1. considering the performer and his musical instrument, and
2. considering the auditory display in human-computer interaction design.

3.1 Fitts' law is good for sound

The above considerations are very close to a musical perspective: the performance of a musician is a particular cooperation and interaction between several muscles, and an analysis in a Fitts' perspective could indeed be interesting. For instance, an analysis of the piano literature in terms of targets or in terms of Fitts' law could lead to interesting developments.

This kind of musical task is very demanding even if is still a single-scale pointing control, and it suggests that the arm can be thought like a multiscale structure: it assigns precise small-scale movements to finger, which are naturally specialized for fine-grained actions, while larger-scale movements are left to arm joints (wrist, elbow, and shoulder), whose natural role is to produce wider motions of the hand. A similar, well founded, observation was carried out in [15] where the authors report the results of a preliminary pointing experiment that shows that users can handle higher levels of task difficulty with two-scale rather than traditional one-scale pointing control. In this case, the most interesting element is the conclusion that the motor control system has a better resolution than the eye in human-manual tasks, and that vision is the limiting factor for tolerance.

Even if this kind of application will not appear to be too useful in a musical perspective,it can still show interesting paths to follow further. These considerations are strictly connected with the kind of gestural control which is very often involved in an audio interface. When performers are pointing a sounding-target the gesture control is previously determined and cannot be modified during the movement itself, for instance by augmenting the size of the target. Gestural expressiveness becomes a very important component of the movement and the same applies to the successful and accurate reaching of the target. Expanding targets seem to be useful if we think at movement in a discrete way: however, it appears that music and sound need continuous control, and furthermore the auditory channel performs a continuous feedback.

Continuous control can be analyzed applying the steering law, derived from the Fitts' law [1], [2], [3], [5], [4]: these authors show that trajectory-based interaction cannot be successfully modeled by the original formulation of the Fitts' law, but nonetheless there are robust regularities in trajectory-based tasks. The authors found out that experiments like "steering through tunnels" are modeled by the steering law which can be very useful, for instance, in designing hierarchical menus, fig.2.

Devices can be evaluated[2] performing this "new" kind of task, now used in HCI as often as target pointing. An interesting application of the steering law is in locomotion, particularly in driving: the results suggest to study this law for other applications, in the perspective of establishing what was found for hand movement for more complicated gestural

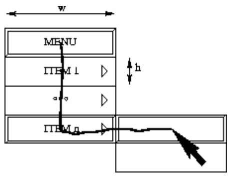


Figure 2: *Interacting with menus*

control too (for instance that implied by a musical instrument); thinking about a trombone or a violin, the kind of metaphors this law suggests are clear: there are path constraints, there is a gestural interaction involving the motor system and also there are time constraints.

3.2 Fitts' law sounds good

In general, auditory display is effective and inexpensive for all aspects of interaction that are inherently temporal. This is another interesting perspective for our research about Fitts' law and sound. Moreover, in the broad field of human-machine interaction researchers are continuously proposing new systems and paradigms, but a quantitative evaluation of the effectiveness of new interactions is often missing. In particular, multimodal communication via continuous interaction is still an area with many under-explored issues. Certainly, the designer would benefit from predictive models (similar to Fitts' law) and guidelines to support and complement intuitions and aesthetic concerns. The literature which concerns directly Fitts' law and sound modeling is scarce. However, some previous results exploring the auditory feedback instead of visual feedback in Fitts' law are worth mentioning: [26], [11], [22]. In these works the auditory feedback proved to be useful as a navigation aid (in a timbre space navigation [26]) or for gestural control (in a mobile device [22]), or was even made to respond to a linear model instead of the Fitts' model (in the non visual bullseye model [11]).

HCI knowledge can be used to consider new musical interfaces [20] which can make good use of different examples of auditory interfaces, for example *earcons* [7] and *auditory icons* [14]. Considering auditory icons, Gaver's work [13] shows convincingly that informative sound feedback is possible and useful. It is also clear that, up to now, the majority of these applications offered just ancillary and poor sonic feedback, while an in depth study of the effectiveness of auditory display in conveying information about processes controlled by continuous interaction [23] is still missing. This kind of feedback is accomplished by what was called by Gaver [12] *evolutionary objects*, where variables controlling the properties of the sound are updated while the sound is playing: physics-based sound modelling can provide a good instrument for this kind of feedback. In particular, simple metaphors like the rolling ball in [23] can be used to provide feedback to simple gestural input devices; in the mentioned case study the control metaphor is that of a rectilinear track to balance a ball that rolls over it: it's a simple metaphor which can applied in many different interaction contexts. It has been observed that the auditory feedback can provide in this case considerable information about the ball and the surface, something that pure visual feedback cannot provide. This form of auditory feedback stemming from physics-based modelling is inspired by the excellent control capabilities of music performers such as, for instance, violin players.

4. CONCLUSIONS

After novel expressivity paradigms and gesture analysis methods have been developed in an effort to map human gestures into quantitative scales (e.g. Fitts' law, steering law, etc.), it has become crucial to deal with new, often specific sets of control parameters.

Also, the role of multi-modality and multi-sensory communication will be central in the design of the next generation interfaces. As a consequence, non-speech communication will play an important role inside the information stream established between machines and users [6, 9].

Furthermore, a sound modeling activity which does not take the display into account, starting from the early design steps, is prone to potential failure. This fact has become particularly evident recently, when a new generation of non-conventional audio appliances (ranging from small portable devices, such as the last-generation mobile phones and palmtops, to multiple-loudspeaker Hi-Fi systems for the home theater) has been made available to large parts of the population. Hence, the aspects of display that must be now taken into account during the development of a sound model have become manifold [10].

Techniques for the evaluation and validation of such interfaces must be further developed in the realm of sound quality research and in interfaces for content-addressable sound database search engines ([21, 19, 25, 24]).

In all these issues, Fitts' law can provide considerable developments through many unexplored paths and the continuous human-artefact interaction appears to be a challenging task.

5. REFERENCES

[1] J. Accot and S. Zhai. Beyond fitts' law: Models for trajectory-based hci tasks. In *Proc. of CHI'97*, Atlanta, Georgia, USA, 1997. .

[2] J. Accot and S. Zhai. Performance evaluation of input devices in trajectory-based tasks: An application of the steering law. In *Proc. of CHI'99*, Pittsburgh USA., 1999. http://www.informatik.uni-trier.de/ ley/db/indices/a-tree/z/Zhai:Shumin.html.

[3] J. Accot and S. Zhai. Scale effects in steering law tasks. In *Proc. of CHI2001*, Seattle, Washington, 2001. .

[4] J. Accot and S. Zhai. More than dotting the i's – foundations for crossing-base interfaces. In *CHI2002*, Minneapolis Minnesota USA, 2002. http://www.informatik.uni-trier.de/ ley/db/indices/a-tree/z/Zhai:Shumin.html.

[5] J. Accot and S. Zhai. Refining fitts' law models for bivariate pointing. In *Proc. CHI2003*, Ft. Lauderdale Florida USA., 2003. http://www.almaden.ibm.com/u/zhai/papers/BivariateFitts.pdf

[6] F. Avanzini and D. Rocchesso. Controlling material properties in physical models of sounding objects. pages 91–94, La Habana, Cuba, Sept. 2001.

[7] S. A. Brewster, P. C. Wright, and A. D. N. Edwards. An evaluation of earcons for use in auditory human-computer interfaces. In *INTERCHI '93*, pages 222–227, April 1993.

[8] P. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, Vol. 47(No.6), June 1954.

[9] F. Fontana, D. Rocchesso, and E. Apollonio. Acoustic cues from shapes between spheres and cubes. pages 278–281, La Habana, Cuba, Sept. 2001.

[10] F. Fontana, D. Rocchesso, and L. Ottaviani. A structural approach to distance rendering in personal auditory displays. In *Proc. International Conference on Multimodal Interfaces (ICMI'02)*, pages 33–38, Pittsburgh, PA, Oct. 2002. IEEE.

[11] N. Friedlander, K. Schulter, and M. Mantei. Bullseye! when fitts' law doesn't fit. In *CHI1998*, Los Angeles CA USA, 1998. http://portal.acm.org/

[12] W. Gaver. *Using and Creating Auditory Icons*, pages 417–446. Addison-Wesley, 1994.

[13] W. W. Gaver. The sonic finder: An interface that uses auditory icons. the use of non-speech audio at the interface. Austin, Texas, 1989.

[14] W. W. Gaver. Synthesizing auditory icons. In *INTERCHI '93*, pages 228–235, April 1993.

[15] Y. Guidard, M. Beaudouin-Lafon, and D. Mottet. Navigation as multiscale pointing: extending fitts' model to very high precision tasks. In *CHI99*, Pittsburgh PA USA, 1999.

[16] I. MacKenzie. *Fitts' law as a performance model in human-computer interaction.* Doctoral dissertation, University of Toronto, Toronto, Ontario, Canada., 1991. .

[17] I. MacKenzie. *Motor Behavior Models for Human-Computer Interaction, Chapter for J.M.Carrol*, pages 27–54. Morgan Kaufman, 2002.

[18] I. MacKenzie, A. Sellen, and W. Buxton. A comparison of input devices in elemental pointing and dragging tasks. In *CHI91*, New York, 1991. http://www.billbuxton.com/fitts91.html.

[19] M. N., S. B., P. D., S. P., and M. S. Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres. *Journal of the Acoustical Society of America*, (103):3005 (A), 1998.

[20] N. Orio, N. Schnell, and M. Wanderley. Input devices for musical expression: borrowing tools from hci. In *NIME01*, Seattle USA, 2001.

[21] S. P., M. S., and W. S. A multidimensional technique for sound quality assessment. *Acta Acustica*, (85):650–656, 1999.

[22] A. Pirhonen, S. Brewster, and C. Holguin. Gestural and audio metaphors as a means of control for mobile devices. In *CHI2002*, Minneapolis Minnesota USA, 2002. http://www.dcs.gla.ac.uk/ stephen/papers/CHI2002.pdf.

[23] M. Rath and D. Rocchesso. Informative sonic feedback for continuous human controlling a sound model of a rolling ball. In *Proc. of International Workshop on Interactive Sonification*, Bielefeld, 2004.

[24] M. S., S. P., K. J., M. N., P. G., R. V., and S. B. A meta-analysis of timbre space: Acoustic correlates of common dimensions. *Journal of the Acoustical Society of America*, 2003. In preparation.

[25] M. S. and W. S. A meta-analysis of timbre space: Multidimensional scaling of group data with common dimensions, specificities, and latent subject classes. *Journal of the Acoustical Society of America*, 2003. submitted.

[26] R. Vertegaal. *An Evaluation of input devices for timbre space navigation.* Doctoral thesis, University of Bradford, 1994. http://citeseer.ist.psu.edu/vertegaal94evaluation.htm.

Studying usability evaluation to improve its practical utility

Erik Frøkjær
Department of Computing
Universitetsparken 1
DK-2100 Copenhagen
+45 35321456
erikf@diku.dk

Kasper Hornbæk
Department of Computing
Universitetsparken 1
DK-2100 Copenhagen
+45 35321425
kash@diku.dk

ABSTRACT
Current research on usability evaluation has several limitations, including focusing on the evaluation outcome and on counting usability problems; a realistic understanding of how usability evaluation is used in practice has been largely ignored. We describe some of our recent work on addressing these limitations, including a diary study of evaluation processes, studying developers' assessments of usability problems, and generating redesign suggestions instead of problems. In addition, we speculate on future research that aims to address the limitations.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Evaluation/Methodology; D.2.2 [Software Engineering]: Design Tools and Techniques—User Interfaces.

General Terms
Measurement, Design, Experimentation, Human Factors.

Keywords
Usability evaluation, redesign, think aloud, metaphors of human thinking, empirical study, usability inspection, diaries.

1. INTRODUCTION
Research in usability evaluation is in a peculiar situation. On the one hand, a substantial number of studies have compared evaluation techniques and developed new ones, for reviews see [2-4]. On the other hand, several authors have pointed to severe limitations of that work. Gray & Salzman [3], for example showed how an often-cited selection of comparisons of evaluation methods suffers from low validity. Problems were identified not only with the statistical tests and the conclusions passed on to practitioners and researchers, but also with the measures used in the comparison of methods. A study of usability evaluation in industry also found that different teams of evaluators identified different usability problems [12]. Recently, Dennis Wixon [14] made the case that comparisons of UEMs do not appreciate that the real goal of such methods is to impact design, not to generate problems. Thus, comparisons may miss to assess properly the practical utility of UEMs.

Recently, we have begun to address these limitations by moving beyond some of the common ways of doing usability research. These include (a) disregarding the evaluation process and focusing only on the outcome of the evaluation, typically on sets of problems; (b) ignore the practical taking up and use of the evaluation's results in realistic system development contexts, and (b) counting usability problems, rather than investigating other outcomes of usability evaluation.

Below we describe studies that each moves beyond one of these common practices. In the final section we speculate on some possible research in which to further addressing the limitations of current usability research.

2. A DIARY STUDY OF EVALUATION PROCESSES
One way to study the evaluation process in more details is to have participants keep a diary. Diaries have previously been used to study the use of evaluation techniques [10,11]. In a recent paper [9] we described a study comparing two psychology-based inspection techniques, cognitive walkthrough and metaphors of human thinking (MOT). In comparison to the existing use of diaries, that study used diaries to compare evaluation techniques and combined diaries with quantitative measures of evaluation performance. In comparison to most of the existing literature in usability evaluation, the study not only focuses on the sets of problems that result from the evaluation, but also on the evaluation process.

Twenty participants evaluated web sites for e-commerce while keeping diaries of insights and problems experienced with the techniques. We will here only discuss the insight into the evaluation process that the analysis of the diaries gave. Especially two findings are relevant to mention.

First, the diaries show that usability problems are found in a variety of ways, not just by using the techniques as prescribed. At least ten participants identify problems already before reading the description of the inspection technique, or while initially orienting themselves on and gaining an overview of the web site. One participant writes, during her first visit on the web site before starting the evaluation procedure:

> Identification of immediate problems and some ideas for tasks. Especially the questionnaire [on the web site] is a disaster. The menu in the left side sometimes disappears. No systematic information on whether a word or a label is clickable…

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

That participant ends up reporting on her problem list three problems regarding the questionnaire. Even after finishing the evaluation procedure, some participants continue to identify problems.

Thus, participants seem to identify problems in many ways, not only through the techniques, reflecting large differences in individual working styles.

Second, during the course of the evaluation participants change their opinion on what they consider a usability problem, e.g. some participants change their opinion about problems when redesigning. One participant writes that

> [I] have come to the conclusion that the buying procedure is really not so complicated that it will give errors for the user.

The same participant had on his problem list noted as a serious problem the cumbersome buying procedure. Conversely, at least five participants identify problems when redesigning, problems they had not previously been aware of, for example:

> Looking at a screen dump makes me aware of new usability problems. What am I to do with problems I have just discovered?

These observations, and other from the paper [9], suggest that the process of usability evaluation are complex, somewhat disordered, and shaped to a high degree by participants' personal working habits. These findings appear to challenge common assumptions of the evaluation process as an orderly progression of steps that reflect the technique being used.

3. DEVELOPERS' ASSESSMENTS OF USABILITY PROBLEMS

In a couple of experiments, we have studied how developers assess usability problems. The main argument underlying these experiments is that developers' assessments heavily influence if a problem is addressed. Developers have a vested interest in minimizing redesign in order to meet time and cost-constraints and thus may be inherently biased in their assessment of usability problems. In practice, however, these are the circumstances that determine which and how problems are addressed.

In one study [8], MOT was compared to heuristic evaluation (HE). An experiment was conducted in which 87 novices evaluated a large web application. Of particular interest here is that the key developer of the web application assess the problems uncovered by MOT as more severe on users and also appeared more complex to repair than the problems uncovered by HE. The key point here is that the developer's assessment of usability problems helped identify differences between techniques; such differences could be relevant when selecting which technique to use.

In another study [7] we investigated how developers of a large web application assess output from usability evaluation. Problems and redesign proposals were generated by 43 evaluators using an inspection technique and think aloud testing. Of particular interest here is that developers' assessments of problems and our subsequent interviews with them provided insights in some of the reasons for taking up or ignoring a problem. For example, developers expressed that those problems which could be fixed easily and quickly were of particular utility. One developer explained:

> Typically if something can be easily and quickly fixed … that is a suggestion which requires four months of development is not as useful as some small suggestion, which corrects a small problem in 10 minutes, then I can correct it immediately

During all interviews, we asked developers if they could recall usability problems and redesign proposals. Usability problems were mostly remembered by developers as classes of problems, the particular instances was forgotten. One developer said that 'yes, there are several of them [usability problems] that I can still remember' and then—surprisingly—went on to expand on how specific redesign proposals on exploring similarities to standard search engines could be incorporated in the design. In contrast, all developers were able to describe in some detail redesign proposals which they had found interesting.

Another interesting finding was that developers find the problems identified to be mainly confirmations of issues they already know. In a comparative usability evaluation, Molich et al. [12] similarly found that only 4% of the problems identified were new to the usability team responsible for the system evaluated. One immediate reaction could be that this is not much. Yet, maybe we should be careful in concluding that developers get few new insights from usability evaluations. The developers in our study actually used the usability problems, and their thinking about the application seemed to have been influenced. Further, developers who for years have worked intensively with the application and its use context will not easily take up results of usability evaluations. On the contrary, changing their understanding is a process requiring time, during which new insights does not appear as something distinct and immediately clear. Rather, developers will experience nagging doubts, small changes in thinking, and challenges to their understanding. Studying how this develops over time would probably give a more valid picture of the impact of usability evaluations.

4. REDESIGN SUGGESTIONS AS SUPPLEMENTS TO PROBLEMS

Usability problems predicted by evaluation techniques are useful input to systems development; it is uncertain whether redesign proposals aimed at alleviating those problems are likewise useful. We have recently investigated this by having developers of a large web application assess usability problems and redesign proposals as input to their systems development [7]—the study also mentioned in Section 3.

Developers assessed redesign proposals to have higher utility in their work than usability problems. In interviews they explained how (a) redesign proposals help understand usability problems, i.e. redesigns contribute to characterizing and making more concrete the problems found, and illustrate why problems are important; and (b) redesign proposals are useful for inspiration and for seeking alternative solutions for problems that the development team has been struggling with. Point (b) is exemplified in the following quote from one of the developers:

> in some situations you may do things one way or the other, and then you can just choose, i.e. whether some list should be alphabetical or just split up…in other situations, like the

> three level hierarchical selection of job titles, no matter what we do we get into some complicated mess…so if one can find some way of making it more intuitive and usable than other ways, then we accept it eagerly, [because] we haven't quite figured out how to do it ourselves

The usability problems supported prioritizing ongoing development of the application and taking design decisions. One developer said that

> usability problems … what one cares about is the extent of them, how many is saying that some thing is a problem and how many is saying that some other thing is a problem, that help me prioritize what I should focus on

These comments do not mean, however, that developers did not appreciate usability problems, especially when they are well argued, clearly described, documented, and easy to fix. On the contrary, all developers wanted both problems and redesign proposals to form part of the input from usability evaluation to systems development.

5. FUTURE WORK

We are continuing to experiment with the problems and ideas introduced above. In addition, we wish to share a few further ideas for going beyond some of the limitations noted above.

The matching of usability problems underlies most usability research. Most usability problems are brief, often quite difficult to understand, and certainly incomplete in expressing the evaluators' thoughts, a problem of understandability more generally discussed in e.g. [13]. Thus, matching of just such problems descriptions form an insecure foundation for usability research. One example of this is the study by Molich et al. [12], where problems founds by different professional teams are matched. Another example is studies of the so-called evaluator effect [5], i.e. the observation that evaluators typically find different usability problems. In both cases, it could be worthwhile to explore if this matching really is sound. This could be done, for example, by (a) involving evaluators more in the matching process. Hertzum et al. [6] finds an interesting difference between 'objective matching' of usability problems and the opinions of the usability specialist who had produced the problems; (b) include different representations in the matching, for example both usability problems and suggested redesigns, and (c) study how this matching goes on in practice, to see if what we think are similar or different problems function in the same way for development teams.

When one looks to the literature on usability evaluation in industry, the results are surprisingly meager. More studies of industrial systems development could help us understand (a) what output is useful from evaluation techniques, and (b) at which stages different evaluation techniques give the best results.

Finally, in-depth studies of evaluation processes seem to give interesting data on evaluation performance. Above we gave one example on diary studies. Another example of an in-depth approach that gave interesting data is Boren and Ramey's [1] study of think aloud. They showed how practical think aloud studies are often far from the original content of the think aloud methodology.

REFERENCES

1. Boren, M. T. & Ramey, J. Thinking Aloud: Reconciling Theory and Practice, *IEEE Transactions on Professional Communication*, *43*, 3 (2000), 261-277.
2. Cockton G., Lavery, D., & Woolrych, A., Inspection-Based Evaluations, in Jacko, J. A. & Sears, A. *The Human-Computer Interaction Handbook,* Lawrence Erlbaum Associates, 2002, 1118-1138.
3. Gray, W. D. & Salzman, M. C. Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods, *Human-Computer Interaction*, *13*, 3 (1998), 203-261.
4. Hartson, H. R., Andre, T. S., & Williges, R. C. Criteria for Evaluating Usability Evaluation Methods, *International Journal of Human-Computer Interaction*, *13*, 4 (2001), 373-410.
5. Hertzum, M. & Jacobsen, N. E. The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods, *International Journal of Human-Computer Interaction*, *13* (2001), 421-443.
6. Hertzum, M., Jacobsen, N. E., & Molich, R. Usability Inspections by Groups of Specialist: Perceived Agreement in Spite of Disparate Observations, *Proc. CHI2002*, (2002), 662-663.
7. Hornbæk, K. & Frøkjær, E. Comparing Usability Problems and Redesign Proposals as Input to Practical Systems Development, under review (2004).
8. Hornbæk, K. & Frøkjær, E. Usability Inspection by Metaphors of Human Thinking Compared to Heuristic Evaluation, *International Journal of Human-Computer Interaction*, *17*, 3 (2004), 357-374.
9. Hornbæk, K. & Frøkjær, E. Two Psychology-Based Usability Inspection Techniques Studied in a Diary Experiment , *Proc. Nordichi 2004*, ACM Press (2004)
10. Jacobsen, N. E. & John, B. E. Two Case Studies in Using Cognitive Walkthroughs for Interface Evaluation, *CMU-CS-00-132* (2000).
11. John, B. E. & Packer, H. Learning and Using the Cognitive Walkthrough Method: a Case Study Approach, *Proc. CHI'95*, ACM Press (1995), 429-436.
12. Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. Comparative Usability Evaluation, *Behaviour and Information Technology*, *23*, 1 (2004), 65-74.
13. Naur, P. *Knowing and the Mystique of Logic and Rules*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1995.
14. Wixon, D. Evaluating Usability Methods: Why the Current Literature Fails the Practitioner, *interactions*, *10*, 4 (2003), 29-34.

Examining the Use of Usability Results in a Software Development Company

Rune Thaarup Høegh
Dept. of Computer Science
Aalborg University
+45 96358894
runethh@cs.aau.dk

Jan Stage
Dept. of Computer Science
Aalborg University
+45 96358914
jans@cs.aau.dk

ABSTRACT
This paper presents the first results of a study on a usability evaluation for a Danish software development company. The use of the results from the usability evaluation is examined through interviews with two developers from the software company. It is through an interview with a project leader from the company found that the traditional usability report plays a very little role for the development team. Initial results suggest that textual feedback proves more valuable when accompanied with video and oral feedback.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation (e.g. HCI)**]: User interfaces- *Evaluation/methodology*

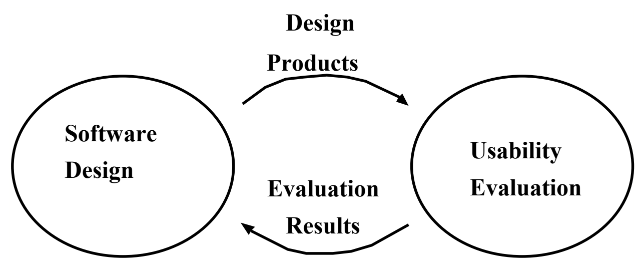
General Terms
Theory and methodology. HCI

Keywords
Usability results, evaluation, Human Computer Interaction.

1. INTRODUCTION
It is a well-known problem in traditional software development that some systems fails to fulfill their goal or has serious limitations [1, 6]. The limitations include that the system does not adequately support the core tasks of the user and unsuitable designs of user interactions and interfaces [3]. These particular problems can be addressed by including usability evaluations as a part of the development.

The interplay between usability evaluation and design can be understood as depicted in figure 1. Two circles represent the design and the evaluation activities. A development team designs a part of the software, and the design is implemented to a product. This product is then usability evaluated by a usability team giving a set of usability results. These results are used to improve the quality of the design, and the circle starts again. The work in this paper seeks to examine the relationship between the usability evaluation and the development of the software. It examines the arrow called "evaluation results" in order to learn if the results are optimal for the development team

Figure 1 – The interplay between iterative design and



usability evaluation

The HCI-lab at Computer Science at Aalborg University has conducted several usability evaluations for companies that wanted to improve their products and make their software user-friendlier. The typical evaluation is conducted by using the think aloud protocol [4]. In such evaluation a set of users are invited into the HCI-Lab to evaluate the software. The user asked to speak out their thoughts while going through a series of assignments taking him through selected parts of the software. The user's interaction and words are recorded on video for further analysis. Some times one or more developers from the software development company are invited to observe the evaluation. The end product of a usability evaluation by the think aloud protocol is a rapport. In that rapport is a summery of the usability issues identified along with a rating of the problems.

This paper follows up on a usability evaluation and examines how the result of the evaluation was used in the software development company. The software company's name is kept anonymous.

2. METHOD
In the fall of 2003 a Danish software company had their software usability evaluated at HCI-lab at Aalborg University. The usability evaluation was conducted using the standard think aloud protocol, and two from the software development team were invited to observe the usability evaluation. They observed six sessions of usability evaluation with different users. After a formal analysis conducted by the usability team, a usability report was sent to the software development company. A few months afterwards the two participating developers were interviewed about how the usability evaluation was used by the software

Fourth Danish HCI Research Symposium, November 16, 2004 Aalborg University, Denmark.

company. They were interviewed individually using a semi-structured interview [2], such that interviews could be compared while still having the option to follow and investigate an interesting line of topics. The interviews were focused on examining how the usability evaluation had influenced the development of the software, and how the usability rapport was used when the software company received it.

This paper presents the results of the analysis of the first interview. The interview was transcribed and analyzed using the Grounded Theory approach [5]. The interview contained a number of statements about how the usability evaluation had effected the development. Using grounded theory analysis a set of categories was found using. These categories where then merged into four general categories. The general categories are presented below

3. RESULTS OF THE INTERVIEW

The four overall categorizes outline the most important issues that emerged from the interview. Each of them describe different aspect of the usefulness of a usability evaluation, but they also present area where improvements can be done.

3.1 The effects of observing an evaluation

The developer that observed the usability evaluation found it to be a very honest experience. The honesty in the experiences consisted of the very powerful images of seeing a user using the software she had designed while being guided by a professional usability team. Seeing the user struggle with certain parts of the software made her realize the usability issues and at the same time get a feeling of how they could be solved. She did however notice that only the most severe issues was remembered later on, and the less critical issues where faster forgotten.

Another benefit observing an interview was that she could remember which part of the user interface that had critical usability issues and why. This meant that she could picture the user use the system when correcting the usability issues and could therefore do a better job on correcting them. Having seen the evaluation also meant that she could explain the usability issues to the development crew better because she did not need to rely solely on the text from the usability report. She could use her memory to remember the issues and therefore relate to them in a better way.

3.2 The immediate effect of the evaluation

The immediate result of the usability evaluation was that the observing developers went back to their company and created a list of the five to ten most severe usability issues based on their own experience. This list was then given to the development team and the issues were addressed in the software. Observing the evaluation therefore meant that the development team could initiate the redesign of the problematic areas the day after the evaluation based on their own observations.

The observing developers also saw that many problems were in the software's dialog. They therefore decided to inform the development team and the local evaluation team to have special focus on this area when testing the software in the future. The project leader also decided that usability evaluations should be included when possible in future software development.

3.3 The use of the usability report

The formal result of the evaluation was the usability rapport. This was given to the company four weeks after the evaluation after a thorough analysis. The report included a number of usability issues and described where they were found along with a rating of their severity. The list was long including more than 50 issues of different severity.

The development team had not used the usability report to improve the software. The usability report was read and used mostly for future reference such that the developer knew what to avoid in future designs. The reason that the list was not gone through in any systematic manner was that the most important issues were already addressed at the time when the report arrived. The remainder of the problems was not addressed in the design for two reasons. When reading through the report the developer could recognize only the most important issues and she felt that the report sometimes described a problem without her knowing how to solve it. The usability report did however remind her of some of the problems she had forgotten while the design team only focused on the most critical issues.

3.4 Suggestions for improved feedback

When asked about the usefulness of the usability rapport, the developer told that it was a resource to avoid similar problems in future designs. The benefit of the report for what was the current design was limited, partially because of the way the feedback was given only in text describing problems and issues.

After having worked with a problem the developer felt unsure if the change she had in mind for the user interface would actually solve the usability issue. Because of that she suggested that the feedback was extended from only being text to also include verbal feedback. This verbal feedback should be in the form of counseling with the usability team after the development team had worked with the issues for some time. This counseling would have the two purposes. One would be of a creative kind, to help the development team to get ideas how to improve usability issues that the development team could not come up with a better solution for. The other purpose would be to go through changes in the user interface in order to evaluate if the changes had solved the problems – without having to go through a new usability evaluation first.

4. DISCUSSION

The result of the analysis of the first interview shows that it is of a great benefit for the development team to have someone observing the usability evaluation. This allows the observer to gain a better insight in the usability evaluation and the team has knowledge of the most important issues immediately, which means they can address the most important issues straight after evaluation. One could be concerned that they development team would focus on issues that were not analyzed yet, but the most critical issues where in this case easy to spot, and there was no doubt if they were worth redesigning or not.

It is worth considering if the usability would benefit even more from getting a copy of the raw videotapes straight after the usability evaluation. With such videotape the entire team would be able to get this very honest feeling it is to see their software in use and being evaluated. The videotape would also be a benefit when the development team received the formal feedback from the usability team. They would then be able to browse through the videotape and see the issues mentioned in the usability report. The risk is however that the amount of videotape overwhelms the development team, because there is easy more than 5 hours of video for a standard evaluation. For that reason an edited tape might be of more help.

The development team used the actual usability report very little, that was partly because they had already addressed the most important issues. The five to ten most severe problems were addressed, but the list with issues was a lot longer than that. The development team may have felt overwhelmed by the list of issues, or they may now have had time or resources to address them. It was also mentioned that they sometimes not had a good idea how to improve the design to solve an issue. It was suggested that the feedback could be extended to also include a session where the development team could discuss possible improvements with the usability team. This would mean that the usability team would have to step out of the role as objective evaluators and into the role as consultants.

5. LIMITATIONS

The study is based on a single interview with a developer from one Danish company. To gain more certain knowledge further studies must be done including an analysis of the other developer, but also including other companies. Expanding the study will reveal if the identified issues are only related to the single company or if the issues influence development with usability evaluations in general.

6. ACKNOWLEDGMENTS

Thanks to the participating employees from the Danish software company and thanks to the participants in the usability evaluations.

7. REFERENCES (skal ordnes alfabetisk)

[1] Frøkjær, E. & Korsbæk H., Informatization Policies in Denmark, Chapter 3 in P.H.A. Frissen et al (Eds): European Public Administration and Informatization, IOS Press, Elsevier, Amsterdam, pp. 25-47, 1992

[2] Kvale, S. Interview : En introduktion til det kvalitative forskningsinterview. Kbh.: Hans Reitzel, 1997.

[3] Landauer, T. K., The trouble with computers: usefulness, usability, and productivity, MIT Press, 1996

[4] Rubin, R. Handbook of Usability testing: How to Plan, Design and Conduct Effective Tests. John Wiley & Sons, Inc. 1994.

[5] Strauss, A. Corbin, J. Basics of Qualitative Research. SAGE Publications, 1998.

[6] Teknologirådet, Bonnerup rapport, 2001

How usability practitioners can get their suggestions implemented in industrial software design

Georg Strom

DIKU, University of Copenhagen
Universitetsparken 1
DK-2100 Cph. O., Denmark
georg@diku.dk

ABSTRACT

This paper describes barriers that make it difficult for usability practitioners to get improvements implemented in interfaces, and it describes how work with usability and HCI best is fitted into product planning and software development. It is based on experiences from eleven years of practical usability work with 27 different projects including the development of mobile phones, professional electronic equipment, large professional systems and web sites and web-based applications.

Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Design, Economics, Human Factors

Keywords

Industrial software development, marketing, system design

1. Introduction

HCI and usability professionals have described the value of usability in numerous interviews and popular articles and in a number of books. However, it appears that these efforts have not led to the desired result: It is still common that managers of software projects do not include any user studies, usability tests or evaluations when they plan a project. Therefore it is worth investigating what the actual problems are and how they can be solved.

2. Method

This paper is based on my own experiences from eleven years of work with interface design and usability in two large companies (1987-1992 and 1996-2002). My work included field studies, interface design, coaching of interface programmers and usability tests. I worked on interfaces for professional electronic equipment, mobile phones, large software systems and web-sites and web-based applications. I have worked on at least 27 different projects, and I have worked with 19 different organizational units (7 development groups, 3 marketing groups, 4 sales groups and 5 internal support groups). To support my memories about the projects I have used notes where I had listed projects I worked on and activities I had done in each project.

The method used in this paper is close to the ideal anthropological study where the researcher is a castaway (shipwrecked) in foreign culture, forced to learn about it in order to survive, and in such a position that his or her presence as an observer does not influence the events [3]. Compared to interviews and brief observations, this method gives an understanding of a foreign culture (in this case that of industrial software development) that is deeper and more diverse.

3. It was difficult to get interaction improvements implemented.

In both companies a substantial part of all projects were done without any planned or professional usability work. In more than one third of the projects where I did usability work, my work did not result in any perceptible improvements of the interaction. See table 1. This is a high failure rate; in particular when taking into account that most of my work consisted of using well-defined methods to produce information that was requested by managers of the organizational units.

4. The value of usability and HCI was in general accepted

My lack of impact was in spite of a general positive attitude towards usability. The first company had a culture with a strong customer- and user-focus. In the second company I gave lectures and briefings, and I felt my effort was supported. It appeared that most people, including managers and software developers, had

Table 1: Distribution of results of usability projects analyzed in this paper

	Dev (eng.) group	Marke-ting group	Sales group	Int. support group
Very successful interaction implemented		1	1	
Perceptible improved interaction implemented	4	3	2	5
None or only insignificant improvements implemented	6	1	1	1
Usability work not completed	2			

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

been annoyed by usability problems.

In both companies sales managers and product managers, a marketing function, had to keep institutional customers as satisfied as possible, which included that they had to solve or smooth out any usability problems reported by customers.

Software developers in both companies often wanted to discuss their own theories about what might cause usability problems, and they expressed an interest in usability that seemed related to the widespread and common fascination with psychology and human thinking. I did not experience the hostility against usability that is reported by Cooper [1].

5. Other factors are often more important than usability

I experienced three web projects with fixed completion dates and tight schedules, but with only very general requirements, and where the project plans did not include any studies of the users' needs or any usability tests. I later did usability tests on two of the developed web sites, when a number of users had complained that it was difficult to use them (and my tests revealed significant problems).

However, from an organizational point of view the priorities made sense. If the management had opened for discussions and not set a fixed completion date, the projects might have dragged on for a long period and possibly never been completed. The project group knew at the same time that they only would get the assignment, if they agreed to a tight time schedule. If you believe that almost anything is better than nothing, it makes sense not to spend time on usability work.

I worked on a number of version 1 projects: Projects to develop software with new and valuable functionality that was not available from any competitors. In such cases the highest priority was to get a software without technical problems out as fast as possible. It is known in marketing and sales that customers will buy such software because they need the functions, almost no matter how bad the usability is.

I experienced in both companies how a core concept in product planning was the product live cycle [2] and the so-called window of opportunity: The time from you launch a product until you must cancel it, because it cannot compete with new and better products. The return on the money invested in the development depends on the length of the window, such that managers must avoid any usability work that may delay the release of a product.

I saw a number of cases where customers focused more on price or style than on usability. In two cases customers even preferred less usable but more stylish versions of a product. More usable products had probably resulted in lower sales.

I found that usability was considered important for the success of a product, when:

- The product was bought by organizational customers who demanded a certain level of usability. Such customers could demand modifications or threaten with legal action if they found that the usability was unsatisfactory.
- There were a number of competing products that offered the same functionality, and there was a public discussion of the usefulness of different products. I experienced that was the case for mobile phones (as it is for professional digital cameras and for software for web or graphical design).

Even though a higher level of usability may be valuable for the society in general, it is sometimes more advantageous for the developing organization to get something out fast; in some cases an increased focus on usability may even reduce the sales and the company's profits.

6. It is organizationally difficult to implement suggested improvements

I was surprised and frustrated by the many cases where I did usability tests or studies, and where almost none of the identified improvements were implemented.

In a number of cases, improvements to the interface required changes to the requirements that already were agreed with the customer, which meant that an implementation of the improvements required a new negotiation of the contract. That was difficult and increased the uncertainty of the project, so both project management and customers wanted to avoid it.

I was sometimes asked to test an interface that was under development. However, it is only possible to test an interface when the development of it is almost finished, and at that point there was only time to make minimal changes, for instance to change the wording of a text that users did not understand.

A similar problem occurred in some cases when I was asked to do a usability test when a project was finished. The project group agreed that the results seemed reasonable and that some changes should be done to the interface. However, it was not possible to get funding for a second project, which should correct the errors made in the first one.

All software projects I experienced was based on a platform or basic application, they were not developed from scratch. The selection of a platform was a high-level decision, it appeared to be based on an evaluation of the company who delivered it, the price of the software license, and the technical functionality offered. I did not observe any cases where possible usability problems were evaluated before a platform was selected, and even if there was a realistic choice between two or more platforms, it is likely that each came with a number of built-in usability problems.

Often problems identified in a usability test could only be fully solved by making changes in the platform or basic application. Even if it was possible to suggest such changes, and even if the suggestions were accepted, they could not be implemented before the project was finished and the result released.

Ironically, I have several times met usability professionals who worked on platform or basic application software, and who complained that it was almost impossible for them to do user studies, because their company had no direct contact with end-customers or end-users.

7. The position in the organization is crucial

When I analyze my results, it is clear that my success rate was strongly influenced by where I was situated in the organization. See table 1. I will therefore describe the advantages and disadvantages of being situated at different positions.

7.1 A separate usability unit is isolated.

In a period of almost three years I worked in an internal consultancy department in the second company; in essence I ran my own dedicated usability group. I experienced that such an organizational position made usability work more difficult:

- In a cost-center structure, which today is common in companies, usability work done by a separate department is a visible and additional cost for the group who requests it. This makes it more difficult to initiate new usability projects.
- It is difficult to get information about new projects. I spend a substantial amount of time trying to trace down coming projects, and in spite of that I frequently only got involved after the point where my results had been most valuable.

However, I found that a separate usability department or group is necessary when the organization consists of a number of small product groups. It is then not feasible for each group to employ their own usability specialist.

7.2 It is difficult to get improvements implemented in a development group

My experience is that the methods and priorities of software development make it difficult to do usability work as part of a development group.

I observed how industrial software development is focused on delivering software that fulfils a set of defined goals before a specific time and below a certain cost. This means that any sensible project manager will try to reduce the complexity and uncertainty and thereby the risk of not meeting the goals. In contrast, user studies or usability tests introduce new information, which in itself increases the complexity of the software development; it suggests that some of the design goals shall be changed and may even suggest that some completed work is modified, increasing the uncertainty and risk of the software development.

As described earlier, the situation is even more difficult, when the proposed changes only can be done after negotiations with the customer and changes to the agreed requirements.

The consequences were that my work in development departments in general only resulted in minor improvements. In particular it was never possible to change the structure of an application, even when it was well documented that the structure created usability problems.

In contrast, I found it very valuable when interface programmers wanted an informal review of their ongoing work. We would then discuss possible problems in the interaction in details, and I observed how a number of possible usability problems could be identified and solved during the design process.

7.3 Marketing and usability speak the same language

When I started working with usability and interface design in the first company it was as product manager in a marketing department. In the second company I did some of my most valuable work in or together with marketing departments.

I found that doing usability work in a marketing department offered several advantages:

- I was part of the group that made the decisions. I found that the easiest and most effective way of ensuring a sufficient level of usability was to get it defined as precisely as possible in the requirements for a new product (including how the level of usability should be tested).
- There was enough time to do usability work. It was possible to do a usability test of the last version of a product and to discuss what should be changed in the next one, and it was possible to do user studies as an ongoing process without being restricted by the schedules of different development projects. That was in particular necessary when the software consisted of components that were developed partly in parallel and combined in different ways to products, which were launched at brief intervals.
- Whereas computer science has a mathematical and technical background, both marketing and usability have adapted methods from psychology and anthropology; both use personas, scenarios, prototypes and interviews. I found it easy to discuss with marketing people, and I found that our work fitted well together. (The main difference was that I focused more on ease of use, whereas they focused more on the user experience.)

However, I found that it was essential that software designers, hardware designers and other relevant developers participated in the group that defined new products. They often had a detailed knowledge about earlier products, and without them it was not possible to take all technical limitations into account.

I also found that it was important that there was an ongoing informal dialogue between marketing and usability people and the individual developers. Without such a contact details were often misunderstood or implemented in a less than optimal manner.

7.4 Sales groups focus on users

As earlier described, design details were often described in a requirement specification that was agreed with the customer. At one time I discussed with a sales group whether it was possible to do user studies before a proposal was made to a customer. They agreed that it was highly advantageous to know the needs and possible situations of use before a proposal or an agreement was made, but were hesitant because user studies at that stage might create legal problems in relation to the current rules for bids and proposals.

It was the task of a sales group to ensure a successful delivery to the customer. The sales group ran a delivery project, which might include training of customer staff, installation, configuration of each system, production of manuals and user guides and necessary technical support.

Each delivery project had its own budget; it was a profit as well as a cost-center, which made it easier to get money allocated to usability work.

I found that it was possible to improve the usability substantially during several delivery projects. Even when it only was possible to make minor changes to the interface, it was possible to make it easier for users to cope with it.

7.5 Internal support groups can be pressured

My analysis shows that the work I did for internal support groups was fairly successful:

- I was mostly asked to do usability work when a group felt they had a problem and that some changes had to be done.
- Users who complained were often supported by their managers and even by the top management. Even though they were forced to use the corporate systems, they had a certain amount of power.

8. The choice of methods does not influence the number of implemented improvements

When I started to write this article I thought about the occasions where I had discussed with a project manager whether a particular change should be made or whether a particular problem should be regarded as a minor or a major error. In such situations I frequently felt that I needed better arguments and more theories that I could use to substantiate my views.

However, when I went through the 27 projects analyzed in this paper, I found that that there was no relation between the quality of my arguments and how well they were accepted. In fact, the two situations where I was unable to finish my usability work were among those where I had the strongest arguments.

When managers or members in a project group thought it was possible to implement my results, they would accept them almost immediately and only discuss details. They would not discuss the methods, theories or observations my results were based on.

One reason was that my suggestions often were seen as common sense. I found that usability problems often appear to be obvious when someone has pointed them out. When I described how a menu-text could be misunderstood, it was rarely discussed whether it would be misunderstood.

In some cases where it was difficult to implement changes managers would also accept my results immediately, but the changes would not be implemented.

It was only in some of the cases with serious problems that it was almost impossible to solve, that managers and project managers asked in details about the methods and theories I had used to reach my results. In such cases I could not succeed, no matter how good my arguments were; The persons I discussed with were searching for something they could use to reject my results.

9. Discussion and conclusion

This paper is based on what I remember (with support from my notes). The projects were done at different times, they included different activities and were done in different organizational units, which make it easy for me to distinguish them. Daniel L. Shachter describes how the general outline or meaning of events tend to be remembered [4], such that my memory about the discussions in and relative success of each project should be fairly precise (even though I cannot remember all usability problems encountered in each project).

The usability work described in this paper was done in two leading software companies, and my own qualifications for doing it was above average. It is likely that other usability practitioners find it more difficult to influence software development and the design of interfaces.

The problems I encountered were in spite of a widespread interest in and acceptance of usability. It is therefore unlikely that a general promotion of usability will lead to a higher proportion of projects where usability work influences the design of interfaces.

It appears that better usability methods do not give usability practitioners more influence. (Even though the introduction of better methods can improve the value of the changes they suggest.)

I found that decision-makers that refused to implement the results of usability work in their interfaces in general had reasons that were rational and sensible considering the actual situation and their role in the organization.

Usability is only one of a number of aspects of a product, and often it is not the most important. It is even likely that some companies do usability work not because it can be justified by its value for the company, but because it is regarded as fashionable or good practice.

The goal of a development group is to deliver a product that fulfils a set of defined goals within a certain time and below certain costs. Usability work is then often seen as something that may delay the project and increase the complexity and risk of the project. In contrast, usability work done in parallel with contract negotiations in a sales organization or during the product planning in a marketing organization, can ensure that the goals of the project fits the needs of the users.

It is today common that software or web companies create new products by combining different components that are developed in parallel and more or less continuously. It is then not feasible to do usability work as an integrated part of different development projects: It must be done as a continuous process in close collaboration with marketing.

My experiences suggest that there are two ways in which usability practitioners can achieve more influence:

- Realize that usability work in some cases cannot add substantial value, and avoid spending any effort on them.
- Learn about marketing and organization and get into a position where it is possible to influence the planning of new products at as early a stage as possible.

10. Acknowledgements

Thanks to Torkil Clemmensen, Copenhagen Business School, who have helped with the discussions that have led to this paper.

11. Literature

[1] Cooper, Alan: *The inmates are running the asylum.* Sams, 1999.

[2] Davidson, Hugh: *Offensive Marketing.* Penguin, 1987

[3] Pratt, Mary Louise: Fieldwork in Common Places, in *Writing Culture, The Poetics and Politics of Ethnography,* ed. by James Clifford and George E. Marcus. University of California Press, 1986.

[4] Shachter, Daniel L.: *Searching for memory, the brain, the mind and the past.* Basic Books, 1996

eCell – making room for collaboration

Christina Brodersen
Department of Computer Science,
University of Aarhus
Åbogade 34
8200 Århus N, Denmark
+45 8942 5659
sorsha@daimi.au.dk

Christian Dindler, Ole Sejer Iversen
Department of Information- and Media Sciences,
University of Aarhus
Åbogade 34
8200 Århus N, Denmark
+45 8942 9257
dindler, sejer@imv.au.dk

ABSTRACT
In this paper we present the eCell; a temporary collaborative learning environment designed for the elementary school. The eCell incorporates a spatial as well as an IT design dimension supporting pupils engaged in project work. The eCell may be situated in the corridors, libraries or in other shared spaces of school environments.

Categories and Subject Descriptors
K.3.1, K.3.m [**Computing Milieux**]: Computer Uses in Education.

General Terms
Design, Experimentation.

Keywords
Project work, collaborative environments, IT supported education.

1. INTRODUCTION
The Danish elementary school system is changing as new educational visions are gaining ground and shaping pedagogical strategies and practice. There is a general movement within the educational practice of the elementary schools towards interdisciplinary project work where the pupils collaborate in groups. The project work aims at creating involvement and relevance in relation to the surrounding environment drawing on resources from society as well as school facilities. Moreover the project work form is characterized by reaching beyond the traditional boundaries of the classroom calling for a more flexible use of the school's physical space and resources. The pupils move between different locations, e.g. classroom, library, hallway etc., utilizing available resources and transporting materials across locations.

In order to support the collaboration of the project group new concepts of utilizing the physical spaces of the school are needed. The project work form calls for flexible spatial arrangements that support the possibility of forming temporary collaborative environments. Moreover the movement towards integrating technology in educational practice emphasizes the need for environments that support collaboration on digital materials that transcend the practice of pupils sitting in front of a traditional PC.

In this paper we present the eCell. The eCell is a temporary, collaborative learning environment where pupils are able to work with digital as well as physical materials. The eCell concept incorporates a spatial as well as an IT design-dimension. The eCell concept is part of the iSchool project within Centre for Interactive Spaces, ISIS Katrinebjerg; an interdisciplinary research centre bringing together architecture, engineering, and computer science with the research mission to create new concepts for future interactive spaces.

2. eCell
The eCell design consists of a private inner display and a public outer display. These displays combine to create a working environment for groups of pupils. The inner display is a pressure-sensitive large screen connected to an ordinary computer. The outer display consists of a projection on a large semi-transparent surface.

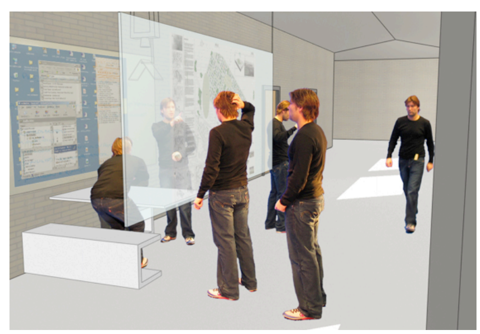


Figure 1 - visualisation of the eCell concept

The group occupying the eCell may work on projects or presentations on the private inner display. On the outer display the group may present their work as it progresses or information about them selves or their project. The outer display is represented on the inner display allowing the pupils to drag pictures or other material to the outer display. The material on the outer display may be organized as a slide show or as a collage. The activity of the group in the eCell is made visible for people passing by. Thus the eCell combines a private working environment for the groups with a public medium that communicates activities or inspirational material back to the surrounding school.

When the eCell is not used by a group of pupils the outer screen may be used as a message board or as a surface where passing pupils can display pictures.

The eCell may utilize some of the areas of the school not being used today e.g. the corridors can be turned into several temporary project rooms where pupils can work with projects and presentations.

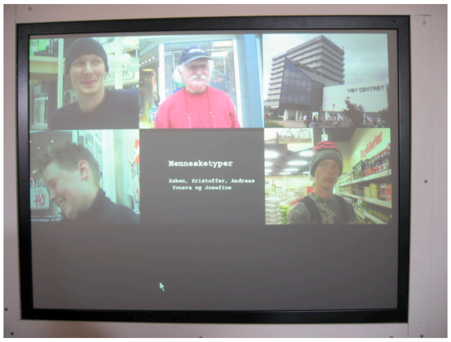


Figure 2 – collage on the public display presenting the project called "Different types of people"

Figure 3 – Pupils from 7. b working in the eCell

Moreover, the perspective of the eCell is to embrace school spaces as an integrated part of the learning environment. Through their project work the pupils gather and produce materials e.g. pictures, text or links to websites. The inner display of the eCell supports the pupils in producing and processing these materials. The outer display may be said to allow the pupils to give material back to the environment i.e. the school surroundings. The prospect is to create a learning space where the activities and cultural materials of the pupils become a visible and active part of the school environment. Teachers as well as other pupils may be inspired form the ongoing activities in the eCell and contribute to a more vibrant educational environment where material and traces of past activities are shared across different class levels.

3. RELATED WORK

Computer supported collaborative learning within educational environments is treated by numerous research communities such as the well established CSCW and CSCL environments. The Classroom 2000 [1] is a prototype classroom environment with special focus on the development of a software infrastructure to seamlessly capture much of the rich interaction that occurs in a typical university lecture. The CSILE research project [4], (Computer Supported Intentional Learning Environments) is an educational knowledge media system that functions as a "collaborative learning environment" and a communal database, with both text and graphics capabilities. This networked multimedia environment lets pupils generate "nodes," containing an idea or piece of information relevant to the topic under study. Nodes are available for other pupils to comment upon, leading to dialogues and an accumulation of knowledge. These projects both work with ubiquitous computing as ways of supporting education and learning, however they do not include the physical space as a factor IT supported learning.

In other pedagogically driven research projects, deployment of smart technology in Classrooms is conducted for further study e.g. [2]. These studies, however leaves out the pedagogical tradition of extra-mural learning (excursions, visits, interviews) and the innovative educational possibilities of digital media and ICTs.

4. CONCLUSION

The eCell is an example of technology that takes advantage of the unused spaces and turns them into a facility for group work. Furthermore, the eCell leaves traces of the ongoing and previous activities and consequently augments the room and gives something back to the physical environment it is placed in. Many attempts have been made to introduce IT into elementary school to support them as learning environments. The majority, however, are concerned with developing software for learning environment or the pedagogical ideals, and they do not take into account how the physical environment can and should be utilized.

5. FUTURE WORK

Several aspects of the eCell are currently being developed. In the first version, the inner display consisted of a large touch-sensitive screen attached to a traditional PC running Windows XP. This solution provided the pupils with the possibility of using most of the applications that they were used to. This platform is however inherently designed for single user interaction and is not the ideal tool for the collaborative setting. We are currently working on systems that will support multi user interaction making it possible for the entire groups to actively participate in the activities in the eCell. Concurrently with the development of this technology, we face the challenge of developing software that supports the collaborative aspect of the eCell environment.

Regarding the spatial design of the eCell crucial safety issues have emerged. One of the perspectives of the eCell is to utilize some of the otherwise unused areas of the school corridors. The fire authorities have relatively strict regulations regarding installations and objects that may hinder free passage through the corridors in the event of a fire. In accordance with this, the outer display of the eCell needs to be retractable in order to ensure free passage. We are currently working on a mechanical solution where the outer display may be retracted towards the ceiling when the eCell is not in use.

6. ACKNOWLEDGMENTS

This work has been supported by Center for Interactive Spaces, ISIS Katrinebjerg (Project #110). We are furthermore indebted to the pupils and teachers at Vestergårdsskolen in Århus for their valuable evaluation and design ideas for the first version of the eCell

7. REFERENCES

[1] Abowd, G. D (1999). Classroom 2000: An experiment with the instrumentation of a living educational environment. In IBM Systems Journal, vol 38, no. 4, 1999, p. 508 - 530

[2] Strong, B. & Kidney, D. (2002). Collaboratively Evaluating and Deploying Smart Technology in Classrooms (The McMaster University Model). In Proceedings of the EdMedia Conference, 2004, p. 175 - 179

[3] Ulicsak MH, Daniels H & Sharples M (2001) CSCL in the classroom: The promotion of self-reflection in group work for 9-10 year olds. Proceedings of the European Conference of Computer Supported Collaborative Learning 2001, Maastricht, pp. 617-624.

[4] "Scardamalia, M., & Bereiter, C. (1994): Computer support for knowledge-building communities. The Journal of the Learning Sciences, 3(3), 265-283

PDA as a 'Cognitive Prosthesis'

Malene Damborg
Institute of Psychology, University of Copenhagen
malene5@hotmail.com

Brith Klarborg
Institute of Psychology, University of Copenhagen
brithklarborg@hotmail.com

ABSTRACT
In this paper problems occurring in the interaction between a commercial Personal Digital Assistant (PDA) and individuals with acquired brain injury are addressed. Reduced long term memory, reduced attention span, and reduced executive functions are related to particular types of errors committed in an experiment by the brain injured users and some ideas for preventing these errors are presented. These ideas could be a field of future research.

1. INTRODUCTION
Individuals with acquired brain injury often experience increased difficulty with cognitive processes. For that reason they might find some relief in the use of cognitive artifacts such as various technological devices; they can use them as a sort of 'cognitive prostheses' [3]. When working within the field of assistive technology the importance of good design and usability becomes evident considering the users' vulnerability and special demands. In this study we have evaluated the usability and design of one such 'cognitive prosthesis': The use of the PDA 'Palm, Tungsten E' as a prosthesis for reduced long-term memory, reduced attention span, and/or reduced executive functions.

Most of the research on this area has concentrated on the advantages of various devices to this type of patients. And most studies do get good results applying technology onto this area of use. However, our intention was to uncover some of the *use related problems* arising from combining a commercial interface and people with acquired brain injury.

Our study is a collaboration with "Center for Hjerneskade, Amager", where we had the opportunity to test and evaluate a number of persons with brain injury learning to use PDAs to compensate for various deficits. Data were collected through interviews, a case study and an experiment designed to test their skills at using the PDA's most common programs and functions. The amount of data is too large to be presented in this paper; therefore we have singled out three examples from the experiment that we find representative of both the patients' deficits and the main issues involved in the interaction between people with brain injury and PDAs.

The three deficits on which we will focus are: reduced attention span (working memory), reduced long-term memory and reduced executive functions. These three cognitive functions contribute to the process of "prospective memory", also known as "the realization of delayed intentions" [1]. In other words, prospective memory is the ability to remember to carry out intended actions, remembering to remember [2; 5]. Since the process of prospective memory relies on a number of cognitive functions, it is amongst the most vulnerable and studies show that persons with acquired brain injury have reduced prospective memory as one of their most common complaints [6]. PDAs are obvious prostheses for reduced prospective memory; they will help the patient retain information and intentions over a period of time and an alarm will help the patient know when the time has come to carry out the intention. However, sometimes the very skill needed to operate the PDA is exactly the skill the patient is lacking. Wilson [9] has described this dilemma in the following way: "The people who need memory aids the most have the greatest difficulty in using them."

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

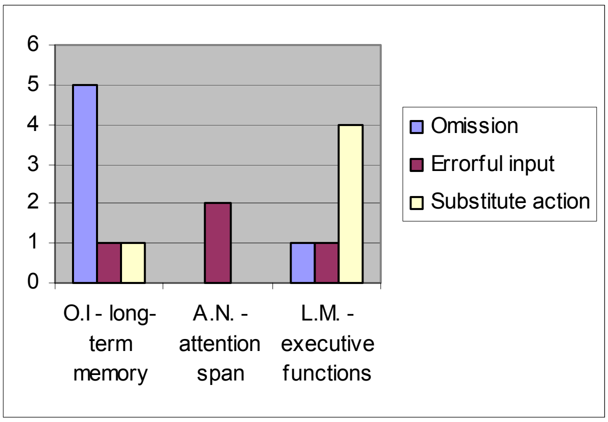
2. PROBLEMS OF INTERACTION
In the experiment patients were asked to complete a total number of seven assignments. Each assignment consisted of entering a piece of information into the PDA (e.g. an appointment, a task, setting an alarm etc.). The mistakes made when entering this information were classified into three groups (table 1):

Omission: Assignment not completed, nothing to be found in the PDA.

Errorful input: Assignment completed in the wrong way, errorful information to be found in the PDA.

Substitute action: Assignment solved in an alternate way than indicated in the instruction, the right information to be found in the PDA, but in the wrong place.

Table 1. Distribution of errors



2.1 Reduced Long-term Memory
The characteristic type of mistake for patients with reduced long-term memory is omission. When the person cannot retain information from one session of use to the next, the PDA will appear unfamiliar on each encounter and the person will remain a novice user. Functions in the PDA that cannot be solved simply by logic approach, but depends on the user's

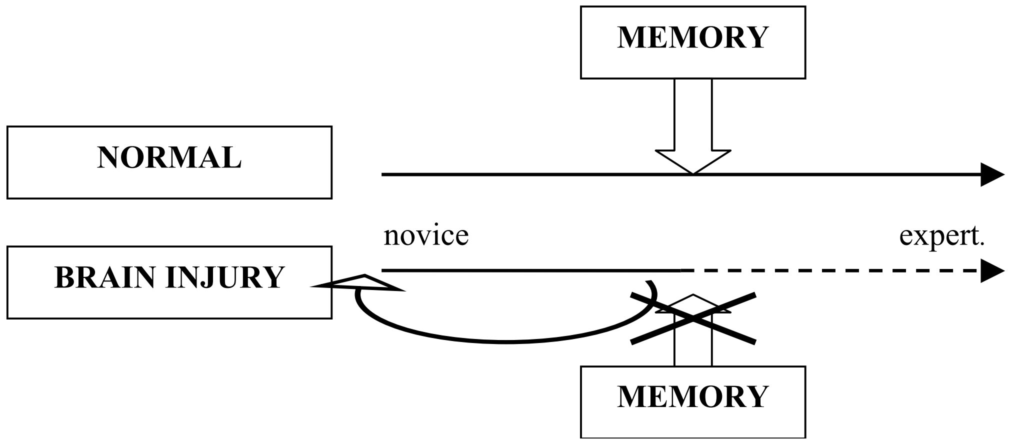


Figure 1. A normal person uses memory to advance from novice to expert. Since a person with memory problems cannot do so, he remains a novice user.

previous experience, become difficult to solve. The patient 'O.I.', who through psychological testing has been diagnosed with severely damaged long-term memory, is after six sessions of tutoring incapable of solving any of the seven assignments correctly. Omission accounts for five of the seven mistakes. Being a novice appears to be incompatible with operating the PDA, but at the same time it is impossible for the patient ever to gain expert status (figure 1).

2.2 Reduced Attention Span

The performance of patients with a reduced attention span (working memory) is characterized by 'errorful input'. With a reduced attention span it is difficult for a person to remember information throughout a sequence of actions and hence information can be lost or mistaken. This type of error dominates the performance of 'A.N'. Despite the fact that she is rather good at operating the different programs and functions included in the test, she does seem to loose herself in some sequences, resulting in entries of errorful information. 'A.N.' has been diagnosed with a reduced attention span and for that reason especially the complex assignments evoke 'errorful input'. E.g. when being assigned to use a repeat option on an appointment she is capable of performing all the right steps, but ends up getting the appointment started one month too early. This could be because the repeat option is quite complex: the action sequence consists of 11 steps involving four different windows.

2.3 Reduced executive functions

'Substitute action' is when the assignment has been solved, but unintended programs or functions have been used instead of the ones indicated in the instructions. 'L.M.' makes particularly many 'substitute actions'. For instance: He enters a reoccurring event at the appointed date for each week instead of using the 'repeat' function in the PDA. In another assignment 'L.M.' enters a new task when asked to attach a note to an existing appointment in the calendar. Furthermore he is by far the slowest at completing the assignments, using close to twice the average amount of time. The psychological testing shows that L.M. has severely reduced executive functions. Executive functions are what allow us to form a plan and 'execute' it as intended. It can also be described as a problem solving skill. When brain injury reduces the executive functions the patient will experience difficulty in forming a plan or strategy and especially shifting between two strategies can be a challenge. The extensive amount of substitute actions as well as time used by 'L.M.' might be due to executive problems. It causes him great effort to form a plan when it implies having to choose among a number of functions and programs.

3. POSSIBLE IMPROVEMENTS OF PDA

To include patients with acquired brain injury in the group of potential users of a PDA it is necessary to rethink interface and software. Measures need to be taken in order to support the three kinds of deficits discussed above: reduced long term memory, reduced attention span, and reduced executive functions. One way to support this group of users is to help them benefit from their remaining resources. People with reduced long-term memory would benefit from an interface which increases the possibility to rely on their problem solving skills (executive functions). In this way they could approach the PDA with logic instead of repeatedly failing to retrieve specific details about use from memory. Another way of making the PDA more accessible to the group is to follow general design principles as they are known from non-assistive technology (i.e. Shneidermans eight golden rules of interface design [8]). For instance shorter action sequences are advisable in order to relieve working memory: "Tasks should be arranged such that completion occurs with few actions, minimizing the chance of forgetting to perform a step." [8] Also information relevant to the user should be kept visible at all times during his steps of action and thereby constitute an external information store. This would help the user orient himself with respect to both entered information as well as available software functions. Inglis et al. [4] pinpoint that one should always pursue that "[...] the most important elements of the system are visible at all times […]" , even if this should reduce function. These suggestions could be accommodated by giving the interface a large surface structure rather than the current tree structure. In order to increase usability for a patient with reduced executive functions the question of how to reduce confusion needs to be addressed. Simplicity in the design is a possibility. This would mean removing redundant options and hence minimizing the time consuming process it is for such a user to operate the calendar. In the PDA used in this particular experiment redundancy has been found on many levels, e.g. overlapping programs, functions and buttons.

A starting point in the remaking of a PDA interface design and software development could be to discuss which metaphor to implement in the design. In order to help users with reduced executive functions we would suggest to maintain a very high degree of consistency and simplicity in the design; something which can be achieved by choosing one single design metaphor and sticking to it. The PDA-program most commonly used by the patients is the calendar program. Certain aspects of it draw heavily on the metaphor of a real paper calendar. However, some aspects remain that rely on a computer metaphor. There are certain advantages of choosing the calendar metaphor. For one it uses a surface structure as opposed to a computer metaphor that relies on a tree structure (cf. computer programs using menus). Therefore a calendar metaphor would be of great help to users with a limited attention span. Secondly a calendar metaphor would be familiar to most potential users, not excluding people without computer experience. Thereby it might possibly solve some learnability problems connected especially to users with reduced long term memory.

4. FUTURE RESEARCH

In this study we have tried to pinpoint some of the areas that might influence on the performance of a person with acquired brain injury when using a PDA as a 'cognitive prosthesis'. However, it is necessary to research each of these areas separately to find out if they really do affect the user in any way. Some of the possible future research projects could include

- Surface structure vs. tree structure
- Calendar metaphor vs. computer metaphor
- Few programs/functions vs. many programs/functions (redundancy vs. more options)

5. REFERENCES

[1] Ellis, J. (1996), *Prospective Memory or the Realization of Delayed Intentions: A conceptual Framework for Research*, Brandimonte, M., Einstein, G.O., McDaniel, M.A. (Eds.), Prospective Memory: Theory and Applications, Lawrence Erlbaum Associates Inc, New Jersey, s.1-22

[2] Eysenck, M.W., Keane, M.T. (2000), *Cognitive Psychology: A Student's Handbook*, 4th ed., Psychology Press Ltd., East Sussex

[3] Gartland, D. (2004), *Considerations in the selection and use of technology with people who have cognitive deficits following acquired brain injury*, Neuropsychological Rehabilitation, 14 (1/2), s. 61-75

[4] Inglis, E.A., Szymkowiak, A., Gregor, P., Newell, A.F., Hine, N., Shah, P., Wilson, B.A., Evans, J. (2003) *Issues surrounding the user-centred development of a new interactive memory aid*, Univ Access Inf Soc, 2, s.226-234

[5] Inglis, E.A., Szymkowiak, A., Gregor, P., Newell, A.F., Hine, N., Wilson, B.A., Evans, J., Shah, P. (2004), *Usable technology? Challenges in designing a memory aid with current electronic devices*, Neuropsychological Rehabilitation, 14 (1/2), s. 77-87

[6] Kliegel, M., Martin, M. (2003), *Prospective memory research: Why is it relevant*, International Journal of Psychology, 38(4), s.193-194

[7] LoPresti, E.F., Mihailidis, A., Kirsch, N. (2004), *Assistive technology for cognitive rehabilitation: State of the art*, Neuropsychological Rehabilitation, 14 (1/2), s. 5-39

[8] Schneiderman, B. (1998), *Designing The User interface*, Addison Wesley Longman, The University of Maryland

[9] Wilson, B.A. (2004), *Recent Developments in Neuropsychological Rehabilitation*, Workshop på Center for Hjerneskade, København d. 26/4-2004

A Framework for the Design of Tangible Interaction for Collaborative Use

Eva Hornecker
Institute of Design & Assessment of Technology
Vienna University of Technology, Austria
Mads Clausen Institute, University of Southern Denmark
eva.hornecker@media.tuwien.ac.at eva@mci.sdu.dk

ABSTRACT
This paper describes a framework (under progress) for designing tangible interaction for collaborative use. Four design aspects/themes should be carefully exploited, if one wants to design a tangible interaction system intended for collaborative use: space and spaciality, tangible manipulation, embodied facilitation and representations and their expressiveness.

Categories and Subject Descriptors
H.5.m [**Information Interfaces and Presentation**]

General Terms
Design, Human Factors,.

Keywords
Tangible User Interfaces, Tangible Interaction, Design, .

1. INTRODUCTION
Tangible User Interfaces (short: TUIs) have become a hot topic in HCI. Until quite recently research was mostly technology-driven, focusing on the development of new systems. A change in focus can be detected from the special issue of Personal & Ubiquitous Computing on "tangible interfaces in perspective". Yet there still is a lack of theory, why "tangible interaction" works and what exactly is important to it [5]. And although cooperation support might be the most important generic feature offered by TUIs, this issue has attracted even less attention. Often people seem to assume that cooperation-specific advantages of physical environments are simply inherited by tangible interfaces. But a union of advantages from physical and digital worlds does not come automatically. We should know which properties of physical environments to maintain or explicitly exploit. Otherwise we risk destroying the resources relied upon in collaboration and diminishing positive effects of co-presence of human actors.

In my PhD thesis [10] I assembled findings from CSCW, work studies, communication research and design disciplines regarding social effects of physical, manipulable 3D media, identifying several lines of reasoning arguing for positive effects of TUIs on collaboration. Part of this project have also been empirical studies of cooperative situations supported by tangible media and a redesign study of a TUI. Ongoing work consists of distilling a design framework from the thesis results while broadening its scope to tangible interaction. This framework is introduced in the remainder of this paper.

Tangible interaction encompasses a much broader scope of systems or interfaces, which are not restricted to controlling *digital data* via manipulation of tangible objects (one can control real devices as well) and to the placement and relocation of tokens, what has been criticized as an imitation of interaction methods from the screen and neglecting the richness of embodied action [2, 4]. Therefore it seemed productive to address this larger design space, which also yields a higher number of systems to consider, leaving the somewhat artificial confines of any definition of TUIs behind.

1.1 Designing for collaboration
Some argumentation may be necessary about why to consider collaborative use. Many researchers agree that TUIs are especially suited to support collocated collaboration and report productive, enjoyable group processes. The number of TUI-systems aimed at collaborative scenarios – often design or group learning situations – documents this belief. Research is also acknowledging that social interaction is an inherent and important part of everyday life and of getting work done. E.g. museum visitors often come in groups and group interaction (also with strangers) plays an important role in the visit experience [3,8]. In work situations implicit communication and coordination, both co-present and distributed in time and space, can be found even in at first sight seemingly individual work.

Designing FOR cooperation is analogue with the understanding within interaction design that one cannot design an experience, only for it – one can create opportunities for experience. Similarly we cannot force people to cooperate, but we can induce it and create a „force field" encouraging collaboration. The framework presented here aims to help in creating such "force fields" by offering "design sensitities" [3] and some (soft) guidelines.

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

1.2 Tangible Interaction

Albeit refraining from a strict definition, we do need some shared understanding of what is meant with tangible interaction. In literature one can find different characterizations of tangible interfaces and of tangible interaction in the broader sense:

- Physical representation & manipulation of digital data [14], respectively interactive couplings of representational physical artefacts with computationally mediated digital information [9]
- Input by physical manipulation with hands, this being sensed, followed by system feedback. The more embodiment (spatial ties of in/output) and metaphor used in shape or movement, the more tangible [7]
- Bodily interaction with physical (graspable) objects [4]
- A combination of real space and real objects with virtual displays ([1] on Interactivating Space)
- Interactive systems, physically embedded within real spaces, which offer opportunities for interacting with tangible devices, and so trigger display of digital content or reactive behaviors ([3] on interactive spaces)

These do match nicely with my framework, although the frameworks aspects were originally developed in a different way (combining results from literature analysis, theory review and own empirical work).

2. THE FRAMEWORK

2.1 The four aspects / themes

Tangible Interaction Systems for collaborative use should carefully exploit

- Tangible Manipulation
- Space and Spatiality
- Embodied Facilitation
- Representations and their Expressiveness

These aspects each again consist of three to four more specific aspects. These aspects lead to a varying number of concrete guidelines. Within the scope of this paper, only a selection can be presented. First I will explain the four themes. References to the characterizations of tangible interaction are written in italics.

2.1.1 Tangible Manipulation

Tangible interaction is bodily interaction with physical (tangible) objects/devices. Input is done by physical manipulation.

This is more than using only physical props within Virtual Reality worlds or using mouse & keyboard. It is also more than simulated haptic feedback. Tangible interaction is direct manipulation of physical objects. These physical objects ARE the interface (and not just an intermediary tool) and ARE interaction objects.

In addition tangible interaction is about the kind of action and interaction taking place – it is bodily interaction, using the hands in varied ways and often interacting with the entire body - and also on the reaction of the objects.

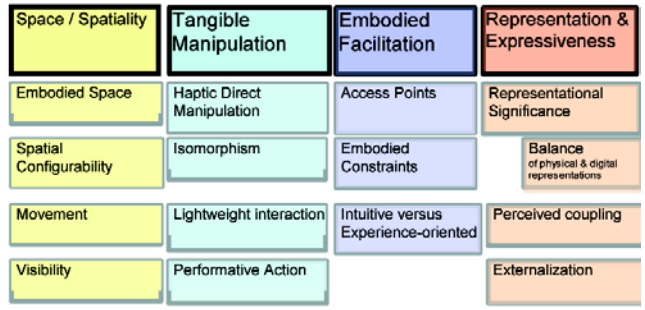


Figure 1. The complete framework with subthemes. (Representational significance and Balance are interconnected topics, handled as one subtheme)

2.1.2 Space and Spatiality

Tangible interaction systems are physically embedded in real space and combine real space with virtual displays.

Real space is 3D. It is inhabited, lived space (not abstract coordinates). Phenomenology talks of situated space, which receives orientation from an embodied here. “Situated space” further means that every room is marked by its atmosphere, the previous usage experience and its surrounding context. Situated space resembles the notion of places [3].

“People and physical space are made of the same stuff, but people and virtual space are not” [13, p.308]. Humans are spatial beings. Our perception is tightly coupled with action and movement. The body is a reference point for perception and thinking.

Space also offers a multitude of qualities or resources, like distance between objects, size, closure and openness, the ability to be filled with material entities, allowing movement and so on.

2.1.3 Embodied Facilitation

The space of the system is both a literal one (*tangible interaction systems being physically embedded in real space*) and metaphorical (the system space). Both are spaces for interacting in, allowing some movements and prohibiting or hindering others. We can interpret systems as spaces or structures to act and move in, thereby determining usage options and behavior patterns. They enforce social structures and direct user behavior.

Sometimes we stumble upon unintended side effects of design regarding social interaction, e.g. not sharing information, reduced awareness etc. This theme proposes to utilize these phenomena by intent. One can learn from didactics and facilitation methods how structure, both physical and procedural, can be shaped to support and direct group processes. This starts with arrangement of rooms and seating, provision of work materials, and goes on to deliberate adoption of game-like interaction rules. Both interaction design and didactics/facilitation can be interpreted as designing spaces for interaction and experience [12].

With Tangible Interaction Systems structure is not only in software, but also physical. They can truly embody facilitation methods. The way we can read and interact with representations is part of this structure.

2.1.4 Representations and their Expressiveness

Tangible Interaction is about physical representation of data.

Tangible objects stand in for - respectively represent - digital functions and data, or they represent other physical objects (and interact with them), or they simply stand in for themselves (with tangible appliances this can be the case). Often there are hybrid ensembles (or collections) of physical and digital (but perceptible) objects, each with different (representational) qualities.

Representations communicate to us; they have expression. In interaction we „read" and interpret representations. In interaction we act on, modify and create representations, permanent and fluent ones.

3. Selected Sub-Themes

Figure 1 shows the complete framework with the subthemes. Within this paper and due to the work in progress nature of the framework only a few themes and resulting guidelines can be presented in detail. I chose giving different details on these, with some only presenting the subtheme and giving the guidelines, and explaining a selected few with short examples.

3.1 Tangible Manipulation: Haptic Direct Manipulation

Guidelines around haptic direct manipulation suggest to

- Allow users to grab, feel & move the "most important stuff"
- Make tangible interaction the dominant mode of interaction
- Be pragmatic, usability goes first (no dogmatism about haptic directness)

3.2 Tangible Manipulation: Lightweight interaction

Lightweight interaction means to allow "conversation with the material". Users should be able to express and to test e.g. design ideas quickly, without cognitive overhead. This encourages participation and gives everyone similar chances. Guidelines supporting this are:

- Give constant, legible feedback
- Allow small iterative steps

3.3 Space and Spatiality: Embodied Space

We encounter objects and people in space. They have material presence (and demand our attention) - we meet them face to face, feel their aura and resonate with them [15]. Concrete space is always situated: we experience and create places [3]. This implies multisensory experiences, also in the embodied sense that space always surrounds us. Social effects of sharing space are intimacy, social nearness and a higher tendency to cooperate. Being in the same place is a reciprocal situation where seeing implies being seen. This creates both vulnerability and trust [15].

Guidelines around the theme of embodied space suggest to

- Enable co-presence of people & objects
- Exploit the relation of the human body to space (embeddedness, distance, left-right back-forth, big-small, enclosed-open)
- Turn space into places

3.4 Space and Spatiality: Movement

There is currently only one specific guideline

- Support bodily interaction

Bodily interaction is experienced as *enlivening*. It heightens the inner activity level, stimulates mental energy, creates mental-bodily engagement. Bodily interaction is *expressive*. It is part of expert skill, and also is a means of personal self-expression. Bodily interaction is highly *performative* and often a part of implicit coordination with other people. Bodily interaction is *observable*, fostering group awareness and attracting attention.

A bigger interaction space enforces more & larger movement and thus intensifies interaction. It also encourages more expressive gesture. Movement leads to bodily appropriation of space (taking ownership). The dynamics of group discussion often can be read from bodily interaction rhythm (see e.g. [6]).

Figure 2. Visitors at a city festival dancing on the clavier within the nightly "Sensoric Garden" installation

Figure 2 shows the clavier path, one installation from the "Sensoric Garden". This was the result of a one-year student project, being an ensemble of interactive installations installed in a public park during a festival in Bremen and open during night hours [12]. Walking along this path (thereby interrupting light sensors) triggered colored spotlights where one walked and different drums and beats.

Here the expressive & performative aspects of interaction are salient. This is part of the fun of interaction. People try out different things, walking back and forth, using their umbrellas to trigger light sensors, jumping and dancing. We could observe people dancing for extended periods of time and dancing in groups. The picture demonstrates that it also seems to be fun to interact as a group. As the light sensors are located directly beneath corresponding spotlights and the designers had taped stripes on the floor to highlight the clavier analogy, it was intuitive to understand general concept. Yet for a good performance practice or skill is needed.

In the assessment of two different versions of the EDC, a system for participatory urban planning relying on an augmented game board, I also found movement to be a crucial element [6]. The different size of the system versions had marked effects on group behavior. With the larger system we observed large gestures, people taking ownership both bodily and mentally and a very lively group. The group using the smaller system used only short, tiny gestures and behaved almost timid and quiet, not appropriating the system space. This is visible in posture. Whereas the first group often leaned out wide over the system, the second tended to lean back and to use self-blocking postures (elbow on table and chin leaned upon while drawing).

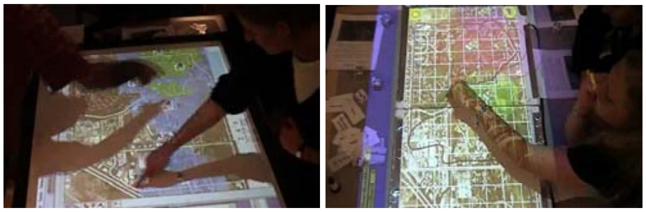


Figure 3. Comparing movement and posture of user groups at two system versions of different size [6]

A big interaction space necessitates big movements. In this study we could observe how bodily interaction seemed to trigger engagement and mental movement. The visible movement (on the video of the sessions) gives a reliable indicator for the discussion intensity and imagination taking place, while quietness is linked to phases of reflection & abstraction. Big movements also enhance expressiveness (gestures). In the EDC study we found gestures to be important for building a shared mental image. These big movements and gestures also were important in fostering taking ownership of system space.

3.5 Space and Spatiality: Visibility

Central guideline here is:

- Ensure visibility of 1) objects 2) actions and 3) effects

Visibility makes actions observable and improves legibility. It contributes to account-ability [13]. Because of the implicit force to be able to explain publicly visible actions, people tend to reflect more what they do. Seeing actions while they are being done and seeing prepraratory movement aids anticipation and improves (peripheral) awareness, supporting coordination. Seeing an action and observing the effect also enables learning by observation. Reciprocity (seeing and being seen) is important for social interaction. Visibility of objects provides focus and shared reference points. It calls for attention.

In the EDC evaluation it was salient how people would always talk and explain when they were sketching a solution idea. The drawings gave visible traces of actions, while gestures served as non-permanent, slowly fading memory help. Without visibility of emerging ideas (one system version had no sketching facility), the shared mental image suffered, resulting in insecurity and less thorough discussion of alternative solutions. Yet both system versions effectively served as a focus for discussion, creating a shared (bodily and mental) orientation.

3.6 Embodied Facilitation: Access Points

Access points refers to the options people have to access and actively manipulate the system. Access is an issue of power, highly influencing group dynamics. Guidelines are:

- Give multiple points of interaction
- Give equal access and no privileges
- Implicitly produce a shared transaction space, that is a space where the participants action and attention areas overlap (usually the forefront of bodies)
- Allow for simultaneous action

In assessing the EDC [6] we found that privileged access of facilitators to system functionality in the facilitated participatory session affected the power play. Changing the system to a setup, where crucial system function could be accessed from the table (where the group was sitting) established equality and enabled everybody to take over system control.

4. ACKNOWLEDGMENTS

All people supporting my thesis project and beyond, especially Jacob Buur for hosting me at the MCI and supporting this work.

5. REFERENCES

[1] Bongers, B. (2002) Interactivating Spaces. Proceedings of the Symposium on Systems Research in the Arts. Informatics and Cybernetics Baden-Baden, Germany.

[2] Buur, J., Jensen, M.V., Djajadiningrat, T. (2004). Hands-only scenarios and video action walls: novel methods for tangible user interaction design. Proc. of DIS 2004. pp. 185-192

[3] Ciolfi, L. (2004), "Situating 'Place' in Interaction Design: Enhancing the User Experience in Interactive Environments". Ph.D. Thesis, University of Limerick

[4] Djajadiningrat, T., Wensveen, S., Frens, J., Overbeek, K. (2004). Tangible products: redressing the balance between appearance and action. Personal and Ubiquitous Computing 8(5). pp. 294-309

[5] Dourish P. (2001). Where the Action Is. The Foundations of Embodied Interaction. Bradford Book, MIT Press, Cambridge MA.

[6] Eden H., Hornecker E., Scharff E. (2002). Multilevel Design and Role Play: Experiences in Assessing Support for Neighborhood Participation in Design. Proc. of DIS'2002, pp. 387-392.

[7] Fishkin, K.P. (2004). A taxonomy for and analysis of tangible interfaces . Personal and Ubiquitous Computing 8(5). pp. 347-358

[8] Heath, C.; Luff, P.; Vom Lehn, D.; Hindmarsh, J. (2002). Crafting participation: designing ecologies, configuring experience. Visual Communication Journal Vol 1(1), pp.9-33

[9] Holmquist, L., Schmidt, A., Ullmer, B. (2004). Tangible interfaces in perspective: Guest editors' introduction. Personal and Ubiquitous Computing 8(5). pp. 291-293

[10] Hornecker, E. (2004). Tangible User Interfaces als kooperationsunterstützendes Medium. PhD-thesis. University of Bremen. July 2004

[11] Hornecker, E. (2004?). Analogies from Didactics and Moderation/Facilitation Methods: Designing Spaces for Interaction and Experience. (in press) Digital Creativity.

[12] Hornecker, E., Bruns, F.W. (2004). Interaktion im Sensoric Garden. Proceedings of Mensch & Computer 2004, pp.65-74

[13] Robertson T. (2002). The Public Availability of Actions and Artefacts. CSCW Journal 11 (3-4), 299-316.

[14] Ullmer B., Ishii H. (2000). Emerging frameworks for tangible user interfaces. IBM Systems Journal 39 (3 & 4), 915-931.

[15] Waldenfels B. (2000). Das leibliche Selbst. Vorlesungen zur Phänomenologie des Leibes. Suhrkamp, Frankfurt a/M.

How Can HCI/Usability and Computer Games Enrich Each Other: A First Look

Anker Helms Jørgensen
Center for Computer Games Research
Department of Digital Aesthetics and Communication
IT University of Copenhagen
anker@itu.dk

ABSTRACT

The fields HCI/usability and computer games have existed for a few decades with virtually no mutual interaction. However, in recent years, a number of exchanges have appeared, both in academia and in practice. This paper presents a preliminary account of this development. Exchanges in both directions seem viable: evaluation methods from HCI/usability towards games and interaction techniques and supporting user communication from games towards HCI/usability. The paper concludes with a discussion of the differences and similarities between the two fields.

Author Keywords

HCI, usability, computer games, video games, interaction.

ACM Classification Keywords

H5.2. User Interfaces.

INTRODUCTION

The two fields HCI and computer/video games have been around for a few decades. Both have gained substantial foothold and have definitely come to stay. In computer science curriculae in the US, HCI has become one of the ten most popular areas [9]; every system developer knows about usability but does not necessarily agree or abide with it; a vast number of usability activities are conducted every day; numerous textbooks in HCI and usability have appeared; and a good number of conferences are held every year.

Similarly, the area of computer games has exploded in the last decade: a large proportion of the population plays regularly; alledgedly, the game industry in the US is the fastest growing sector in IT and is now economically equivalent to Hollywood; and massive multiplayer online games with hundred of thousands or even millions of players have appeared - such as Everquest and Lineage.

A generic feature of the two fields is the dedication to providing the users with what they want, but nevertheless there has been very little interaction between them. An exception in the HCI field is the 1982 paper by Tom Malone [10] addressing how software can be made more enjoyable by adopting ideas from games. Another exception is the extremely popular game *The Sims* developed by Will Wright. Richard Rouse [13] comments on this game

> *"The Sims' user interface is a beautiful example of how to do an interface correctly - the user interface is so simple and intuitive and the tutorial addresses how to play the game, not how to manipulate the interface … no doubt the result of rigorous playtesting"*

An interview with Will Wright reveals in fact that the user interface went through 11 iterations with about 100 playtesters where the developers sat down and watched players' mistakes and misconceptions [13].

But within the last few years, a remarkable change has emerged. In the game industry, Microsoft established a Playtest group a few years ago with quite remarkable results [11]. In academia, an article in the respected journal Human-Computer Interaction based on a Ph.D. thesis by Carlo Fabricatore appeared [7]. In spite of these developments, Pagulayan et al. conclude [11] *"This relationship between theories of game design and traditional HCI evaluation methods has yet to be defined but definitely yields an exiting future."*

This paper shares this point of view and takes up this challenge. First the paper outlines the background and a framework. Next contributions from usability to game and vice versa are outlined, followed by a discussion of the differences and similarities of the two fields. On a related note, usability has recently been applied successfully to interactive art by Höök et al. [8]. This is relevant as many computer games do have an element of artistic expression.

BACKGROUND AND FRAMEWORK

In software engineering and the emerging field of HCI/usability[1], a host of books on user interface design emerged in the 1980's and 1990's. Similarly, in game development, many books have seen the light of the day since Chris Crawford's seminal book *The Art of Computer Game Design* from 1982 [4, 5,13,14]. It is striking that - to the best of my knowledge - no books on interface design in games have appeared although the players' interactions in games is so intense.

An interview with the highly experienced game designer and author Ernest Adams [1] provides a strong indication of the realm of game design:

> *"I've been working for a major game developing company for 8 years and I've never seen a methodologically sound study of who the players are ….. game design is based on common wisdom and guesses - designers build games for themselves."*

Against this background the paper illustrates the state of art by discussing five recent examples of exchanges between HCI/usability and computer games, see table 1 below. The selection of material included warrants a comment. I have focussed on the contributions in the literature that best support the points made in the paper. I do acknowledge that other work in the field exists, for example by Melissa Federoff.

Paper	Contribution	Route Usability Games
Pagulayan et al (2003) [11]	Evaluation methods	⇒
Christensen et al (2003) [2]	Participatory Design methods	⇒
Bekker et al (2004) 16]	Validity of usability evaluation in games	⇒
Rouse (2001) [13]	User needs/conceptions	⇐
Dyck et al (2003) [6]	Interaction techniques & player communication	⇐
Chao (2001) [15]	Game interface to prod. software	⇐
Fabricatore et al. (2002) [7]	Empirically founded design guidance	⇔

Table 1 Selected examples of HCI/usability-games interaction

[1] I have chosen the akward terminology "HCI/usability" - as opposed to either HCI or usability - as both areas are relevant in the discussion.

ROUTES FROM USABILITY TO GAMES

Pagulayan et al**.** [11] report an impressive endeavor undertaken at Microsoft Game Studio in Washington. The rationale for their work is that games *are* doing very well on the market, but would in all probability do even better if the games were more usable. There is also a growing recognition that even successful games such as Halo suffer from poor usability, for example in terms of tutorials and weapons control [3]. Established in 2000 and employing a handful of game developers, psychologists and HCI specialists, the Playtest group has tested more than 70 games with more than 10.000 participants. The group employs quantitative and qualitative methods and address initial player experience as well as deep gameplay. The group is also involved in development of the games. They employ an iterative approach denoted RITE: Rapid Iterative Testing and Evaluation based on short cycles. The following example illustrates their work [12]. In a combat flight simulator most players had difficulty with the term "AI level" in the difficulty presets (meaning "Artificial Intelligence level"). Several interpretations were offered:

- "A one level … ?"
- "Al level … ?"
- "Altitude level … ?"
- "Anti aircraft … ?"

The redesign solution was to rename "AI level" to "Enemy level" in three settings: rookie, veteran, ace; this remedied the problem completely. Did anyone say *"Classic usablity wisdom: 'Speak the users' language'"?* To the best of my knowledge this endeavour at Microsoft is the only example of major game developers having seriously taken up usability approaches.

Christensen et al. [2] report on using techniques and methods from Participatory Design and usability in developing a hybrid game called Takkar. Takkar combines live action role playing games with virtual role playing games, allowing playing on-line inbetween the much less frequent live action role play sessions that take place outside with dresses, equipment, scripts, roles, etc. a few times a year. The development included substantial prototype development, numerous playing sessions both on-line and in the real world, expert reviews, and in-depth debriefing interviews. The iterative development revealed a range of issues such as walking and talking simultaneously with your fellow players. This is *completely* natural in the real world while not at all easy to transfer to the virtual game world due to game engine constraints.

Bekker et al [16] report on a quantitative study of the number problems identified in using usability evaluation techniques in childrens' initial and extended use of computer games. They found that seven users were required to identify 85% of the known problems in initial use and eleven users in extended use – contrasting the commonly held – but also contested figure of five users as advocated by Jakob Nielsen.

ROUTES FROM GAMES TO USABILITY

The opening chapter of Richard Rouse's book *Game Design – Theory and Practice* [13] is delightful reading. It is called *What users want.* It lists 16 principles such as *Players want a challenge* and *Players expect to fail* – much like principles or guidelines in HCI/usability literature. A considerable overlap with conventional HCI/usability approaches appears, but the differences are far more striking: The description – almost a narrative - entails the player in flesh and blood. It is clearly written by a highly experienced game designer who knows both the design and the play side of gaming. Contrasting this, most HCI/-usability textbooks perceive users in the light of cognitive science: mental models, memory capacity, attention span, features of the human visual system, etc. The Rouse opening chapter is the chapter I always wanted on users in HCI/usability books. On this note, I believe that many HCI/usability books are written by researchers with little or no design experience.

Dyck and colleagues [6] undertook a comprehensive study of interface features of contemporary computer games. The background to their study is the observation that games developed in their own direction, e.g., no windows and no widespread standards, and there is a strong focus on novelty, user performance and user satisfaction. Thus, games can be seen as adopters of new interaction technologies. Dyck and colleagues studied 13 games while employing 5 methods: game playing, keeping diaries, observations, reviews, and analysis sessions.

Their findings address four areas:

- *Effortless community*. Games makes it easy to form, join, and participate in communities of users. An example is the successful integration of the natural community in the game world in massive multiplayer games.
- *Learning by watching*: gamers help people learn the application by watching 'over the shoulder' of more experienced users. As an example, in multiplayer games all players can readily observe other players' actions and thereby reduce obstacles in learning.
- *Deep customizability*: games give users the power to modify and extend any aspect of the user interface and allow them to share those modifications. As an example, in Everquest, players can readily store an action sequence in a new button with a few mouse clicks as opposed to for example Microsoft Word, where 7 actions are required to record a macro and 5 more to place it in a toolbar.
- *Fluid system-human interaction*: games communicate information to users in ways that do not demand the users' attention and do not interrupt the flow of work. As an example, in many games, system messages are delivered in an unobtrusive way and do not require the player to acknowledge or dismiss them.

Dyck et al. conclude that games provide a wealth of successful radical and novel interaction concepts that might benefit users of productivity software. On a final note, I should add that some find that these conclusions paint a too glamourous picture [J.H. Smith and L.J. Christensen, personal communication, June 20th, 2004].

Chao [15] reports an interesting application of game elements to productivity softweare. He reshaped the traditional user interface to process management in UNIX to a Doom-look-a-like, where a monster corresponds to a UNIX application process. Killing a monster corresponds to killing a process, while reducing the power of a monster corresponds to reducing the resources of a process, such as slowing it down. The interface was disseminated via the Internet and soon found wide acceptance among UNIX system administrators.

ROUTES BOTH WAYS

Carlo Fabricatore and colleagues [7] asked the question: *What do players want?* and answered it by empirically addressing players' playability preferences. They selected the genre action game and selected 39 popular games. They had 53 experienced players play each of these games for roughly two hours while tape-recording their comments, logging their interactions, and conducting interviews after the sessions. The results were analysed in depth and resulted in a *game reference model* with 3 categories (entities, scenarios, and goals) broken down further. The guidance includes mandatory prescriptions and voluntary recommendations. An example from the category *role* (part of *identity* and *entities*) is the prescription *Allow the player to understand the role easily* and the recommendation *Offer the player the possibility of selecting the initial role of the protagonist.*

This work is groundbreaking in that it provides an empirical basis for design guidance by way of a comprehensive study of actual player behaviour. Hence Fabricatore and colleagues supplemented the experience-based game design evidence with systematically, empirically derived design evidence. As the evidence has been established in a fashion acceptable by the HCI community, one might hope that the evidence may flow back to HCI/usability and support designers looking into ways of making productivity software more enjoyable.

DISCUSSION AND PERSPECTIVES

As indicated above, exchanges in both directions seem viable: for example evaluation methods from usability towards games and interaction techniques and supporting user communication from games towards HCI/usability.

There are a range of commonalities between productivity software and computer games (and more broadly entertainment software): learning, motivation, mental models, control, interaction, feedback, spatial navigation, linguistic and visual expressions etc. These seem to be within the scope of traditional usability approaches in both camps.

There are, however, also substantial differences that call for thorough consideration when trying to marry usability and games. The most important is probably that of *challenge* in games by way of an intended difficulty. This is witnesses by the game design slogan *easy to learn but difficult to master*. This is basically handled by balancing the game-play – a feature that can be compared to usability efforts towards meeting the users' needs in terms of effectiveness, efficiency, and satisfaction.

Another striking difference is freedom of use and acquisition. Productivity software is largely used mandatorily, typically in work settings. In acquiring and using productivity software, the user or consumer has little or no choice due to de-facto market standards (in web-applications, the range of choices is, however, often somewhat wider). Contrasting this, use of computer games is featured by being almost exclusively voluntary and hence Darwinism in the marketplace has a more pronounced role at the low level. However, as in the early days of computers and computing, the quality usability of games seems so far to have had little impact in the market.

In many computer games, huge virtual universes are created and players can contribute substantially to these universes. In spite of the richness, diversity and vastness of these universes, the action takes place almost entirely *within* these universes, encompassing both the digital world and the real world (with player communities, bying/selling avatars and objects, exchange of level designs, etc.). In productivity software there are substantial implications *beyond* the software: *"Can I finalise the budget for the next quarter before the management meeting tomorrow morning?"* and *"Did I manage to get the right plane ticket on this odd airline web-site?"*

The implications of these differences – and other fundamental differences between productivity software and computer games – are currently not clear in terms of the potential of usability methods in development of games and the other way: what productivity software can learn from computer games re. engagement and enjoyment. Perhaps a starting point can be the maxim in computer games (and indeed games in general): *easy to learn, but difficult to master* that reflects and contrasts conventional usability evidence *easy to learn and easy to master*. This is where the *substance* is regarding purpose of the software: the fundamental *challenge* for the player and the fundamental *utility* for the user.

ACKNOWLEDGMENTS

I am grateful for insightful comments to an earlier version of this paper from J.H. Smith, G. Frasca, L.J. Christensen, and two anonymous reviewers.

REFERENCES

1. Adams, Ernest, personal communication, Nov. 4th, 2003.
2. Christensen, L.J., Jørgensen, T. T-Y., and Jørgensen, A.H.: Developing a hybrid of MMORPG and LARP using usability methods: the case of Takkar. *CD-ROM Proc. Conf. Level Up*, Utrecht, Holland, 2003.
3. http://crankyuser.com/halo.multiplayer (June 18th, 2004).
4. Crawford, C.: *The Art of Computer Game Design: Reflections of a Master Game designer*. McGraw-Hill, 1982.
5. Crawford, C.: *Chris Crawford on Game Design*. Prentical Hall, 2002.
6. Dyck, J., Pinelle, D., Brown, B., and Gutwin, C.: Learning from Games: HCI Design Innovations in Entertainment Software. *Proc. 2003 Conf. On Graphics Interface (GI'03)*, Halifax, 2003
7. Fabricatore, C., Nussbaum, M. and Rosas, R.: Playability in video games: a qualitative design model. *Human-Computer Interaction*, vol 17 no 4, 311-368, 2002.
8. Höök, K., Sengers, P., and Andersson, G.: Sense and sensibility: evaluation and interactive art. *Proc. CHI 2003*, p. 241-248, 2003.
9. King, John, personal communication, March 7th, 2003.
10. Malone, T.W.: Heuristics for designing enjoyable user interfaces: lessons from computer games. *Proc. Human Factors in Computer Systems*, Gaithersburg, Maryland, p. 63-68, 1982.
11. Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R., and Fuller, T.: User-centered design in games. In J. Jacko and A. Sears (Eds.), *Handbook for Human-Computer Interaction in Interactive Systems*. Erlbaum, p. 883-906, 2003.
12. Pagulayan, R.: Testing for Enjoyment. NNGroup workshop, London, Oct 6th, 2003.
13. Rouse, R. III: Game Design – Theory and Practice. Wordware, 2001.
14. Rollings, A. and Adams, E.: Andrew Rollings and Ernest Adams on Game Design. New Riders, 2003.
15. Chao, D.L. Doom as an interface for process management. (2001). In Proc. CHI'2001,. New York: ACM Press, p. 152-157
16. Bekker, M, Barendregt, W, Crombeen, S & Biesheuvel, M (2004): Evaluating usability and challenge during initial and extended use of children's computer games. Proc HCI 2004 UK, Leeds, 331-345, Springer.

Design Through Engagement

Lene Nielsen
Dept. of Informatics, CBS
Njalsgade 80, trappe 18
2300 Copenhagen S
+ 45 38253358
Ln.inf@cbs.dk

ABSTRACT

Within user-centered design personas and scenarios are seen to help designers imagine the users and aid exploration of design ideas. From the descriptions of fictitious users, the designer can identify with the users [6] and predict what the product will do for the user. The experiences from a workshop setting points to the limitations and problems with written persona descriptions in facilitating engagement. Furthermore the results suggest that the descriptions should be accompanied by plenary discussions in the design team.

General Terms
Design

Keywords
Personas, scenarios, design, user-centered design

1. INTRODUCTION

The method of personas has become widely appreciated [5] as a method that helps designers understand the users and keep focus on the user in the design process.

Most recommendations of how to create personas seem to include a written description either in the form of a bulleted list [8] or a longer narrative. Often photos depicting the fictitious user accompany the descriptions.

"Because personas must provoke empathy, they must have attributes of a real person" [3] p. 12. As the quote indicates real people can create empathy, but what attributes needed is not explained. Unfortunately the HCI literature is sparse on how a written description can create empathy and the literature does not examine the process of understanding another person from a written description.

In earlier work [7] I have looked at film scriptwriting to find theories and recommendations dealing with how written descriptions of users can be engaging. With inspiration from this area and from narrative theory I have developed the concept of engaging personas and narrative scenarios that forms the spine for both the workshop presented and this paper.

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

1.1 Workshop on personas, situations, and scenarios

In December 2001, the e-business group at AstraZeneca were in the middle of launching a web service for asthma patients – LinkMedica[1] - and wanted to gain information about their users; the users' needs and their use of the site. It was decided that personas and scenarios should be used as methods to provide the e-business group with the insight they lacked.

LinkMedica is a web service for asthma patients and health care providers. Asthma patients are able to monitor their asthma using an asthma diary, and health care providers can access the patients' diary data. [2]

To gain insight into the users, both patients and health care providers, four user studies were carried out. The studies focussed on the use of the site, the use of Internet, the patients' relationship to their illness, and the health care providers' perception of which patients would benefit from using the site.

Two workshops for the e-business group were initiated. The workshops focussed on creating personas and scenarios from field data. Both workshops were video filmed. The videos and the written material from the workshops have provided the analytic material for this paper.

The workshop included three steps: development of personas, definition of each persona's needs and the situations in which the persona would use the website, and finally development of scenarios for each need/situation.

2. DEVELOPING ENGAGING PERSONAS AND NARRATIVE SCENARIOS

The framework that was followed in the workshops consists of three steps:

1. Defining engaging personas and writing of these
2. Defining needs and situations for each persona
3. Writing scenarios for each situation or need

2.1 Engaging personas

The persona as perceived by Alan Cooper [3] is defined by goals. The engaging persona is defined by character traits[2]. This defines the persona as a rounded character [4 that enables empathy and sympathy thus creating engagement [9].

---

[1] See www.linkmedica.dk

[2] See [7]

The character possesses a variety of elements, both personal (inner) and inter-personal (social, public, and professional). The uniqueness of the character traits creates the difference between each persona. The personas have inner needs and goals as well as interpersonal wishes and professional ambitions.

The engaging persona description is characterised by:

1. Body: A description that allows the reader to understand the character as a human in a bodily sense.
2. Psyche: A description of lasting psychological traits.
3. Background: A description that includes cultural and social background.
4. Emotions: Current emotions.
5. Cacophony: Two oppositional character traits

The term cacophony can be expressed as two oppositional character traits and a peculiarity[3]. This creates a dynamic character with potential for development. With the oppositional character traits the development of the character is so to speak embedded in the character from the beginning.[4]

2.2 Needs and situations

The understanding of the personas is further developed in the definitions of needs and situations. The situations are decided upon with a point of departure in the persona's needs or a situation that will develop a need. The situation unfolds within the given design area and serves as a starting point for a scenario.

2.3 The scenario

The scenario is written as a story with a beginning, middle, and end. The plot of the story is formed by the persona's needs, the actions in causality derived from obstacles that need to be overcome, and goals that need to be reached.

In the scenario, the design area is explored from the persona's point of view. Design ideas are created and evaluated against the needs and character traits of the persona.

The information is validated by how well the events are explained and fully covered as an effect of a previous event or by the character's motivations for action (closure) and how well the story hangs logically together (coherence) [1].

3. THE WORKSHOP

3.1 Creating personas

An analysis of the material produced in the workshop showed that the engagement in the persona did not come from the written persona descriptions, but from the discussions of personas.

There were two kinds of discussions during the process:

- Discussions between designers while defining the persona
- Discussions after the personas had been written when they were used in the next steps: discussions of situations and discussions of the individually written scenarios

The videotaped discussions between the designers while defining personas showed that the transition from discussion to writing was difficult and that information was lost in the process. During the sessions of defining and writing the personas, the participants showed two different strategies:

1. Discussing, taking notes, and writing when the discussions had finished.
2. Discussing, writing, discussing, writing.

The latter strategy seems to have the least loss of information.

3.2 Defining needs and situations

During the discussions of the situations the participants made a lot of effort to find the persona's motivation for using the application. While discussing, the participants formed an understanding of the engaging persona enabling them to understand the motivation for the use. In this way, the discussions became just as important as the written material and compensated for the loss of information in the writing process.

The participants used their own experiences to support the creative process, but as these were brought to the design process through discussions, there was a constant adjustment of information to the design area, to the other participants' experiences, and to the user inquiries.

The understanding of the personas furthered engagement. As the excerpt below shows, the designers were now able to construct each persona as a human agent and through the discussions place themselves in relation to the persona's action and emotions.

"E: Then she is in control again
A: Yes she works for the tax authorities so she is 'hard-nosed'. (…)
E: I don't think that if you ask her, she will say she is 'hard-nosed'. Because she has talked to someone at her work even though she prefers systems.
D: It has something to do with the knowledge she has herself, she wants it confirmed or denied. It is this way of thinking 'hard-nosed'."

As the personas were discussed and perceived in a collective process, the participants acquired a common ownership of the personas. This is in contrast to the individual ownership that was observed in the writing process.

The thorough understanding of the engaging personas became the source from which the needs, in accordance to the specific design area, could be seen to originate.

The engagement in the personas was tested in the discussions about which situations should be chosen to start the scenarios. During the discussions, the participants were unable to understand the situations without a frame of reference to both the persona's needs and his/her motivation for action.

---

[3] Film scriptwriters express this as 1+1+1. The 1+1 are the two character traits in opposition. The last +1 is a peculiarity (workshop with M. Rukov 2003)

[4] (See http://ccm.redhat.com/user-centered/personas.html for descriptions of users with only one character trait).

3.3 Creating scenarios

In the process of writing the scenarios, the persuasiveness of the story structure became evident. In one of the scenarios, the participants had relied heavily on their own knowledge of people similar to the persona in question and created a character that was unable engage the readers. The scenario made a perfect story, but during discussions the participants realised that the events in the scenarios did not resemble the circumstances in the field studies.

This shows that knowledge of the design area is important and that the scenarios have to be discussed. The written scenarios created design solutions from the persona perspective and allowed each participant an engagement in the persona. In contrast, the discussions opened up for a common understanding of the functionalities and the problem spaces connected to these.

4. WHAT DID THE ENGAGEMENT BRING TO THE UNDERSTANDING OF THE PERSONA

For the design group, the discussions functioned not only as definitions of the personas, the situations in which the personas would use the system, and the scenarios; they also created a broader understanding of the personas and the use situations.

The power of engagement was tested when the discussions took a new turn and moved away from a personas perspective to a perspective of an unknown and general user.

"D: You could make a function in the medical practitioner's programme where you can write a message and then you can tick, if you want to send it to patients as an e-mail or SMS."

The excerpt shows that the name of the persona has disappeared and the discussion focuses on an unknown medical practitioner. The design ideas were no longer validated by the persona. Instead the functionalities drove the story of the scenario forward.

5. MODEL FOR ENGAGING PERSONAS AND NARRATIVE SCENARIOS

The framework I present in the following is based on the concept of engagement in the persona, prescriptions derived from film scriptwriting and narrative theory, and experiences from the workshop. The framework includes the discussions as it, in the workshops, became evident that the discussions test the designers' presumptions and bring these into the open [7]. During the discussions the understanding that the designers have of the step just finished is aligned.

The process of creating engaging personas and narrative scenarios includes three steps and five phases. The three steps are:

Step 1: Persona development.

Step 2: Definitions of a number of situations and/or needs. Each definition forms the basis for a scenario.

Step 3: Development of a scenario connected to a situation and/or need.

Between each step are discussions that evaluate the previous step and create a foundation for the next step.

The information used in the discussions consists of information from three different sources:

- Information from field studies.
- Information derived from the participants' knowledge of the design area and persons similar to the personas.
- Fictional information created to suit the traits of the persona and the story structure.

6. CONCLUSION

To develop a system with a focus on the user requires that the designers are able identify with the user and understand the user's motivation for using the system.

The concreteness of the engaging persona seems to make it difficult to maintain a system focus. Instead it forces the designer to understand the system from the user's point of view.

Not all interaction designers and usability consultants are good writers, and the process of writing may create a barrier where information is lost. The plenary discussions among all partners involved support the written documents. Through the discussions, the engagement is developed, refined, and tested.

The value of discussions cannot be underestimated. They provide a means to bring the designers' individual assumptions into the open and test these against the character traits of the persona and against the field studies.

The discussions also create an awareness of the relationship between facts, cultural knowledge, and fictional information.

7. ACKNOWLEDGMENTS

My thanks to the HCI Research Group, Dept. of Informatics, CBS, for support.

8. REFERENCES

[1] Abbott, H.P., *The Cambridge Introduction to Narrative*. The Cambridge Introduction to. 2002, Cambridge: The Cambridge University Press.

[2] Anhøj, J. and L. Nielsen*, Quantitative and Qualitative Usage Data of an Internet-based Asthma Monitoring Tool*. Journal of Medical Internet Research, 2004. 6(3).

[3] Cooper, A., *The Inmates Are Running the Asylum*. 1999, Indianapolis: SAMS.

[4] Horton, A., *Writing the Character-Centered Screenplay*. 1999, L.A.: University of California Press.

[5] Manning, H., *The Power of Design Personas*. 2003, Forrester: Cambridge, MA.

[6] Nielsen, L. From User to Character - an investigation into user-descriptions in scenarios. *Proceedings of DIS2002* - Design of Interactive Systems. 2002. London.

[7] Nielsen, L., Engaging *Personas and Narrative Scenarios*, ed. CBS. Vol. 17, PhD Series. 2004, Copenhagen: Samfundslitteratur.

[8] Pruitt, J. and J. Grudin. *Personas: Theory and Practice*. http://www.research.microsoft.com/research/coet/Grudin/Personas/Pruitt-Grudin.doc

[9] Smith, M., *Engaging Characters: fiction, emotion, and the cinema*. 1995: Clarendon Press.

Studying User Experience in Public Places

Jeni Paay [1,2]

[1] Department of Information Systems and
Faculty of Architecture, Building and Planning
The University of Melbourne
Parkville, Victoria 3010
Australia
jpaay@unimelb.edu.au

[2] Human-Computer Interaction Group
Department of Computer Science
Aalborg University
DK-9220 Aalborg East
Denmark
jeni@cs.auc.dk

ABSTRACT

The development of a location-based pervasive computing solution requires a solid understanding of the needs of users within the context that the system will operate. A human-centred situated approach to designing mobile computing that is indexed to the users physical and social environments will help build a digital system that can augment user experience of a public place and reduce the complexity and amount of information needed for the human computer interaction.

This paper presents the outcome, in the form of an analytical model, SOPHIA, derived from field studies that were designed to capture this user experience of public places in the built environment, and to create a representation of this understanding in a form that will bridge the gap between analysis of the user's context of use, and directly applying this knowledge to inform design of a location-based context-aware information system.

Author Keywords

User experience, contextual interview, context-aware mobile computing, built environment, indexicality.

INTRODUCTION

The importance of this research is that it provides an understanding of the human experience of a place, by applying methods and tools for gathering this information and presenting it in a form that can inform the design of context-aware applications which are indexed to both the physical and the social aspects of user experience in a public place. Despite many research projects in the mobile HCI community that look at issues associated with designing context-aware mobile computing appliances, very few index to human understanding of the information that already exists in the built environment in which they are operating. Current research into the use of indexical information for mobile device interfaces is focused on spatial and temporal indexicality, but since context is more than just space and time, a need for more research in this area to gather information about the user's tasks and activity has been identified [5], so that social aspects of the use context might also be indexed to. Understanding the role of human activity within the built environment is needed to inform design of information systems that can augment and enrich human experience of a space [3].

Access to the users "knowledge-in-the-world" [8] reduces the need for the representation of complex and extensive information in the user interface. In this way, interface designs for mobile devices, which typically have small screens and limited means of input, can be simplified. To gain this access to user's understanding of their current environment, we need to study how physical and social affordances of a place influence the situated interactions that occur there. To be able to design a mobile computer system that is truly aware of both its physical and its social context, we must first understand the relationship between the built environment and the social roles and rules of the people who inhabit that space [1]. We need to understand the social processes that surround our everyday interactions with others [11].

OVERVIEW OF STUDY

McCullough's [6] typology of everyday situations gives a starting point for a vocabulary that can be used to analyse the social activities of a place, specifically those situated interactions associated with being out on the town: eating, drinking and talking; gathering; cruising; belonging; shopping; sporting; attending; and commemorating. Observing these situated interactions in the field, formed the basis for a study of social interaction in a public place, Federation Square, Melbourne, Australia.

Field Observations

Field observations were carried out using contextual interview [2] and observational ethnographic techniques [3]

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

on location in Federation Square. The participants were three different established social groups of three people who had a shared history of socialising in Federation Square. The first group were observed and interviewed during two three-hour visits to the square, the second and third groups completed only one visit each, the duration of which was also about three hours. The elimination of the repeat visit with the second and third groups was possible due to the fact that the method of conducting the field visit was refined, ensuring that the required variety of situated interactions experienced by the group could be achieved in a single visit. It was also observed that the first group participating in a certain type of activity during a repeat visit acted in a similar way as the first visit, making the same sorts of comments and actions, so the interaction data collected during these repeat activities was not significantly unique for the purposes of this study.

The outcome from these visits was: 6 hours of digital video, recording all questions and responses, initiation of activities, and movement of the group around the square; field notes; and diary reflections about each visit, recorded immediately after that visit. The digital video camera was used as the only sound recording mechanism. These videotapes were then transcribed, detailing all language interactions between interviewer and participants in respect to interactions related to place. Some gestures and actions that seemed important were also recorded in the transcript.

Figure 1. Contextual Interview with participants at Federation Square.

Data Analysis

Initial analysis of the transcript involved open coding, from the grounded theory analysis method [10]. From this process, using axial coding, a series of higher level themes were extracted from the data. These were then transferred to individual pieces of paper for the process of affinity diagramming [2] to draw successively higher levels of abstraction from the data, by grouping and sorting until a small set of high level concepts, representing the essence of the data, and encompassing all lower level themes, was extracted.

The analytical model that emerged from the affinity diagramming process is called SOPHIA (SOcial PHysical Interaction Analysis) and as a formalised understanding of social aspects of the user experience of a public place, is available for use during the design and development of a system that supports and augments the user experience. SOPHIA consists of seven high level concepts, further grouped to represent the three main aspects of social interaction in the physical setting of a public place. The main aspects and their associated concepts are: Knowledge (Knowledge-in-the-world, History); Context (People, Situation, Surrounds); and Actions (Reflection, Extension).

Figure 2. Affinity Diagramming to extract high level concepts during analysis of data.

Design Workshop

A design workshop was then used to take SOPHIA, and the outcome from an earlier expert analysis of the physical environment as experienced by people, the architectural and informational descriptive frameworks called MIRANDA [9], and devise from these representations those aspects of understanding that were most useful for informing design of a system to augment the user experience of Federation Square, as a case study.

Key findings from both SOPHIA and MIRANDA were identified and represented in diagrams and descriptions in the form of "design opportunities". The ethnographer was present during this process to assist in the interpretation of themes based on observational data. These design opportunities were not in the form of design requirements for a specific system, but identified underlying concepts, drawn from that data, that could be used to shape the intentions of design requirements. These opportunities were in themselves a further abstraction of the data, and yet an abstraction using representations more familiar to designers, such as graphs, diagrams and semantic rules, than the high level social and physical models represented as SOPHIA and MIRANDA.

Figure 3. Design Workshop to identify design opportunities from MIRANDA and SOPHIA.

RESULTS AND FUTURE WORK

This process resulted in a rich understanding of the user's context. Not only through the extraction of themes, which provided a rich qualitative story themselves about how people experience physical space and interact with each other in those spaces, but also through the refinement of those themes into high level concepts that can be used to extract design opportunities for use in the creation of more specific design requirements. This study supports the contention of the importance of fieldwork and understanding of mobile contexts as the antecedent to the design and evaluation of these systems [6]. These opportunities could not be identified from casual observation of the place, or even from the interview transcripts, and it is only through the process of grounded theory analysis that this data could be reduced to a quintessential form for informing the design process, and in this way help bridge the gap between user studies and the design of a system that supports those users in the context of use.

To realise the validity of this process, a prototype system is being designed and evaluated using the outcomes from this study, the "Just-for-Us" system. This involves defining system requirements, in light of the design opportunities already identified by the process and as this system will be embedded in a mobile form factor, some functionality will be drawn from existing mobile guide designs. This mobile system will be taken into the field study location, Federation Square, Melbourne, Australia, and evaluated with users to measure its applicability, usefulness, and usability, in augmenting the user experience of that public place.

ACKNOWLEDGMENTS

I would like to thank the Department of Computer Science at the University of Aalborg for hosting my study visit, during which time the design workshop was accomplished. Also, Jesper Kjeldskov for his supervision and assistance during my stay in Aalborg, and Steve Howard and Bharat Dave for their supervision during my PhD research program.

REFERENCES

1. Agre, P. Changing Places: Contexts of Awareness in Computing. *Human-Computer Interaction 16* (2001), 177-192.
2. Beyer, H., and Holtzblatt, K. *Contextual Design: defining customer-centred systems*. San Francisco, Calif., Morgan Kaufmann, San Francisco, Calif., 1998.
3. Blomberg, J., Burrell, M., and Guest, G. *An Ethnographic Approach to Design*. Lawrence Erlbaum Associates, Inc., New Jersey, 2003.
4. Ciolfi, L. Understanding Spaces as Places: Extending Interaction Design Paradigms. *Cognition Technology and Work 6, 1* (2004), 37-40.
5. Kjeldskov, J. Just-in-Place: Information for Mobile Device Interfaces. In *Proceedings of 4th International Symposium, Mobile HCI 2002,* Springer-Verlag (2002), 271-275.
6. Kjeldskov, J. and Graham, C. A Review of MobileHCI Research Methods. In *Proceedings of 5th International Symposium, Mobile HCI 2003*, Springer-Verlag (2003), 317-335.
7. McCullough, M. On Typologies of Situated Interaction *Human-Computer Interaction 16* (2001), 337-349.
8. Norman, D. *The Design of Everyday Things*. Basic Books, New York, 1988.
9. Paay J., and Kjeldskov J. Understanding and Modelling the Built Environment for Mobile Guide Interface Design. To appear in *Behaviour and Information Technology* (2004).
10. Strauss, A., and Corbin, J. Basics of Qualitative Research, Sage Publications, Newbury Park, CA., 1998.
11. Tamminen, S., Oulasvirta, A., Toiskallio, K., and Kankainen, A. Understanding Mobile Contexts. In *Proceedings of 5th International Symposium, Mobile HCI 2003*, Springer-Verlag (2003), 17-31.

Multimodal interactions in physically modeled sound synthesis

Stefania Serafin Juraj Kojs Amalia De Goetzen Rolf Nordahl

Medialogy
Aalborg University Copenhagen
Lautrupvang 15,
Ballerup, 2750
sts@media.aau.dk

ABSTRACT

We propose new approaches to interact with physically modeled sound models. Tangible interfaces are used to control physically modeled sonic models. Moreover, a visualization technique allowing the user to experience immediate feedback from the model is proposed.

Categories and Subject Descriptors

H.5.5 [**Sound and Music Computing**]

General Terms

Performance, Design, Human Factors.

Keywords

Physically modeled models, tangible bits, playability.

1. INTRODUCTION

In the recent years, the physically modeled sonic models have attracted substantial amount of attention in the sound synthesis and computer music community [1, 3]. This attention lies in the fact that the parameters of the physically modeled models can be controlled in the similar fashion as their physical equivalents. The performers are invited to interact with these instruments in a manner they are accustomed to with the physical instruments, whether these are music instruments or everyday life sonic objects.

In fact, the playability regions of the physically modeled instruments are directly derived from their physical counterparts. Yet, despite the user-model intuitive interaction instigated from such a relationship, the sound designers remain struggling in applying the physically modeled models in the multimedia applications.

One of the reasons for such hesitance is perhaps conditioned by unpredictable complexity of sonorities produced by the model with varying parameters. The topic is discussed in [5]. Additionally, the models' control has stimulated discussion. Frequently traditional interfaces such as mouse and keyboard have been adopted. These, however, essentially diminish the models' expressivity along with the possibility of natural motoric interaction between the user and the model.

To address the control difficulties, we propose a visualization-assisted strategy for physically modeled sonic models. We will discuss how the technique operates with the physical model of a violin, although it is suggested that similar approach can be extended to other musical instruments and sounding objects in general.

Interactive visualization-assisted technique allows visual representation of the behavioral variation in the model's parameters. Technique called playability region has been employed to validate physically modeled models. Here the model's behavior was analyzed and compared to the behavior of its physical counterpart. Accordingly, we propose to use the playability region to visualize the output waveforms of physically modeled models. In practice, the user attains immediate and intuitive feedback of the resulting sonority while perceiving the visually displayed waveforms—the result of the input parameters.

2. TANGIBLE INTERFACES

Tangible bits were previously proposed as controllers of physics-models in [1]. Everyday objects are preferred in such interaction (as opposed to sophisticated musical instrument controllers), as they are both easily accessible and intuitive in contact to a variety of unskilled users.

Tangible bits can be assigned to perform various tasks and mapped into a variety of instrument's parameters. These are described and evaluated in the following sections.

Tangible User Interfaces (TUIs) [1] are an emerging field in Human Computer Interaction. Physical objects are used for the control of digital information. Different musical applications have already applied TUIs as a control interface such as the Audiopad [2], the Reactable [3] and the Audio d-Touch [4]. In these systems, tangible objects control sampled sounds and signal processing techniques. To our knowledge, there exists no system in which tangible objects would control physical models. In accordance with the model's design, the physical motion of the tangible objects can be effortlessly mapped to its sonic characteristics. In this way, a larger body of users can interact with the virtual sonic models via simple tactile objects.

3. FRICTION MODELS

Friction is a result of force between objects in contact. In most engineering applications, friction is reduced to and identified as a source of noise and unstable vibrations. From a sonic point of view, friction stimulates the sound production of numerous musical instruments such as bowed violin and everyday objects such as musical saw, Tibetan bowl, train wheels, squeaking door and rubbed wineglass [5].

In this paper we proceed to look at possible applications of these models in multimodal HCI.

4. VISUALIZING PLAYABILITY

Playability can be defined as the region of the multidimensional space given by the input parameters of a physical model where high-quality tone is obtained. Traditional playability evaluation techniques propose a sound model that parameter space is visualized in the time domain. Schelleng [8] proposed an evaluation for the playability of a real bowed string instrument.

He suggested plotting the playability in a two dimensional space, in which the bow velocity was fixed and bow force and position varied between 0 and 5 N and 0 to 0.4 respectively (where the bow position is normalized, i.e. 0 represents the nut, that side of the string attached to the upper end of the instrument, and 1 represents the bridge, while 0.5 represent the middle of the string). Schelleng proved that there is also defined region in which the Helmholtz motion can be achieved.

The Helmholtz motion is recognized as the "sawtooth like" motion—an ideal motion every player is trying to achieve.

The playability region as defined by Schelleng is shown in Figure 1.

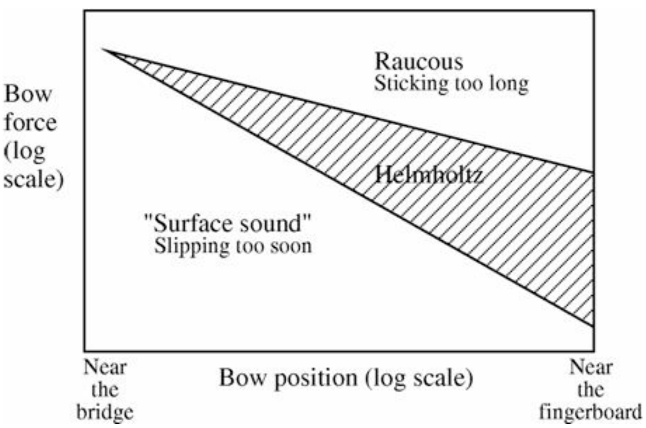


Figure 1. The Schelleng diagram

Figure 2 shows the playability region of the simulated bowed string physical model described in [6]. The playability region was calculated by maintaining a constant bow velocity of 0.05 m/s, and changing the bow force and bow position inside the parameters range described above. It is noticeable that there is an area of the playability region (which is the one denoted by Helmholtz in Figure 1) in which "good tone" is obtained. Above such region, noisy waveforms are obtained due to the fact that the player is using a too high bow force to control the model.

Below the "good tone" region, so-called surface sounds are obtained, due to the fact that the bow force is not high enough.

It can be observed that the simulated playability region displayed in Figure 2 is in good agreement with the playability region measured by Schelleng on a real instrument, and displayed in Figure 1.

This shows that the simulated bowed string physical model behaves in the same way as the measured instrument.

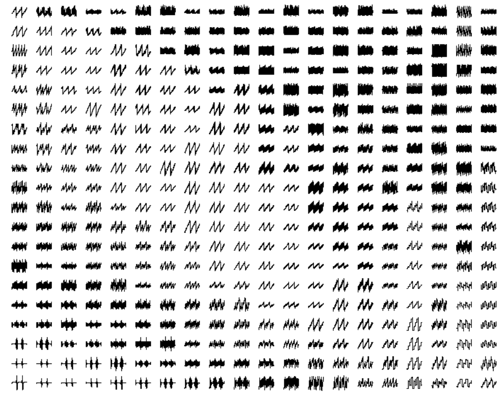


Figure 2. Playability waveforms of a violin physical model, in the position versus force plane. A constant bow velocity of 0.05 m/s is kept in the simulations.

5. TOWARD MULTIMODAL PHYSICS BASED INTERFACES

We implemented our system using the Max/MSP and Jitter [7] software package.

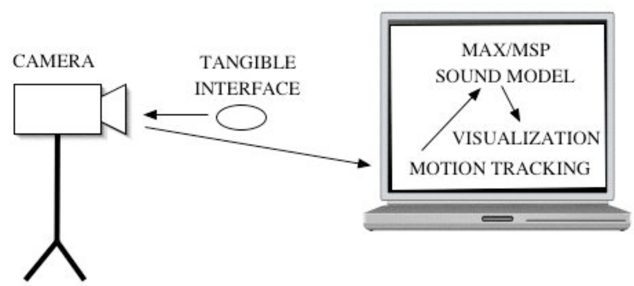


Figure 3. The structure of the multimodal interface. A camera tracks the position of the tangible interface, which controls the physical model whose parameters are visualized in real time.

As tangible objects, we used hand-handled therapy devices such as squeeze balls.

In the process, video camera tracks the tangible objects, and motion tracking algorithms implemented in Jitter allow detection of the variation in position of such objects.

Such variations are mapped into sonic parameters of the physical model. In the same time, the parameters are mapped in the visualized playability region similar to the one shown in Figure 2.

6. APPLICATIONS AND TASKS

The multimodal interface described in the previous section can be used in different applications, from interactive computer music to new explorations in HCI. Such applications are described in the following.

6.1 Interactive computer music

The development of new interfaces for musical expression has seen an increase of interest in the computer music community. Often the interest is focused on interfaces, which mimic existing musical instruments.

One of the disadvantages of new interfaces for musical expression is the fact that most interfaces are not used by a broad spectrum of users. More often, interfaces are the product of a single developer and are hard to reproduce. We believe that by involving everyday objects the development of new musical instruments can be shared by a wider range of users and audience. Engaging unskilled participants in computer music performance was previously employed in Garden of the Dragon. [7]. In this interactive composition, the singing tube, a plastic corrugated tube utilized for pluming and insulation, serves both as an instrument and control interface for a singing tube physical model. The participants were involved in motoric manipulation of the tube and transparent control of the model's sonic parameters. The performance succeeded in bridging the interactive process between a sonically expressive physical instrument and its computer correspondent.

The singing tube interface was characterized by a steep learning curve, as it allowed different fairly unskilled musicians to become engaged in new music performances. We observed that the use of different tangible bits such as everyday objects in harmonization with physical models proved to have a strong potential for HCI music performances. Due to their simplicity, they allow broader audience to become engaged and interested in dexterous computer-assisted music functioning.

The performances have further potential to evolve into a task-oriented games and navigational systems. In particular, we find our proposals suitable for further explorations in audio-visual comprehension of physical and virtual music relations.

7. ACKNOWLEDGMENTS

The authors would like to thank Professor Jim Woodhouse at Cambridge University for stimulating discussion on the playability of virtual bowed strings.

8. REFERENCES

[1] Ishii, H., Ullmer, B., "Tangible bits: towards seamless interfaces between people, bits and atoms", in *CHI'97*, pages 234 -- 241, 1997.

[2] Pattern, J., Recht, B., and Ishii H. Audiopad. A Tag-based Interface for Musical Performance. In Proceedings of NIME Conference, Dublin, Ireland, May 24-26, 2002.

[3] Jordà, S.: *Sonigraphical Instruments: From FMOL to the reacTable**. The Proceedings of the 3rd Conference on New Interfaces for Musical Expression (NIME 03), Montreal, Canada, 2003.

[4] Costanza, E., Shelley, S.B., Robinson, J., "Introducing Audio d-touch: a Novel Tangible User Interface for Music Composition and Performance", in Proceedings of the 6 th international conference on digital audio effects (DAFx-03), London September 2003.

[5] Serafin, S. *The sound of friction: real-time models, playability and musical applications.* Ph.D. Thesis, Stanford University, Stanford, CA, 2004.

[6] S. Serafin, J. O. Smith, III, and J. Woodhouse, ``An investigation of the impact of torsion waves and friction characteristics on the playability of virtual bowed strings'', in *IEEE Workshop on Signal Processing to Audio and Acoustics, New Paltz, NY*, New York, Oct. 1999, IEEE Press.

[7] J. Kojs, S. Serafin, "The voice of the dragon: A physical model of a rotating corrugated tube. Proceedings of Digital Audio Effects Conference, London, 2003.

[8] Schelleng, J.C., "The bowed string and the player," J. Acoustical Soc. of America, vol. 53, no. 1, pp. 26–41, Jan. 1973.

[9] Zicarelli, D. An Extensible Real-Time Signal Processing Environment for Max, Proceedings of the International Computer Music Conference, Ann Arbor, Michigan, 1998.

Design Souveniring: An Experimental Approach Aimed at Improving Cross-Cultural Awareness in Global User Interface Design

Lori Webb
Gothersgade 145, 2. sal
1123 Copenhagen K
+45 33 32 88 31
lmwebb@yahoo.com

ABSTRACT
In this paper, I will describe the conceptualization of an experimental design approach aimed at improving cross-cultural awareness in global interface design. The concept is based on an interdisciplinary approach that builds upon global user interface design, design research, experience design, culture theory, intercultural communication, cultural probes adaptations and credibility.

Keywords
Global interface design, cultural dimensions, intercultural communication, creative design practice, cultural probes, design souveniring.

INTRODUCTION
This paper will present various sections instrumental in the concept development for the *design souveniring* approach. This includes global user interface design, culture vs. intercultural communication theory, cultural probes, design souveniring, conclusion and further considerations.

BACKGROUND
The need for the development of new global design tools has been documented in previous literature. [1, 2, 3, 4] Current trends in interface design lean towards the design of more user appropriate design experiences. [5] New trends in design research suggest increased innovation and interdisciplinary involvement in creating design solutions. [6]. Countries around the world have varying cultural values and aesthetic perceptions [7, 8].Intercultural communication theory, cites awareness and experience as two crucial elements for successful communication between cultures. [9, 10]. The need to improve online credibility is a growing issue which designers increasingly need to address [11, 20].

Fourth Danish HCI Research Symposium, November 16, 2004
Aalborg University, Denmark.

And finally, the integrated use of creative design methods such as Cultural Probes has proved to not only be useful in inspiring design and informing design [12] but, is also an enjoyable method of conducting design research [13, 14, 15].

Although web design literature encourages designers to consider intercultural needs [16, 17]; the majority of web site designs are not reflective of these directives.

The increasing pressure on designers to produce more and more user-tailored designs [5] grows everyday in tempo with user needs, expectations, demands, interest and technological savvy.

Design is a communication form and by ignoring diverse communicative and cultural needs of international users condemns global user interfaces to failure.

The aim of this project was to create an easily implemented and interesting approach for integrating cultural awareness into creative design practice as a means of improving the level of user appropriateness [1], user experience [5] and credibility [11] for global user interfaces.

Global User Interface Design
Website design literature is quite plentiful but, literature on global, international, cross-cultural, intercultural, multiethnic interface design is greatly limited with tools to improve global interface design virtually none existent.

Cultural background is a key determinant of aesthetic perception. National aesthetics are often rooted in impact, artisan influence, trends and styles of historical design movements. The clues of which exist in a cultures' architecture, art and design. Things such as composition, color, typography and geometry can vary greatly from country to country providing important insights into cultural preferences. [1]

Culture Theory vs. Intercultural Communication Theory
Dutch theorist, Geert Hofstede conducted a study for IBM in the 1970's that established 4 cultural dimensions: Power Distance, Individual vs. Collective, Masculine vs. Feminine

and Uncertainty Avoidance. A later study, provided the fifth dimension, Long Term Orientation.

Hofstede's dimensions have been applied to varying research areas including some relevant to the area of web design practices and the scope of this project. The influence of culture on Human Computer Interaction [19] and Interface Design [8,18] each utilized aspects of Hofstede's Dimensions in pursuing their research.

While Hofstede's cultural dimensions provide some insight into more homogenous national cultures it is just one snapshot of any given cultural picture.

Edward T. Hall presents cultures complexities. describes culture as a communication system incorporating three aspects: its general structure, its various elements and its content. Content, can then be further subdivided into three areas: sets, isolates and pattern [9]. Hall also views cultures in terms of time, space and context. All of which would seem to be a germane application of culture theory to global user interface design considering that the world wide web exists within an multi-cultural communication forum.

Culture can also be divided into two areas. Implicit culture, which contains the elements of culture that are most obvious and explicit culture that is harder to describe and most entrenched in our cultural composition [9]. Understanding the differentiation of these two areas can be a critical factor in the success of intercultural communication [10] as well as the development of user appropriate global user interface design.

Cultural Probes

The concept behind Cultural Probes evolved out of an artistic rather than scientific background. Cultural Probes like astronomical probes are released into the environment to collect and return data. The intention of the probes were initially created to only provide design inspiration.[12] but, have since been used in conjunction with other methodologies to inform as well as inspire design.

Use of adapted forms of Cultural Probes [13,14,15], attempted not only to create more user appropriate designs but, also to create more user, as well as designer, appropriate methods. The empathy probe, adapted the nature of its probe to provide empathic understanding through the integration of experience designer, Elizabeth Sander's say, do and make approach [14]. In the Design with Care project, cultural probes were integrated with ethnographic methodologies to navigate sensitive settings [15]. In the NetWorking.Kids project, the probes were adapted for improved and natural accessibility to their target group, kids, by integrating use of the mobile telephone into the project. Kids mobile phone usage in Denmark [13], where the study took place, is prolific and therefore an appropriate extension of the probes.

What is Design Souveniring?

Design Souveniring is a creative design approach, in the same vein as Cultural Probes (12). Like "probes" it aims to inspire design process but, unlike "probes" it also aims to inform design knowledge, broaden design exposure and stimulate designer awareness and sensitivity of cross-cultural design issues.

In the context of this project, the concept of "souveniring" is a broad use of the travel metaphor to encourage and enable cross-cultural design awareness and experience. The term Design Souveniring, therefore, is to be understood in two parts:

- The process of collecting digital or actual souvenirs to inform cross-cultural experience.
- The process of "stealing inspiration" from these digital or cultural artifacts to enhance cross-cultural design practice.

The *design souveniring* approach provides quick exposure to culturally relevant clues relevant to cross-cultural design. The *design souveniring* data as well as process can provide valuable "snapshots" into the global user experience. The designer interaction with the digital or actual souvenirs will provide texture rich cultural input which could later be utilized to inspire or inform design.

Similar to travel, there are many ways to arrive at your destination. Design souveniring can be designer lead or target user lead. It can be embarked upon individually or collectively. A combined approach, rooted in intercultural communication [10] can provide both an emic (insiders view) and etic (outsiders view) of culture that can provide a more holistic picture to better enhance the design of global user interfaces.

In real life, travel and souvenir collecting occur at various budget levels. *Design souveniring* is similarly linked to priorities and resources. Time can also be a factor and is something of a commodity for everyone in the 21st Century and therefore a relevant concern when developing design development approaches. Integration of *design souveniring* in the initial stages of development, have the potential to provide deep perspectives into cross-cultural design solutions that can save valuable time and resources later in the process. Continued use of the *design souveniring* approach throughout the designer research journey can also prove to be a media rich resource that provides an opportunity to expand design experience in tangent with designer goals, interest and investment.

Designer lead

The aim of the designer lead use of design souveniring is to encourage cross-cultural design understanding through personal experience. And to provide the designer with a rich, self dictated opportunity to amass not only design relevant but, personally interesting artifacts.

The actual experience of the *design souveniring* process has the potential to provide quick snapshots into the often, frustrated reality, of being an intercultural web site visitor, in much the same way as Empathy Probes provides an empathetic experience for its' users [14].

A mixed media approach incorporating both the digital and physical world is the most ideal. The physical artifacts gathered can provide a richer textural element that would be limited in the digital world. While the digital world provides an extensive and economical accessibility to artifacts with the added advantage of easy distribution to, say, international colleagues, researchers or clients.

Design souveniring can be readily used in the short term for quick insight into cross-cultural design related issues. Idealistically, it will be integrated into international and global user interface designer consciousness and iteratively practiced to enhance or enlarge the cultural profiles. Created over time, this could result in a heightened cultural sensitivity in international design practice on the web. The artifacts created through these processes can be used to deepen the level of the intended design, encouraging improved intercultural communication and self reflection.

The *design souveniring* data, can then be utilized to better inform interviews, questionnaires and focus groups with target users for validation.

To Inform Designers:

-Online "travel"assistance

-Search Engines/Directories

-National Tourist Boards

-National Post Offices

User lead

The aim of User lead Design Souveniring is to initiate a dialogue between the user and the designer on relevant intercultural design and communication issues. It involves users in an iterative process from the earliest stages which can further assist in site appropriateness and increased site credibility.

User lead *Design Souveniring*, is procedurally close to the execution of the Cultural Probes method [12]. Users are presented with information and then asked to embark on a cultural journey (domestic-self reflective or international-cross-cultural experience).

The participants actively document their experience by creating *design souvenirs* that designers later can use to remind them of user presence, needs and desires. The information can be used to inspire design, to analyze design and to qualify or quantify design research. See below for further exemplification. Effective planning of the sample target group can improve design integrity by providing a range of representative perceptions. Single cultural targets can provide depth into cultural understanding and design inspiration. Cross-cultural targets can provide insight into the communicational barriers of intercultural designs.

Plans for a Focus Group Approach

Groups can be used to introduce new input on a culture and confirm or dissipate stereotypical perceptions and cultural bias. Group input can provide rich, textural artifacts to expand typical (in the box) design processes into creative problem solving atmosphere. Group input can provide valuable insight into the macro and micro views of cross-cultural interaction.

Groups should be composed of geographical and/or cultural representatives to ideally represent the country or culture. For example, a Danish group would ideally consist of representation from Zealand, Jutland, Funen, Bornholm, the Færø Islands and Grønland. This to provide a broader understanding of the "total" cultural representation and issues.

Materials

Informed consent, video, scissors, paper, glue, tape, stapler, notebook pages, "own data", tourist bureau materials (maps of country, maps of capital, various regional and tourist attraction brochures. postcards, Stamps (stamp collage), artifacts from famous national writers, poets, musicians, philosophers, scientists, statesmen, artists, designers, music, pop culture representation and thank you gifts.

Procedure

•Contact regional representatives

•Give or send "Souveniring" Kit (see contents below)

•Give participants fixed time frame in which to collect data (i.e. week)

•Have participants send back "Souveniring" Kit

•Send "Thank You" gift to participants

• Follow up interviews for analysis validation

Souveniring Kit Contents

•A Presentation of Design Souveniring

•Presentation of Kit

•Informed Consent

•Tourist literature (maps, brochures, postcards, photos)

•Notebook

•Stickers (to notate areas on map—round colored)

•Highlighter

•Assorted Paper (color/texture)

•Camera (documentary: your culture/your connection to other cultures)

•Questionnaire:

•Own culture commentary (emic view)

•Other culture commentary (cross-cultural issues) (etic view)

•Similar and different (for testing intercultural issues)

•Pre-postage paid envelopes

•Pre-postage paid postcards

•Disposable Camera

•Recorder

•Keep a Log of SMS and data collection

•Designer Loggers: Catalogue activity.

CONCLUSION

Cross-cultural design is influenced by experience. Cross-cultural experience provides insight to cross-cultural expectations and conventions. Cross-cultural observation provides windows into cross-cultural life; the people, the environment, the needs, the frustrations, the desires, the dreams. Improved cross-cultural awareness can aid in improved cross-cultural understanding and better intercultural interactions.

FURTHER CONSIDERATIONS

The *design souveniring* approach is still in the conceptual phase and needs further use in the field to ultimately evaluate its validity and reliability as a developing method for design research. The next phrase of the process, involves testing the design model and analyzing the feedback.

Recommendations for further studies would be to enlist or form interdisciplinary cross-cultural design teams to use and review the *design souveniring* approach, helping to evaluate its operational use and to increase the awareness of the need for improved cultural sensitivity in international design practice.

A collaborative cross-cultural effort, would allow for data to be analyzed with a more holistic global reflection and allowing for further refinement and development of improved creative practices addressing global user interface design and enhanced intercultural communication for a truly world wide web.

REFERENCES

1. Fernandes, T. Global Interface Design. Academic Press. Chestnut Hill, MA. 1995
2. Santos, C. Hispanic Culture in Design Research: Design Research Methods and Perspectives. Ed. Brenda Laurel. MIT Press. Cambridge, MA. 2003.pp.55-62.
3. Lipton, R. Designing Across Cultures. HOW Design Books. U.S. 2002.
4. Nielsen, J. Designing User Interfaces for International Use. Elsevier Science. Amsterdam, The Netherlands. 1990.
5. Shedroff, N. Experience Design 1. New Riders.Indianapolis, IN. 2001.
6. Laurel, B. Design Research Methods and Perspectives. MIT Press. Cambridge, MA. 2003.
7. Hofstede, Geert. Kulturer og organisationer-overlevelse i en graenseskridende verden.Grafik A/S Denmark. 1991.
8. Marcus, A. Cultural Dimensions and Global Web Design. AM + A. CA. 2001
9. Hall, Edward T. The Silent Language. Anchor Books. New York. 1990.
10. Herlitz, G. KulturGrammatik. Munksgaard. Denmark. 1999.
11. Fogg, B.J. Persuasive Technology. Morgan Kaufmann. San Francisco, CA. 2003.
12. Gaver, B. Dunne, T. and Pacenti, E. Designing Cultural Probes. In Interactions, Jan/Feb.1999. pp.21-29
13. Iversen and Nielsen. Digital Cultural Probes. DK. 2002.
14. Mattelmäki, T. and Battarbee, K. Empathy Probes. In: Proceedings of the Participatory Design Conference, 23-25.6 2002, Malmö Sweden. 2002.
15. Crabtree et al. Design with Care.UK. 2003
16. Lynch, P. and Horton, S. Web Style Guide. Yale University Press. 1999.
17. Nielsen, Jakob. International Usability.Nielsen Norman Group.US. 2000.
18. Dormann, C. and Chisalita, C. Cultural Values in Web Site Design. Retrieved www.cs.vu.nl/~claire/Hofstede-dormann.pdf. Amsterdam, NL 2003
19. Yeo, A. World-wide CHI:Cultural User Interfaces, A Silver Lining in Cultural Diversity. SIGCHI. Vol 28. No.3. July 1996.
20. Fogg,B.J.Prominence-Interpretation Theory: Explaining How People Access Credibility. A Research Report from the Stanford Persuasive Technology Lab, Stanford University. Retrieved March 1st, 2003 from http://credibility.stanford.edu/pit.html.20