Proceedings of

Umeå's 13th Student Conference in Computer Science

Edited by

Johanna Högberg

DEPARTMENT OF COMPUTING SCIENCE
UMEÅ UNIVERSITY

Department of Computing Science
Umeå University
SE-901 87 Umeå, Sweden

ISSN 0348-XX
UMINF 09.13

Umeå Tryck & Media, Umeå, 2009

Preface

Umeå's Student Conference in Computer Science is the highlight of the conference course in our Computer Science curriculum. The objective of the course is to give the students a forum where they can participate in scientific research. The conference format was chosen to provide a realistic environment in which the research results can be presented.

A student who participates in the course first selects a topic and a number of research questions that he or she is interested in. If the topic is accepted, then the student outlines a paper and composes an annotated bibliography to give a survey of the field. The main work consists in answering the research questions and reporting the results in a draft version of a paper. The draft receives at least two reviews written by department staff. If the reviews are favourable, the student is allowed to present the work at the concluding conference, and to submit a final version to the proceedings. The course thus gives an introduction to independent research, scientific writing, and oral presentation.

This year was the thirteenth offering of the course, with a total of 27 registered students. Of these students, 20 participated actively in some part of the course, and 18 eventually submitted a full paper. Out of the these submissions, 15 were accepted for publication in the proceedings.

We are grateful to the reviewers who helped us to review the submissions within a very short time frame.

Umeå, May 2009

Johanna Högberg
Program Chair
USCCS 2009

Program committee

Johanna Högberg
Håkan Gulliksson
Dipak Surie

Reviewers

Suna Bensch
Erik Billing
Henrik Börstler
Henrik Björklund
Håkan Gulliksson
Johanna Högberg
Lars Karlsson
Lena Palmquist
Dipak Surie

Contents

Properties of “Good” Java Examples

Nadeem Abbas

Abstract: Example programs are well known as an important tool to learn computer programming. Realizing the significance of example programs, this study has been conducted with a goal to improve the quality of example programs. We make a distinction between “good” and “bad” examples, as badly designed examples may prove harmful for novice learners. We investigate example quality and argues about those factors which impact it, either positively or negatively. We provide a set of desired example quality attributes, and describes problems and mistakes observed in commonly used example programs, in context of object oriented programming. Finally, we provide a comprehensive checklist of measurable criteria and suggests appropriate metric to give quantitative measure of the example quality.

1 Introduction

Learning something new is a human instinct and great fun, but it often involves complications and challenges. Being students of computer science, we learn lot of new concepts, tools and technologies, and often face problems as well. Whenever there is a problem and a will to resolve it, there is a way, i.e. problem solving. Programming, in fact, is nothing but a problem-solving approach. However learning how to program in itself is a big problem confronted by many people.

Learning by examples is a well established pedagogy [11]. While learning computer programming, example programs play a vital role and can be used as an effective tool to explain complex concepts which otherwise are difficult to comprehend. According to Börstler et al. [4], example programs act as templates, guidelines and inspirations for learners. Examples are generally believed to make learning easier by reinforcing fundamental concepts and eliminating confusion or misconception. However examples are not always equally good for learning [27]. Badly designed examples can mislead students to build myths and misconception. Use of good programming examples is extremely important as the hypothesis is supported by [4, 27, 3]. With in this context, the key research question of this study is:

What are the problems with commonly used programming examples? and what should be the desired attributes for good examples, in context of object oriented programming?

In order to proceed and address the research matter, three primary objectives are identified, as listed below:

- Review the literature to understand the relationship between example quality, comprehension and complexity; principles and guidelines for object-oriented pedagogy and software development in general.
- Identify common problems or difficulties while learning fundamental object oriented programming concepts, by analyzing existing text book examples and collecting student comments.
- Propose desired attributes for examples programs that can be used to distinguish between "good" and "bad" examples.

Here the terms *example* and *example program* are interchangeable and object-oriented is the intended programming paradigm. Example program refers only on actual source code excluding the supporting textual description or any visual aids, such as UML diagrams.

Section 2 provides annotated description about related literature and shows achievement of first objective. Section 3 provides explicit definition of good and bad examples. Section 4 describes example quality and its relationship with complexity, while its subsections 4.1 and 4.2 provides a candidate list of example quality attribute and example quality measures, respectively. Section 6 shows concern about empirical evaluation and validation of proposed example quality attributes. General discussion about related issues is given in Section 7, while Section 8 presents overall conclusion about the study.

2 Related Work

Researchers in academia and industry have done plenty of work regarding software quality. A large number of software metrics and models [29, 9, 22, 8, 18, 16, 15, 28] have been proposed, which certainly helps to improve software quality and maintenance efforts. However there is relatively little work done in the academic context, to determine desirable characteristics of example programs used as learning beacons in academic world. There is no universal standard defined for example programs used by educators during lectures, tutorials, and lab-exercises.

Malan [27] and Kolling [20] has pointed out typical problems with example programs, particularly with respect to object oriented paradigm. Malan and Halland urge educators and researchers to carefully design examples programs for teaching purpose [27]. They claim to treat examples as products selling concepts to the novice learners. Their work [27] identifies four common problems. The first issue pointed is that of too simple and abstract examples, which follow a non-realistic general approach, with no particular objective. The second problem identified is of example programs which are more realistic and too complex to understand. Such examples have unnecessarily complex constructs that create troubles for novice learners. The third issue is that of inconsistency, where example programs overthrow earlier taught concepts, instead of reinforcing them. The last issue concerns those examples that are so badly constructed that they even undermine the concepts they are trying to teach. However Malan and Halland work lacks empirical evaluation and does not provide any explicit solution or measures to improve the quality of example programs.

This study is principally inspired from the work [4, 3, 5] of Börstler et al. In these articles, they have focused to investigate problems and desired attributes of "exemplary" examples, which can be applied to formulate a measurement framework to indicate quality of the example. Understandability, effective communication and adherence to external factors like accepted design principles are the three basic desired properties described by [3]. The article [3] provides a good picture to distinguish good from bad example by comparing two example codes "Beauty" and "Beast". It proposes a software readability metric called "Software Readability Ease Score" (SRES) based on the idea of Flesch Reading Ease Score [13]. Their work has greatly focused on readability property while overlooked certain other attributes of quality e.g. simplicity and adherence to the problem domain. The article [4] provides more comprehensive classification of quality attributes into three categories of technical, object-oriented and didactical quality and presents empirical results of tests performed on selected five candidate examples. However the study is required to be conducted on a large scale, evaluating a good number of text book examples, to actually validate the proposed measurement instrument.

Nordström's licentiate thesis [30] is closely related to this work and really a useful study to proceed. It provides a survey of the literature, helps to identify characteristics and principles of object orientation and proposes six heuristics for designing object oriented examples. Although it briefly describes different object oriented metrics but does not provide explicit mapping between proposed heuristics and these object oriented metrics to have a quantitative measure of example quality. However this study identifies quality attributes for example programs as well as suggests their corresponding quantitative measures in Section 4.2.

3 What is Good and What is Bad?

It is bit tricky to define or find exact definition of "good" or "bad" example. Börstler et al. has defined good example as:
"An example that relates to the needs of novices in programming and object-orientation" [5].
Following is the definition of "bad" example, derived from [3]:

Bad Example An example program which imposes risk of misinterpretations, erroneous conclusions and may lead towards misconceptions causing ultimate problems in learning.

This on the other hand can be extended to define good example as:
"An example program which does not have any risk of misinterpretations, misconceptions and erroneous conclusions, is a good example".

"Good" and "bad" are two mutually exclusive adjectives of quality. Classification of good and bad examples can be simplified by identifying quality factors or attributes. These quality attributes combined with above given definitions, results in more comprehensive definition of "good" example, as below:

Good Example An example that truly satisfies desired quality attributes, learning needs of novices, without any risk of misinterpretations, misconceptions and erroneous conclusions.

4 Example Programs's Quality and Complexity

"Good" and "bad", the two attributes used to describe example programs in this study, are in fact measures of quality on an ordinal scale. One may find a number of different software quality definitions in the literature. A comprehensive definition of quality with respect to computing is given by ISO8402. It defines quality as: *the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs. Not to be mistaken for degree of excellence or fitness for use, which meet only part of the definition.* In context of this study, the product or service is an example program.

According to the above definition, there are two important aspects of quality, the first one is: totality of features and characteristics, while the other is: to satisfy the stated or implied needs. Characteristics and features of an example program are actually the quality attributes described in Section 4.1. *Needs* for an example program are mainly the learning objectives with respect to the particular programming paradigm and computer science in general. A joint task force of the ACM and IEEE Computer Society [40] has defined detailed learning objectives for Programming Fundamentals course. Example programs are expected to satisfy these objectives, therefore. Hence learning objectives must be considered while establishing example design heuristics as well as defining example quality metrics.

Quality of an example program is severely affected by its underlying complexity. Complexity is analogous to "beauty", as they say *"beauty lies in the eyes of the beholder"*. The same is true for the complexity, because the level of complexity perceived varies from person to person. An example program that looks simple to some of the students may create problems in understanding for rest of the students. In literature, program understanding is often referred as program comprehension. *Program Comprehension* is a cognitive process as a result of which one learns, acquires knowledge and gains understanding. In his article [34], Rugaber describes program comprehension as a process through which one acquires knowledge about computer programs. This study treats difficulties in understanding or learning form an example programs as a complexity of that example.

Basili [1] defines complexity as *a measure of the resources expended by a system while interacting with a piece of software to perform a given task.* According to Kearney et al. [18], If the interacting system is a computer, then complexity can be characterized in terms of execution time and storage required to perform the computation, but if the interacting system is a programmer, then complexity can be characterized in terms of difficulty of performing tasks such as coding, debugging, testing, or modifying the software. So we can say complexity of an example program is the measure of difficulties faced and cognitive resources consumed by a learner while using that particular example program.

This study identifies two main categories of example complexity, structural complexity and cognitive complexity. Structural complexity covers difficulties due to syntactic structure, program size, control flow, decision structures and class hierarchy. Therefore while designing a metric to measure example program complexity, one should take care of these factors of structural complexity. Halstead's software science [15] and McCabe's cyclomatic complexity [28] are the mostly used and good choices to measure example complexity based on program's syntactics and decision structure. In order to measure structural complexity of object oriented programming examples, one should also consider the impact of object orientation. Inheritance, polymorphism, data encapsulation and abstraction have a dominant affect on program structure, that impacts structural complexity of the program as well. One may use measures of *Depth of Inheritance Tree (DIT), Coupling between Objects (CBO) and Lack of Cohesion in Methods (LCOM)* defined by Chidamber [12] to evaluate structural complexity. Table 1 and Table 2, enumerate objective measures for structural and cognitive complexity respectively.

Cognitive complexity deals with cognitive problems in learning. Amount of Information, familiarity and recognizability of the concepts, functional and variable dependencies, cuing level, ambiguity factor and expressional complexity are the dominant factors adding towards cognitive complexity of an example. Cant et al. in [10] refers to cognitive complexity as those software characteristics which interact with programmer characteristics to affect the level of resources used. In the same article [10], they have described mathematical formulas to measure cognitive complexity of a program. In theory their measures of cognitive complexity are very comprehensive and logical, however bit hard to understand and implement. We are unable to find any system implementing these measures of cognitive complexity. In future we are interested to implement and evaluate these measures of cognitive complexity by Cant et al.

Complexity of example program is affected by the principles of a particular programming language or paradigm. As in case of object oriented paradigm, programs are modularized on the basis of objects, where modularization results in less chunking but more tracing efforts. The data encapsulation property of object oriented paradigm helps to reduce cognitive complexity by reducing both forward, to follow ripple effects, and backward tracing efforts, to resolve variable dependencies [9]. Inheritance and polymorphism, being key concepts of object orientation, have almost balanced effect on example complexity. They increase cognitive complexity by making functional dependencies difficult to resolve, but at the same time they reduce structural complexity by improving semantic consistency and decreasing the number of lines of code. Inheritance results in functional dependencies, where a method implementing some behavior might be contained in the parent class, and so on. Method calls also become difficult to trace as a result of polymorphism.

4.1 Example Quality Attributes

Example quality attributes is a set of desired properties, which a "good" example program is required to satisfy. The IEEE-1061 standard defines software quality as the degree to which a software possesses a desired combination of attributes. These attributes can be used to evaluate quality of example programs, and can be mapped to define quantitative measures of quality. In Section 4.2.1 of [5], Börstler et al. classify example quality attributes into three categories of technical, object-oriented and didactical quality. Technical quality deals with technical aspects and considers three quality factors of *implementation versus problem, content and style*. Object-oriented concepts and principles are taken into account under the category of object-oriented quality. While didactical quality considers instructional design, comprehensibility and alignment with learning objectives [4].

A list of example quality attributes is given below. Since example program is a special case of software, so most of these attributes are inspired from general software quality attributes.

1. **Readability:** Readability comes first. It is the most important and basic quality attribute for example programs to follow. Readability means that an example program should be well written and easy to read. If an example program is not even readable, how can one understand its underlying concepts. In beginning, novices tend to understand example program by reading them in smaller parts, similar to reading an ordinary text. Therefore basic syntactical units of an example program must be easy to spot and recognize, so that students can easily establishing a meaningful relationship between these elements [3]. Inappropriate code indentation, bad choice of identifier names, meaningless comments and non standard code conventions are the major problems that make example programs difficult to read. Readability of an example can be improved by using meaningful identifier names, good use of comments, proper code indentation, following standard code conventions and removing unused or noisy code elements. Students' familiarity of the concept, background knowledge and personal interests may few external factors that may affect readability of an example program. One may consider readability as a sub unit of understandability or communication, because of its close relationship to these attributes. However, realizing the importance of readability, we count it as an essential prerequisite and an explicitly separate quality attribute.

Börstler et al. has proposed a readability measure of *Software Readability Ease Score (SRES)* [3] based on the Flesh Reading Ease Score (FRES). SRES counts lexical units such as syllables(lexemes), words (tokens/syntactic categories) and sentences (statements or units of abstraction) to measure the readability of an example program.

2. **Understandability:** A good example must be understandable by students and obviously by computers [3]. Understandability is a cognitive process during which students employ their knowledge, skills and available resources to recognize and understand the elements of an example program. It comes after the basic quality attribute of readability. Readability is both a physical and a mental activity [33], whereas understandability is purely a cognitive process. Example programs having poor readability, meaningless identifiers names e.g. *A, B, x, y*, unnecessarily complex code structure and non standard code conventions make understanding really a difficult task for novice learners.

   Program comprehension is the other name for program understanding in literature. Classical models of program comprehension [24, 37, 7, 6, 39, 38, 31, 32, 2, 26, 43, 14] study the process of understandability for testing, modification and maintenance related tasks mainly from professional programmers' point of view. In future our aim is to study cognitive processes involved in example program comprehension from students' point of view, so that we may formulate effective measures to improve example program comprehension.

3. **Simplicity:** A good example should be as simple as possible, suitably abstract, neither too complex nor too much simple. It should not expose more or less elements, e.g. lines of code, concepts and identifiers, than what is actually required [4]. However readability, understandability and other quality attributes should not be compromised for the sake of simplicity. It should not be misused to write too much abstract example programs without any real life application. "Too abstract" or simple examples is one of the four problems reported by Malan and Halland in [27]. In order to attain simplicity, sometimes example programs are written without any particular application or relevance to any real life objects' interaction. Such examples may help students to learn syntax of the programming language but they do not really help students to learn the principle of object oriented or programming paradigm of choice. Therefore simplicity here does not mean that educators should use too much abstract example, rather simplicity here means that example programs should not be complex above the cognitive capabilities of target students.

4. **Communication:** A good example must have effective explanation of

the key concept, reinforcing earlier taught principles and concepts, without resulting any confusion, ambiguities or misinterpretations [3]. One of the essential properties of good example programs is that they do not let students to end in erroneous conclusions. Rather good example programs should facilitate students to clearly comprehend complex concepts and principles. It might be difficult for novices to learn different uses of inheritance or polymorphism by reading lecture notes or text book. But a well designed example programs can effectively communicate and help students in learning such key concepts and their principles.

5. **Consistency:** Example programs should apply concepts of the particular programming paradigm in a consistent manner. In case of object oriented paradigm, every example should take care of the object oriented design principles. For example, an example program to demonstrate how methods operate on data fields, should implement the operations according to the principles of encapsulation. Malan and Halland, in their article [27], has reported an example program where member data is being passed to member methods of a same class as a formal parameter, against the principles of object oriented programming. Therefore both educators and students should be careful with their choice of example programs to make sure that example programs are consistent with respect to all learning objectives and principles of a particular programming paradigm.

6. **Reflectivity:** A good example should reflect the problem domain in question [4]. In case of object oriented programming, example classes and their members, data fields as well as methods, should reflect real world object's states and behavior. Name of the identifiers should possibly match to the objects from real world problem domain or corresponding concepts in the programming paradigm. It helps to improves quality of an example program by making it easier to read and understand. Students can easily spot and recognize syntactic units to establish a relationship between them, that makes understanding of the concepts easier.

7. **Beauty:** Webster's online dictionary defines beauty as a quality that gives pleasure to the senses. It is simply natural to feel and realize the beauty, but hard to describe it in words. Object's structure, appearance, design and inherent characteristic plays vital role to make it beautiful. Beauty of an example program is that along with satisfying all other quality attributes, it should be attractive and interest provoking. Students should tend to play and learn from example programs without feeling any stress.

Although the above enumerated properties are comprehensive enough, and covers wide spectrum, this is not likely to be a last and a final word. Rather, it is flexible and can be extended to fit a different context or programming paradigm.

4.2 Example Quality Measures

After identifying a list of example quality attributes, our next step is to find appropriate measures. In order to have a quantitative measure of example quality, we have classified four major categories of: structural complexity, cognitive complexity and consistency. Tables 1 provides measurable criteria and corresponding metrics for readability. Table 2 has a list of measurable factors and corresponding metrics to measure the structural complexity of example program. Table 3 deals with measurable criteria and suitable metrics for cognitive complexity. Table 4 shows how we can measure consistency attribute to determine quality of the example programs. Using a combination of these quantitative measures, one can evaluate quality of example program on ordinal and interval scales.

Quality Criteria	Measures
1. Code Style 1.1 Proper Indentation. 1.2 Standard Code Conventions. 1.3 Line numbers (can be omitted by assuming it a part of code editor/IDE). 2. Familiarity of Identifier names, reflecting problem domain. 3. Appropriate use of Comments.	- Software Readability Ease Score [3]. - No. 1 and 3 can be measured using static code analysis approach as do many open source code analyzers like JCSC, PMD, Check Style etc (http://javasource.net/open-source/code-analyzers). - No. 2 can be measured using vocabulary size and reuse factor, vocabulary size measure is defined by Halstead [15].

Table 1: Quality Attributes and Measures of Readability

5 Problems with Programming Examples

Example programs that fail to satisfy above specified basic properties of example quality may result in a number of problems to the students. Malan and Halland [27] had specified four typical problems with example programs in learning. Below list describes mistakes and problems observed with examples programs.

1. **Mishandling “Objects First”**: The phrase “Objects First” means to introduce objects earlier to make them a fundamental concept, rather

Quality Criteria	Measures
1. Control Flow 2. Depth of Inheritance Tree 3. Coupling 4. Cohesion	• No. 1 can be measured using McCabe's Cyclomatic Complexity Number [28]. • No 2, 3 and 4 can be measured using DIT (Depth of Inheritance Tree), CBO (Coupling between Objects) and LCOM (Lack of Cohesion in Methods) metrics defined by [12].

Table 2: Quantitative Measures of Structural Complexity

Quality Criteria	Measures
1. Cognitive Complexity-Chunking Effort [10] 1.1 Chunk Size 1.2 Control Structure 1.3 Expression Complexity 1.4 Recognizability 1.5 Visual Structure 1.6 Functional and Variable Dependencies 1.7 Familiarity 2. Cognitive Complexity-Tracing Effort [10] 2.1 Localization 2.2 Ambiguity 2.3 Spatial Distance 2.4 Cuing Level 2.5 Familiarity 3. Amount of Information 4. Number of New Concepts	- Halstead's [15] measures of Length, Difficulty and Effort can be applied to measure cognitive complexity. - Cognitive Functional Size (CFS) [36] also gives measure of cognitive complexity. - Another choice for measuring cognitive complexity is the metric defined by [22]. - However most comprehensive cognitive complexity model for object oriented paradigm is defined by Cant et al. [10, 9]. - No. 3 can be measured by Kushawa's equation [22]: info= f(identifiers, operators). - No. 4 is bit tricky to measure, but can be handled by setting context for each example, and then analyzing the example contents, if they match with the context or not.

Table 3: Quantitative Measures of Cognitive Complexity

Quality Criteria	Measures
1. With respect to Object Oriented Design Heuristics. 2. With respect to Learning Objectives.	- No. 1 and 2 can be addressed by adopting code smells [41] based approach.

Table 4: Quantitative Measures of Consistency

than following procedural approach starting with basic concepts such as control structures, primitive data types. Using an example code that creates a complete class from scratch, and deals with just one object and method for the sake of simplicity, does not really introduces true picture of the object oriented approach [25]. Instead of creating a complete new class from scratch, just handling single object and method, example templates with spaces to fill the code having multiple methods and interacting objects would be more appropriate to introduce object oriented approach [25]. In this case students are not required to write a complete class from scratch, but they have some partially defined template classes, where they can put some member data and methods to complete them.

2. **Too Abstract/Meaningless Elements**: An example having purposeless elements or meaningless member methods or data fields falls in a category of too abstract bad examples [27]. Such examples have no specific application and tend to make it difficult for learners to understand the underlying concept(s) being explained e.g. simple *Hello World* example. Such examples might be good to explain java syntax, but they don't really teach principles of object oriented design, where each object and its members has well defined purpose.

3. **Too Complex**: In order to make example programs look more realistic and meaningful, sometimes educators end up with too complex examples. This kind of example programs do not match with previous knowledge and cognitive capabilities of the target students and are really harmful in learning, specially when introducing some new concepts to the students. This problem is supported by [27].

4. **Noisy Examples**: Noisy example is a one containing unnecessary elements (e.g. methods, data members, too much comments, unnecessary I/O statements) which distract learner's attention from basic concept being taught.

5. **Inconsistent Examples**: Instead of reinforcing the earlier learned principles and concepts, carelessly designed example programs may overthrow or contradict previously taught concepts, in a deceptive attempt to make new concepts simple and easier to understand [27].

6. **Poor Reflection**: Such example programs fail to reflect data and operations according to the problem domain. Students have to face difficulties in recognizing the concepts as they can not relate examples to the particular problem or underlying concept [4]. Poor reflectivity of example programs may results in misconceptions, ambiguities and erroneous conclusions.

6 Evaluation

Evaluation and validation of the example quality attributes and complexity measures is extremely important. Misuse of example quality attributes and complexity measures can be dangerous by honoring bad examples and casting down potentially good examples [18]. Example quality attributes described in Section 4.1 are selected by a wide and thoughtful study, supported by related literature. However the proposed example quality attribute list is not truly reliable, because it has not been empirically evaluated to prove its validation. Empirical evaluation and validation of the study is planned as a future work. According to the evaluation plan, a candidate pool of java example programs from the course CS1, selected from mainly used text books and lecture notes, will be analyzed with the help of a code quality analyzer tool which implements example quality attributes. Then the results of the code analyzer will be used to evaluate example quality attributes.

7 Discussion

Learning "How to Program" is a two-fold process where student learn language syntax as well as underlying programming methodology such as object-oriented, procedural and non-procedural. Most difficult part in learning object oriented programming is object oriented design, which requires students to truly understand problems and design their object oriented solution. This can be simplified by teaching programming with use of carefully designed example programs. This study mainly focuses example code, while explanatory textual or visual aids are not explored in depth. But it admits importance of the explanation given in text books or lecture notes. An example's description may greatly impact its quality by facilitating students to clearly understand any complex part, acting as a shield against misinterpretation.

Understanding the problems and complexities confronted during program development has remained a topic of great interest for researcher during the last few decades. Earlies attempts, such as those made by McCabe [28] and Halstead [15], primarily focused on static physical structure of the program rather than cognitive issues. Learning "how to program" is a cognitive process, commonly referred as program comprehension. Being a cognitive process, program comprehension varies from person to person, based on the limits of individual's experience, interest and cognitive capabilities. Under such circumstances, it is very difficult to find a standard measure of example complexity, and therefore quality of the example program. However it can be fine tuned by finding certain empirically proved factors related to cognitive process, such as experience factor, interest factor etc.

Apart from "bad" examples, there are several others reasons of student problems in learning object oriented programming [20, 25, 21, 17, 35, 23, 19, 42]. These problems include conceptual myths and misconception about object orientation and ineffective pedagogy. However this study focus only on role of example programs, related issues and problems. The example quality attributes, proposed in this study, have not been experimentally proved to verify their completeness and correctness, therefore, their reliability remains in doubt.

8 Conclusion

This paper advocates the importance of example programs as a valuable learning tool, that unfortunately lack attention from educational community. We make a distinction between "good" and "bad" examples, based on established criteria. One of the key findings is to identify the harmful problems observed with commonly used example programs, which may badly impact student's understanding. Apart from identifying bad examples, we contemplate and propose a comprehensive set of example quality attributes, which can be applied to produce high quality example programs. The proposed example quality attributes are aimed to satisfy learning objectives and principles of object-orientation while maintaining effective communication, reflecting problem domain, improved cognition and comprehension without resulting any misconception or wrong interpretation. Another significant contribution is the measurement of example quality, under which it specifies not only a checklist of the factors but also suggests corresponding metrics. It lacks in experimental validation but this is not due to neglect, rather we aim to empirically evaluate the proposed attributes and problems, in our future work.

9 Acknowledgements

I am really grateful to Johanna Högberg for kind support and pivotal guideline throughout the course. Special thanks to Dipak and Johanna for valuable review and constructive feedback. I am much indebted to Jürgen Börstler for showing confidence and encouraging me to do this work. He helped me a lot to understand the things by sharing his worthy knowledge and experience.

References

[1] VR Basili. Qualitative Software Complexity Models: A summary. *Tutorial on Models and Methods for Software Management and Engineering. IEEE Computer Society Press, Los Alamitos, Calif*, 1980.

[2] VR Basili and HD Mills. Understanding and documenting programs. *IEEE Transactions on Software Engineering*, pages 270–283, 1982.

[3] J. Börstler, M.E. Caspersen, and M. Nordström. Beauty and the Beast–Toward a Measurement Framework for Example Program Quality. 2007.

[4] J. Börstler, H.B. Christensen, J. Bennedsen, M. Nordström, L.K. Westin, J.E. Moström, and M.E. Caspersen. Evaluating OO Example Programs for CS1.

[5] J. Börstler, M. Nordström, L.K. Westin, J.E. Moström, H.B. Christensen, and J. Bennedsen. An Evaluation Instrument for Object-Oriented Example Programs for Novices.

[6] R. Brooks. Using a behavioral theory of program comprehension in software engineering. In *Proceedings of the 3rd international conference on Software engineering*, pages 196–201. IEEE Press Piscataway, NJ, USA, 1978.

[7] R. Brooks. Towards a theory of the comprehension of computer programs. *International journal of man-machine studies*, 18(6):543–554, 1983.

[8] Raymond P.L. Buse and Westley R. Weimer. A Metric for Software Readability. In *ISSTA '08: Proceedings of the 2008 international symposium on Software testing and analysis*, pages 121–130, New York, NY, USA, 2008. ACM.

[9] SN Cant, B. Henderson-Sellers, and D.R. Jeffery. Application of Cognitive Complexity Metrics to Object-Oriented Programs. *JOURNAL OF OBJECT ORIENTED PROGRAMMING*, 7:52–52, 1994.

[10] SN Cant, DR Jeffery, and B. Henderson-Sellers. A Conceptual Model of Cognitive Complexity of Elements of the Programming Process. *Information and Software Technology*, 37(7):351–362, 1995.

[11] M.T. Chi and M. Bassok. Learning from Examples via Self-Explanations. 1988.

[12] SR Chidamber, CF Kemerer, and C. MIT. A metrics suite for object oriented design. *IEEE Transactions on software engineering*, 20(6):476–493, 1994.

[13] R. Flesch. A new Readability Yardstick. *The Journal of applied psychology*, 32(3):221, 1948.

[14] TRG Green. Cognitive approaches to software comprehension: Results, gaps and limitations. *Experimental Psychology in Software Comprehension Studies, Limerick, Ireland*, 1997.

[15] M.H. Halstead. *Elements of Software Science (Operating and programming systems series)*. Elsevier Science Inc. New York, NY, USA, 1977.

[16] S. Henry and D. Kafura. Software Structure Metrics Based on Information Flow. *Transactions on Software Engineering*, pages 510–518, 1981.

[17] Simon Holland, Robert Griffiths, and Mark Woodman. Avoiding object misconceptions. In *SIGCSE '97: Proceedings of the twenty-eighth SIGCSE technical symposium on Computer science education*, pages 131–134, New York, NY, USA, 1997. ACM.

[18] Joseph P. Kearney, Robert L. Sedlmeyer, William B. Thompson, Michael A. Gray, and Michael A. Adler. Software Complexity Measurement. *Commun. ACM*, 29(11):1044–1050, 1986.

[19] Päivi Kinnunen and Lauri Malmi. Why students drop out cs1 course? In *ICER '06: Proceedings of the second international workshop on Computing education research*, pages 97–108, New York, NY, USA, 2006. ACM.

[20] M. Koelling and J. Rosenberg. Guidelines for Teaching Object Orientation with Java. *ACM SIGCSE Bulletin*, 33(3):33–36, 2001.

[21] M. Kölling. The problem of teaching object-oriented programming. *Journal of Object Oriented Programming*, 11(8), 1999.

[22] D.S. Kushwaha and AK Misra. Improved Cognitive Information Complexity Measure: A Metric that Establishes Program Comprehension Effort. *ACM SIGSOFT Software Engineering Notes*, 31(5):1–7, 2006.

[23] Essi Lahtinen, Kirsti Ala-Mutka, and Hannu-Matti Järvinen. A study of the difficulties of novice programmers. In *ITiCSE '05: Proceedings of the 10th annual SIGCSE conference on Innovation and technology in computer science education*, pages 14–18, New York, NY, USA, 2005. ACM.

[24] S. Letovsky. Cognitive processes in program comprehension. In *Empirical Studies of Programmers*, pages 58–79. Intellect Books, 1986.

[25] John Lewis. Myths about object-orientation and its pedagogy. In *SIGCSE '00: Proceedings of the thirty-first SIGCSE technical symposium on Computer science education*, pages 245–249, New York, NY, USA, 2000. ACM.

[26] D.C. Littman, J. Pinto, S. Letovsky, and E. Soloway. Mental models and software maintenance. In *Papers presented at the first workshop on empirical studies of programmers on Empirical studies of programmers table of contents*, pages 80–98. Ablex Publishing Corp. Norwood, NJ, USA, 1986.

[27] Katherine Malan and Ken Halland. Examples that can do Harm in Learning Programming. In *OOPSLA '04: Companion to the 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications*, pages 83–87, New York, NY, USA, 2004. ACM.

[28] T.J. McCabe. A Complexity Measure. *IEEE Transactions on Software Engineering*, 2(4):308–320, 1976.

[29] S. Misra. An Object Oriented Complexity Metric Based on Cognitive Weights. In *6th IEEE International Conference on Cognitive Informatics*, pages 134–139, 2007.

[30] M. Nordström. He[d]uristics - Heuristics for Designing Object Oriented Examples for Novices. Technical report, March 2009.

[31] N. Pennington. Stimulus Structures and Mental Representations in Expert Comprehension of Computer Programs. 1985.

[32] N. Pennington. Comprehension strategies in programming. In *Empirical Studies of Programmers: Second Workshop*, pages 100–113, 1987.

[33] D.R. Raymond. Reading source code. In *Proceedings of the 1991 conference of the Centre for Advanced Studies on Collaborative research*, pages 3–16. IBM Press, 1991.

[34] S. Rugaber. Program comprehension. *Encyclopedia of Computer Science and Technology*, 35(20):341–368, 1995.

[35] Kate Sanders and Lynda Thomas. Checklists for grading object-oriented cs1 programs: concepts and misconceptions. *SIGCSE Bull.*, 39(3):166–170, 2007.

[36] J. Shao and Y. Wang. A new measure of software complexity based on cognitive weights. In *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, volume 2, 2003.

[37] B. Shneiderman and R. Mayer. Syntactic/semantic interactions in programmer behavior: A model and experimental results. *International Journal of Parallel Programming*, 8(3):219–238, 1979.

[38] E. Soloway, B. Adelson, and K. Ehrlich. Knowledge and processes in the comprehension of computer programs. *The nature of expertise*, pages 129–152, 1988.

[39] E. Soloway and K. Ehrlich. Empirical studies of programming knowledge. *IEEE Transactions on Software Engineering*, 10(5):595–609, 1984.

[40] CS2008 Interim Review TaskForce. Computer Science Curriculum 2008: An Interim Revision of CS 2001. Technical report, December 2008.

[41] E. van Emden and L. Moonen. Java quality assurance by detecting code smells. In *Reverse Engineering, 2002. Proceedings. Ninth Working Conference on*, pages 97–106, 2002.

[42] Phil Ventura and Bina Ramamurthy. Wanted: Cs1 students. no experience required. In *SIGCSE '04: Proceedings of the 35th SIGCSE technical symposium on Computer science education*, pages 240–244, New York, NY, USA, 2004. ACM.

[43] A. Von Mayrhauser and A.M. Vans. Program understanding: Models and experiments. *Advances in Computers*, 40:2–2, 1995.

The role of auditory feedback in mobile devices

Lars Ahlin

Abstract: Advances in the mobile device field has led to more powerful devices and as a result of that, more complex and demanding applications. The rapid progression results in smaller and smaller devices. The lack of screen estate quickly becomes a limitation for a designer. From an application designers point of view this limitation demands a lot of attention in order to create good usability of the application. This paper examines one way to handle this limitation, auditory feedback. It regards in which way sound can be implemented to enhance the usability and what effect sound can have on the end result. The paper also regards some issues connected with auditory feedback.

1 Introduction

With the technical progress striving for smaller and more computetionally powerful things the mobile device industry is no exception. Today, mobile phones, PDAs and similar hand held mobile devices have the capability to perform almost the same tasks as an ordinary PC. However, even if it is possible to do so, there are issues regarding the use of a mobile device. The size of the mobile display is limited in order to make the device possible to carry around. This limitation demands a lot from an interface design´s point of view. With a restricted display, minimizing the amount of visual impressions is a must. In order to deal with these limitations the amount of visual information has either to be stripped down, removed or offloaded into none visual alternatives. One way of doing this is to sonically enhance the application. By offloading information into audio the user will be provided with auditory feedback instead of visual.

This paper will regard this type of auditory feedback but it will also encounter more simplistic auditory feedback such as sound alert and sound menu navigation. The study can be seen as an investigation of designing sonically enhanced applications and it will discuss what role auditory feedback has in mobile devices and in which ways it can enhance the usability of such a device. It will also regard different issues directly connected to this kind of feedback such as environmental context awareness and disturbance.

The paper starts with a background study of usability, sound techniques and interfaces in section two. In section three an overview of related work is given to provide information about what has been done within the paper´s topic. In the fourth section the procedure of how this paper was conduced is given. The following three sections, five, six and seven contains the result of the study, a discussion regarding the results and some final conclusions.

2 Background

In order to grasp the concept of auditory feedback in mobile devices a couple of things has to be known. Usability is a term with many associations and it requires an declaration from the mobile device development point of view. This section also discusses the usage of earcons and spercons and also the cornerstones of designing mobile device interfaces.

2.1 Usability in mobile devices

In the beginning of 1980 the term usability was coined. The term was set to replace the old expression "user friendly" because of its vague meaning. However, usability is also not unambiguously defined which has resulted in many different approaches to what is meant by it. The three main approaches of how to measure usability is the product-oriented view, the user-oriented view and the user performance view [1].

When talking about usability of handheld devices the approach mentioned last, the user performance view, is the approach used in the articles resulting in this paper and is therefor the approach of choise for this paper. The user performance view states that usability is measured by observing how the user interact with the device, if the interaction is satisfying and if the result from the interaction is satisfying [1]. When focusing on the role of auditory feedback in mobile devices the measuring of usability is a way to describe how the user interacts with the mobile device and what this interaction results in, with audio feedback as well as without.

2.2 Earcons and spearcons

An earcon is the word for non-speech audio used in applications and interface to indicate something through sound. Earcons are synthetic tones that can be used to create auditory messages [3]. Earcons for this purpose have always been used but not always for the same purpose as they are used today. In traditional application design the earcons were used almost exclusively to indicate errors or confirming pressed buttons. The rapid evolution of interfaces have evolved the eracons to contribute with more than just alerts and confirmations sounds [4].

Another technology are spearcons. They are created by speeding up spoken words or phrases. The spearcon is speeded up to the extent that it no longer is categorized as speech. The spearcon is often used as feedback in auditory menus where it outshine earcons by far [13]. The difference in learning rate between the two in such a auditory menu is shown Figure 1 below, Adopted from Dianne K. Palladino and Bruce N. Walker [9]. The two technologies were tested for two types of menus, list 1 and list 2. The bars are showing the mean number of tests that had to be conducted in order to get a perfect score navigating through the menus. As can be seen in the figure the difference between spearcons and earcons are large. The tests with earcons needed between four and seven times as many trys as the tests with spearcons to reach perfect score.

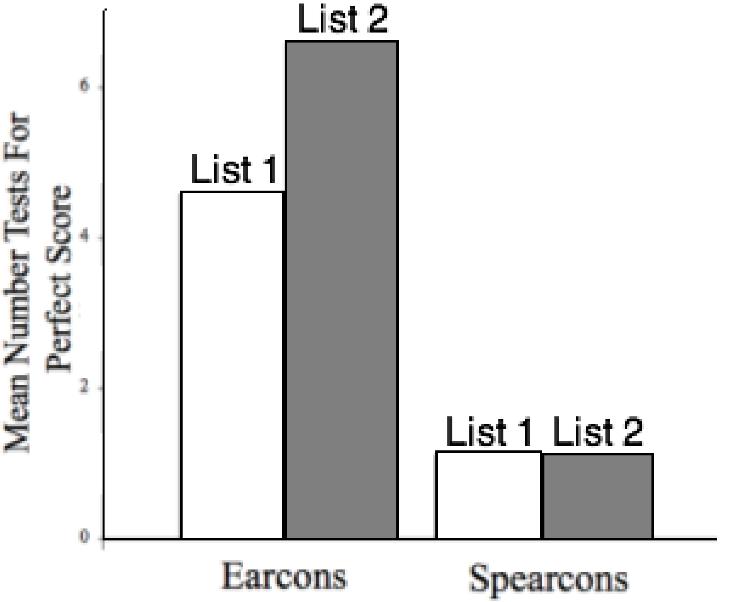

Figure 1: Bar chart showing the mean differens in performance between spearcons and earcons in a auditory menu [9].

Along with the advancements and the increasing complexity among interfaces, the usage of earcons and spearcons has evolved. Nowadays earcons can be implemented as information providers and can in that sense reduce the amount of graphical information by offloading it into auditory information [2].

2.3 Design of mobil device interfaces

To make sure that a user conveniently can interact with a device a interface are implemented. It is through the interface the user feeds input to the device and in reverse it is through the interface as the device provide feedback as output to the user. Different devices have different interfaces and when talking about mobile devices such as cell phones and personal digital assistants (PDA) the design of the interface is quite demanding.

When talking about interface design in general there is one person that stands out in particular, Ben Shneiderman. Shneiderman´s eight golden rules has been adopted almost as a standard for this type of design [11]. The eight rules that Shneiderman stated are as follow.

- **Strive for consistency.** Everything within a interface should be consistent. Consistent commands and consistent terminology should be employed throughout the whole interface and action sequences should be similar or the same for all parts of the interface.

- **Enable frequent users to use shortcuts.** Users that frequently operate a system tend to strive for ways to dismiss unnecessary steps and information. The design should therefor be appointed macros, hotkeys and hidden commands in order to for fill the users needs.

- **Offer informative feedback.** With every operator action the system should provide feedback to the user. This should not only be implemented for errors but for all operator actions.

- **Design dialog to yield closure.** When a operation is completed some kind of dialog should make the user aware of the completion. This is of importance for not creating confusion to users waiting for something to happen.

- **Offer simple error handling.** A system should be design so that user cannot harm the system with serious errors. If a error does occur, the system should detect the error and offer information and error handling.

- **Permit easy reversal of actions.** Implemented reversibility is strongly recommended. It allow users to undo changes or revert to a previously saved version and is a key feature if the system requires a lot of input.

- **Support internal locus of control.** To satisfying this rule the user has to be in control of the system and not vice versa. In other words, the user has to be in charge of the system.

- **Reduce short-term memory load.** Because of the limitations of human memory the system should be implemented in a way that reduces the amount of information a user has to remember when operating the system.

In order to apply these on mobile device interfaces some modifications has to be done. Some of the topics might need a bit of redefinition but as a entirety Shneidermans guidelines are fit for mobile device interfaces too. A mobile device obviously has a size limitation which automatically effects the interface.

According to Jun Gong and Peter Tarasewich et al. [5] the eight rules is not enough so the following topics has to be taken into consideration in order to grasp all aspects of a mobile device interface.

- **Design for multiple and dynamic contexts.** Environmental conditions are inevitable and should not affect the usability of the device.
- **Design for Small Devices.** Devices as small as rings or key chains with no room for buttons might need to take speech as input and sound as output.
- **Design for Limited and Split Attention.** The design should strive to minimize the attention required to use the device.
- **Design for speed and recovery.** It should be easy and fast for a user to access applications and recover previous settings and work.
- **Design for "Top-Down" Interaction.** Hierarchical mechanisms or multilevling sholud be implemented if the information flow is to large in order to let the user decide whats relevant to take part off.
- **Allow for personalization.** The application has to be able to be personalize because different users uses different settings.
- **Design for Enjoyment.** To invoke a positive feeling and an enjoyable experience it is important to take aesthetics in to consideration. [5]

3 Related work

There have been a lot of studies conducted to test the performance of applications with the involvement of auditory feedback [14, 3, 5]. The studies have almost exclusively been focused on the user´s results and testing the hypothesis that users perform quicker and more effective with the involvement of earcons or spearcons. One thing that almost all of the studies had in common was that they tended to focus only at one particular application, not making any afford to grasp the whole phenomena of auditory feedback. Various guidelines on how to design parts of sonically enhanced interface has been made but small efforts has been made to construct guidelines for sonically enhanced interfaces as a unit. An exception is the US government that stated in a report that all visual menus should have a non-visual alternative and with that report provided some guidelines for this purpose [6].

Pavani and Walker et al. [13] who designed guidelines for advanced auditory menus state that a lot of attention needs to be pointed towards choosing suitable spearcons and earcons, emphasizing that the sound is crucial for a satisfied usability. Sound that does not match its purpose will have a reverse effect on the usability and make the system harder to use [8]. Lepltre and Brewster et al. [8] have provided a framework for earcon design and studies on this framework has proven that users made use of less key presses to complete the given tasks. The study also showed that the users were able to complete more task successfully with usage of sound.

4 Procedure

The study was conducted through a literature review of the importance of sound in interface design. The literature review was based on articles and books regarding different usability test on various applications as well as on articles regarding existing guidelines of interface design. When choosing articles it was important that the focus of the article was on the sonically enhancement of the application and not just a study of a application containing audio. It was also of importance that the subject of the article was relevant for todays technology. To encourage diversity the articles were chosen from a wide selection of writers but the extensive work of Stephen E. Brewster has not been ignored, resulting in a more references to his work than to any other writers work.

5 Results

When it comes to applications with auditory feedback there are a wide range of implementations providing information to the user by sound. A common implementation that almost everyone has come across is button sound. When dialing a phone number a sound indicates that a button was pressed and by that provides feedback to the user about a pressed button. Another more advanced area of auditory feedback is sound graphs. The purpose of such a graph is to give the user information by sound instead of otherwise presenting the information in graphics. This is done in order to enable the user to focus his attention on other things at the same time.

5.1 Usability effects

While it can seem simple and unnecessary with earcons indicating if a button is pressed, Brewster et al. [3] show that the effect sound have on the performance in that case is significant. With the presence of sound the study showed an improvement in usability. The amount of data the user entered was increasing by the decreasing of workload for the user. This is consistent with the "Reduce short-term memory" rule under the Design of mobile device interfaces chapter. As shown in figure 2, which is adopted from Brewsters study [3], the performance is almost the same for ordinary buttons with no earcons compared to buttons reduced to half the size with earcons. The bar chart clearly states that the buttons enhanced with sound has a significant improvement of usability compared to the ordinary buttons. The fact that size can be reduced is in line with the "Design for small devices" rule and the improvement in usability is proven by the "Offer informative feedback" rule.

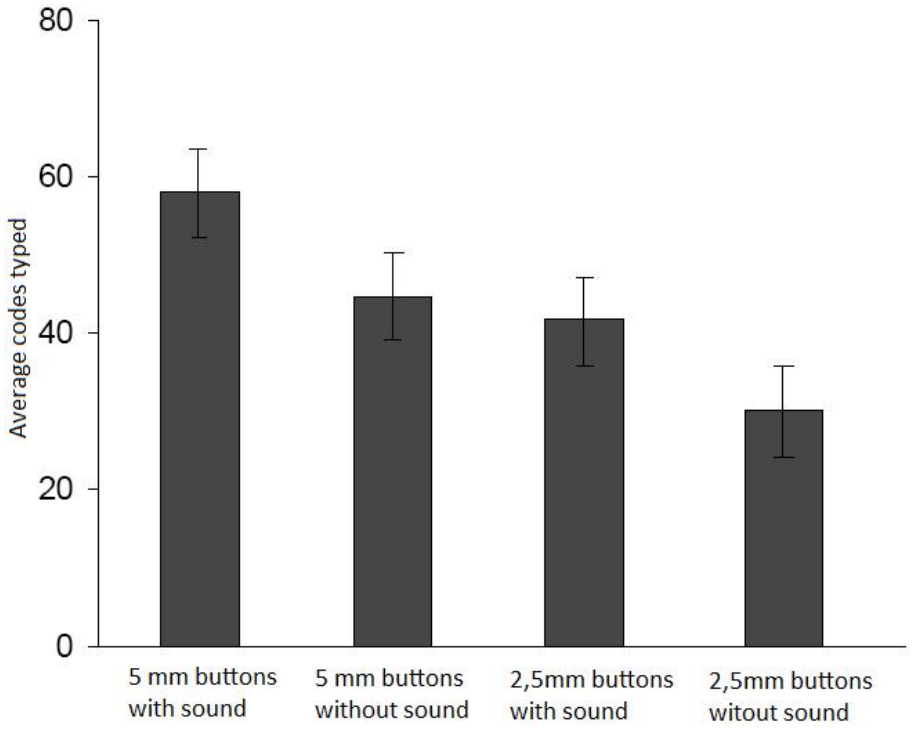

Figure 2: Bar chart summarizing Brewsters study of overcoming the lack of screen space on mobile computers [3].

To present dynamic data to a user sound graphs are documented to be a suitable substitute for a visual graph [2]. In the study, Brewster and Murray et al. [2] used a stock trading application to examine the usability effect of such a graph with positive results. The usage of sound graphs was proven to drastically enhance the usability compared to the visual choice. As a part of that, sound graphs were able to reduce the workload invested by the user. The different design rules supporting this result are the "Reduce short-term memory load" rule, because of the reduce in work load, and the Òdesign for limited and split attentionÓ rule due to the fact that the user do not need to focus any perceptive attention towards a visual graph.

Another way to enhance the usability is to provide navigational feedback in order to help the user navigate the device. This setting has often been implemented to navigate through menus. Studies has proven a more accurate performance with an audio enhancement than without [14]. Another aspect to navigating a menu with the help of sound is the aspect of people with visual impairment, for them the auditory menu is essential for their ability to use a system. In their case auditory notification in every part of the interaction is needed. Without feedback, users can lack a bit of confidence using the device [10].

Auditory navigation is not only used for menu navigation. Another implementation field that was a subject for this article is physical navigation. To support usability in this case the device has to support navigation and awareness of the correct route. An application that is implemented for this purpose is the mobile route guide [12]. It shows that for the majority of the users the usability of finding the correct route is enhanced with audio feedback. The essential design rule for this application is the "Design for limited and split attention" rule. To be able to navigate in the physical environment the attention has to be divided among the device and the actual route.

5.2 Issues

Designing sonically enhanced applications attracts some issues. A big part of the usability aspect is that the device should be free from annoyance and disturbing behavior. For some users sound can provide an annoying experience, an example stating this is provided by Kurdyukova, Hahnen, Prinz and Wirsam, where a mobile outdoor training assistant is evaluated at a group of people. While most of the test subjects where fond of the audio guidance system, a few athletes experienced a hard time trying to coup with the sound guide. [7]

An even bigger issue is context awareness. This is an important aspect to the subject discussed in this article because of the amount of different contexts a mobile device can be found in. Different contexts provide different conditions and there are by far not all contexts that are suited for the presence of sound. The issue with context awareness is therefore to provide information about the surroundings to the device. This should be done in order for the device to adept to the required settings without losing any functionality.

Finally interference between applications can be classified as an issue. If several applications on the same device are running at the same time then sound may not be to the users advantage. Different sounds interfering with each other may certainly create confusion instead of the purpose they were created for.

6 Discussion

So what role does auditory feedback play in mobile devices? As stated throughout the previous sections, auditory feedback enhances the usability in many different ways. Not only to visually impaired people who are forced to rely on it, but to all kind of users. When the screen estate is limited studies as Brewsters et al. [3] show that simplistic features can make a difference. By halve the button size, audio can provide a support for the application to still remain almost the same usability and create a lot more room for other information. To receive guidance and information by sound can not only be seen as a usability enhancement but also as a safety provider. In situations when its crucial that attention is directed on other things, like driving a car, earcons make it possible to navigate or receive information without shifting attention from the road to the device, jeopardizing the users safety.

Because of the fact that perception space of audio is much smaller than the perception space of vision the different types of implementations that has been regarded in this article is in my opinion representing the boundaries of what can done with todays technology.

After reflecting strictly on the effect audio feedback has on the usability it is important to consider the issues it attracts. As the results speaks for themselves when considering the improvements of usability, the studies examined in this paper does very little to reflect on context awareness. A lot of different contexts are not suitable for sound; as a matter of fact the authors seems to lack a bit of interest in this. With a few exceptions almost none of the studies reflect over where the device would be used. In general the impression is given that it is up to the user to avoid usage if the device not suits the context. Many of the applications studied in this report rely entirely on sound and are useless if the situation demands for silence.

To avoid interference among different applications there has to be a hierarchical structure providing information about and sorting out intergroup rankings. A more important system in such a structure has higher priority than a less important system. Systems with low ranks should basically mute when a more important system gets activated. In the end it is my opinion that the consumer of the device is the one deciding these ranks. After all, as the "Allow for personalization rule" state, the application has to be able to be personalized because different users use different settings.

7 Conclution

Auditory feedback should always be implemented as a support for the user and by that I mean it should always be an option. No device should depend on the fact that the user is present in a context where it is possible to use sound. A good implementation can provide auditory feedback as a positive enhancement of the usability of the device. As the devices tend to strive for as small as possible the demand for non visual ways to interact with it will increase. For people entirely relying on auditory feedback there is a tradeoff between using headphone and not using the applications. Auditory feedback has a lot to offer and if all aspects that are discussed in this paper are taken into consideration it will have an imported role to play in the usability of mobile devices.

References

[1] Nigel Bevana, Jurek Kirakowskib, and Jonathan Maissela. What is usability? *Proceedings of the 4th International Conference on HCI*, 1991.

[2] S. A. Brewster and R. Murray. Presenting dynamic information in mobile computers. In *Personal Technologies*, pages 209 – 212, 2000.

[3] Stephen Brewster. Overcoming the lack of screen space on mobile computers. In *Personal and Ubiquitous Computing*, pages 188–205, 2002.

[4] Stephen A. Brewster, Peter C. Wright, and Alistair D. N. Edwards. A detailed investigation into the effectiveness of earcons. In *Auditory display, sonification, audification and auditory interfaces. The Proceedings of the First International Conference on Auditory Display*, pages 471–498, 1992.

[5] Jun Gong and Peter Tarasewich. Guidelines for handheld mobile device interface design. *In Proceedings of the 2004 DSI Annual Meeting*, 2004.

[6] USA government. Us goverment homepage, 2009. `http://www.access-board.gov/telecomm/rule.htm`, accessed 2009-05-10.

[7] Ekaterina Kurdyukova, Jochen Hahnen, Wolfgang Prinz, and Wido Wirsam. Evaluation and design of auditory feedback for a mobile outdoor training assistant. *InÊProc.ÊMenschÊandÊComputerÊ2008*, 2008.

[8] Gregory Lepltre and Stephen A. Brewster. Designing non-speech sounds to support navigation in mobile phone menus. *ICAD2000*, 1998.

[9] Dianne K. Palladino and Bruce N. Walker. Learning rates for auditory menus enhanced with spearcons versus earcons. In *ICAD 2007*, pages 274–279, 2007.

[10] P. Pirhonen, S. A. Brewster, and C. Holguin. Gestural and audio metaphors as a means of control for mobile devices. In *ACM-CHI'2002*, pages 292–298, 2002,.

[11] B Shneiderman. Designing the user interface - strategies for effective human-computer interaction,. In *Addison-Wesley*, 1998.

[12] Markku Turunen, Jaakko Hakulinen, Anssi Kainulainen, Aleksi Melto, and Topi Hurtig. Design of a rich multimodal interface for mobile spoken route guidance. *INTERSPEECH*, 2007.

[13] Bruce N. Walker, Amanda Nance, and Jeffrey Lindsay. Spearcons: Speech-based earcons improve navigation performance in auditory menus. *Proceedings of the International Conference on Auditory Displayd*, 2006.

[14] Pavani Yalla and Bruce N. Walker. Advanced auditory menus. In *GVU Center Technical Report*, pages 1–6, 2007.

Context Awareness by Sensor Fusion from Sensors in Clothing

Ola Andersson

Abstract: Through context awareness computer systems can respond to changes in the environment. By integrating sensors in clothing and combining the data from them by sensor fusion, the clothing can become aware of its context. In this paper, context aware clothing is presented with the use of three concepts: context awareness, sensors, and sensor fusion. The three concepts together with context aware clothing are explained and a set of recomendations for context aware clothing is presented.

1 Introduction

The use of context awareness in clothing serves several purposes. Clothing is a useful medium for computing since it is something that people wear most of the time. Clothing with sensors combined with computing power can be integrated into the daily life without hindering the user [1]. This paper is based on four concepts: context awareness, sensors, sensor fusion and context aware clothing.

The use of *context awareness* in computer systems can improve the users interaction with the system. A system that is aware of its context can respond to it and make choices that are appropriate for the user. If this is extended to clothing, clothes can be aware of how they are used and help the user in everyday activities. Two examples of studies involving clothing combined with sensors can by found in the paper by Lehn et al. [2] and the paper by Edmison et al. [3].

If a system wants to receive information from the environment, it needs *sensors*, or access to the data that have been collected by another system. There exists many different types of data that can be collected by sensors, but also the quality of the individual sensors contribute to how the system perceive the world. There exist several techniques for processing the data, this paper focus on one of them, sensor fusion. *Sensor fusion* is a technique where the data from many sensors are combined and fused together. The data can be from different types of sensors or from many similar, but the idea is to provide a better picture by combining the data than the individual sensors can give on their own. There exists some related work that deals with sensor fusion and context awareness. Van Laerhoven et al. describe a study with a sensor system distributed over the human body [4]. In another paper by Van Laerhoven at al. a study with several types of distributed sensors and sensor fusion is described [5]. The work by Gellersen et al. describes sensor fusion and a study in context awareness [6].

The main question this paper tries to answer is *how a set of distributed sensors can provide context awareness in clothing with sensor fusion.* Each of the four concepts: context awareness, sensors, sensor fusion and context aware clothing is described in the following sections. The first three serve as a background for the fourth concept: *context aware clothing.* The last part of the paper is a set of recomendations that emphasize important points when designing context aware clothing.

2 Context Awareness

Context is described by Gellersen et al. as what surrounds, for example the physical world that surrounds a mobile device [6]. According to Gellersen et al. context can be used to compensate for the abstraction that is required to make systems accessible in changing environments. They also claim that context can be used for different tasks within a mobile device, like power management [6].

There exist many definitions of *context awareness* but one definition that have been refered to by many other authors was presented by Dey et al. as: "A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task." [7]. Another definition of context awareness can be found in a paper by Van Laerhoven [8]. Van Laerhoven writes that many applications classify the information that is recieved from sensors into concepts [8]. This information is then used to get a description of the context that the user is in. Context awareness is defined as: ”complex mapping of sensor data to high-level concepts has notably in the field of human-computer interaction frequently been marked as *context awareness*” [8]. The definition proposed by Dey et al. is more general while the definition from Van Laerhoven gives an idea on how to construct some types of context aware systems since it is tied closer to the hardware of a context aware system.

Context awareness and *context aware systems* are according to Bellotti and Edwards [9] already a part of our environment. Context awareness can make devices respond to changes in the users actions or the enviroment and alter their settings to these changes. Examples of devices or systems that are aware of their context are automatic doors, alarm systems and auto pilots in airplanes [9].

The most common way to identify a physical context is to use sensors together with software algorithms. The human factors are not as easy as the physical context to identify. Bellotti and Edwards claim that unlike computer systems, humans make unpredictable judgements and improvise to get the task done [9]. This is according to Bellotti and Edwards because humans are influenced by for example emotions, dislikes, or phobias and this makes human behaviour very hard to predict for a context aware application [9]. Bellotti and Edwards do suggests a solution to this problem: the system should present alternatives to the user so they can decide how to interpret the data instead of the system [9].

2.1 Recognising Context

Many applications do not use sensor data directly. According to Van Laerhoven sensor data gets classified into concepts that are easier to work with and are more useful [8]. This mapping can be both easy and hard to do depending on the data the system has to work with. While data from a thermometer is one dimensional and can be placed on a line ranging from hot to cold, visual data used in image processing are much harder to map since it can be mapped into several concepts. Examples of these concepts are objects, background, or colour.

One example of how this mapping can be done is presented in the TEA project by Gellersen et al. [6]. TEA is an architecture for context awareness in personal mobile devices. The architecture is built upon three layers, the sensor layer, the cue layer and the context layer. The sensor layer is based on an open ended collection of sensors that can be of different types. The cue layer reduces the volume of the sensor data, independent of the target application. The reduction can be done by calculating the variance of the sensor data, ignoring data under a certain threshold to reduce noise, or by using methods from statistics, calculation of frequencies or patterns of movement. The actual contexts do not appear until in the context layer which combines the cues by the use of several methods, including using rule-based algorithms, statistics or neural networks.

According to Van Laerhoven another way of doing these mappings is to have an internal world model that is trained to recognize contexts when presented with examples of concepts [8]. When the internal world model has learned to recognize a context, it can recognize other similar contexts that is presented to it. Van Laerhoven also mentions that by the use of incremental learning, the world model can be even more flexible and be taught new contexts over time [8]. Incremental learning is described by Russell and Norvig as an approach where "one never has to go back and reexamine the old examples" [10, page 683]. They mention that this is done by keeping a space of possible hypotheses and when presented with new data, hypotheses that does not fit the data are discarded. In the case of contexts, the system can have a number of possible contexts it can be in, and when presented with more data it can discard those that does not fit the concepts formed from the data.

3 Sensors

Van Laerhoven provides three views of what a sensor is; the *first* is that a sensor is a device that is capable of detecting and responding to physical stimuli [8]. This definition is not very specific and everything from a piece of paper to a human being might be seen as a sensor. The *second* view is that a sensor is useless unless it is used in the right environment and observed by someone or something that can respond to it [8]. The *third* view is that a sensor can capture information in the real world and transfer this to the virtual or digital world [8].

The two following subsections describe different sensors divided into two categories and two ways to connect these sensors in networks.

3.1 Direct and Indirect Sensors

One way of dividing sensors into two groups is *direct sensors* and *indirect sensors*. Direct sensors are those that do not rely on external sources for information, and collect all the data directly. Van Laerhoven et al. describes two sensors that belong to this type : accelerometers and ball switches [5]. An accelerometer measures the acceleration an object is exposed to while the ball switch is a binary sensor because it only senses if the sensor is tilted over a certain threshold. The limited output of the binary sensor can be overcome by combining several sensors placed at different angles. According to Van Laerhoven an accelerometer is more accurate than the combined ball switches but requires more resources [5]. Other types of direct sensors are described by Pheifer; these include cameras, infrared and ultrasonic sensors [11]. Indirect sensors can not collect data directly and recieves it from other sensors or systems. Some of these technologies are described in a paper by Pheifer and include RFID, GPS, GSM sub-cell positioning and wireless networks [11]. Many of these technologies require that there is a supporting infrastructure of receivers, satellites or other technologies available.

The notion of direct and indirect sensing can be extended to whole systems. According to Gellersen et al. context aware systems can for example be divided into systems that have *direct awareness* and systems that have *indirect awareness* [6]. With direct awareness the system has one or more sensors, together with models or algorithms for computation of more abstract contexts from the sensor data. A system with indirect awareness depends on the infrastructure for all sensing and processing, and gets the information about its context from communicating with the environment. According to Gellersen et al. recent advances in sensor technology have lead to more direct awareness for mobile devices [6].

3.2 Connecting Sensors

In a system consisting of many sensors, they are usually connected in some way. Van Laerhoven et al. claims that this can be done for example by a network [4]. If the sensors are connected, the data captured by each sensor can be transferred through the network and combined in a form that is useful for the application. Sensors can be connected in many ways, two of these are *wired* and *wireless*. A wired approach is more static since it depends on physical wires that needs to be connected to each sensor. A wireless approach on the other hand is more dynamic but according to Lo and Yang the transceivers for the wireless network requires a lot of power [12]. If a wireless approach is used then the individual sensors require some sort of power supply. Two more recent techniques can be seen in an article by Hanson et al., these are magnetic induction and body-coupled communication [13]. Magnetic induction is a wireless technique that uses magnetic induction from coils of wire. It can be used both for implanted and swallowed sensors. Body coupled communication uses the body as a medium for communication. According to Hanson et al. it has low energy requirements, provide for stable channels and is hard to detect but further research has to determine how safe it is to the users [13].

There exist several ways to connect and manage data in sensor networks. Two ways mentioned by Van Laerhoven et al. are the *centralised approach* and the *distributed approach* [4]. Both approaches along with their advantages and disadvantages are presented below. The centralised approach can be seen as a tree-like structure where data is collected in the leafs at the lowest level of the tree and then propagated up in the hierarchy to the root. The leafs of the tree are sensors, the nodes can be micro-controllers and the root can be a micro-processor. The benefits are that the centralised approach is much easier to implement than the distributed approach and that most systems are based on one computer or one processor, which makes it a good choice. The centralised approach can be seen in Fig. 1. The distributed approach is more robust than the centralised approach since both sensors and micro-controllers can stop working without taking the whole system off-line. It is possible to add, move or remove components in the network without having to reconfigure it manually. The distributed approach also focuses on an emergent self-organisation instead of collecting all the data at a central place in the network. The distributed approach can be seen in Fig. 2.

An article by Hanson et al. mentions two types of network topologies: *star topologies* and *star-mesh hybrid topologies* [13]. The star topology can be compared to the centralised approach proposed by Van Laerhoven et al. [4] as it is based around a central node to which all other nodes connect [13]. The star-mesh hybrid topology is similar to the distributed approach proposed by Van Laerhoven et al. [4] due to that it is not based on a single central node. According to Hanson et al. it is based around a mesh of several central nodes and networks that are connected together [13]. That way, if one central node fails, the network reorganises itself around the remaining nodes and continue to be operational.

A summary to the subject of connecting sensors is that there exist two types of networks: wired and wireless networks. These networks are then connected in a centralised or distributed fashion. There exists at least two network topologies that can be used: the star topology and the star-mesh hybrid topology.

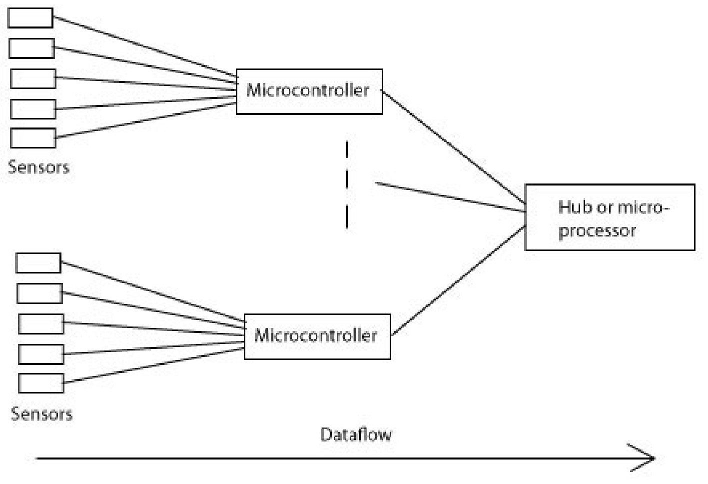

Figure 1: The centralised approach to connect sensors in a network (adapted from Van Laerhoven et al. [4]).

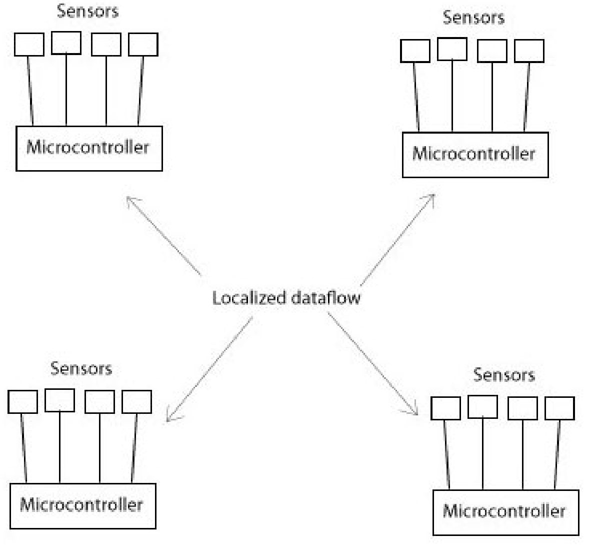

Figure 2: The distributed approach to connect sensors in a network (adapted from Van Laerhoven et al. [4]).

4 Sensor Fusion

Many simple sensors can often be used instead of a single advanced sensor. The data collected from these sensors are then combined. There exists many techniques for combining the data and several parameters that affect the quality of the combined data. Both the theory behind sensor fusion, methods and the parameters that affect the quality are explained in the following two subsections.

Hightower and Borrellio give a definition of sensor fusion as "the use of multiple technologies or location systems simultaneously to form hierarchical and overlapping levels of sensing" [14]. Hightower and Borrellio also writes that this can provide properties that would be unavailible if the systems would be used individually, which may increase accuracy and presision [14]. According to them this is done by integrating several systems with different error distributions and the effectiveness is higher when the techniques are more independent of each other [14].

4.1 Sensor Fusion Theory

The basic idea of sensor fusion is to take several simple sensors and combine the data from them. This is used both in biological systems according to Brooks [15] and in robotics according to Murphy [16, pages 197-200]. According to Van Laerhoven a more traditional view is to have a single or a few sensors at specific places and combine data from them with specialised algorithms [8]. According to Van Laerhoven and Gellersen et al. the data combined from many sensors gives a much more detailed picture than any of the individual sensors can give on their own and might give a better picture than what can be gained from a single more advanced sensor [8, 6]. The difference between sensor fusion and the traditional approach can be seen in Fig. 3.

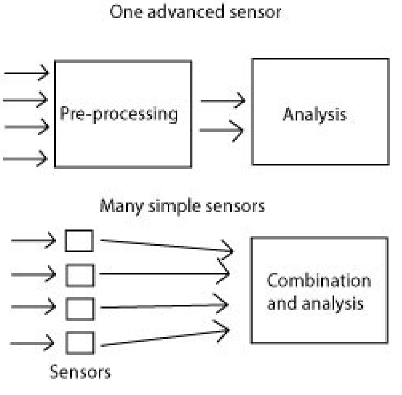

Figure 3: The difference between using a single advanced sensor and many simple sensors (adapted from Van Laerhoven et al. [17]).

The sensors can be spread out over a large area, in the case of a human body they can be spread over the arms, legs or torso. The sensors can then be connected in a network, of either a centralised or a distributed type. The collected data are then combined and analysed using various techniques. In studies done by Gellersen and Van Laerhoven and Gellersen [6, 5] these techniques included rule-based algorithms, statistics and neural networks. There also exist other techniques that are described in various textbooks on sensor and data fusion, like the textbook by Klein [18], but the scope of these books are too large to cover in this paper. The techniques Klein describe can be summarised to: statistics, bayesian networks, Dempster-Shafer theory, neural networks and fuzzy logic [18].

Van Laerhoven describes both the benefits and drawbacks with sensor fusion [8]. According to Van Laerhoven the *benefits* are that the individual sensors are cheap, many sensors make the system robust, the system is distributed and flexible [8]. A simple sensor is usually much cheaper than a complicated one and uses less resources. There is a limit to this though, if the number of sensors increases, the cost and power consumption will eventually be higher than for a single complex sensor. Van Laerhoven claims that since there are multiple sensors that are distributed over a larger area, the resulting system is more robust than a system based on a single sensor [8]. If one sensor is blocked or broken Van Laerhoven claim that there are others that can capture similar information and the system can still produce a good result [8]. According to Van Laerhoven the system is also flexible since more sensors can be added, or the existing sensors can be moved to new places [8]. Van Laerhoven claims the *drawbacks* are that the data from the sensors needs to be combined and analysed in an efficient way and that the software algorithms used are the bottlenecks of the system [8].

According to an earlier study by Van Laerhoven et al. [17] that refers to a book by Mitchell [19] there exists at least two problems with the software for sensor fusion, one is related to machine learning and the other to the number of sensors used. The problem related to machine learning is called *stability-plasticity* or catastrophic forgetting. Van Laerhoven et al. describe it as a trade-off between how flexible or stable an algorithm is [17]. A stable algorithm stops adapting and learning after some time, while a more flexible algorithm will overwrite previously learned data [17]. The other problem is called the *curse of dimensionality*. According to Van Laerhoven et al. it limits the numbers of sensors that can be effectivly combined, as the algorithms quickly becomes much slower when the numbers of sensor increases [17]. In the same study Van Laerhoven et al. describe a way to get around the stability-plasticity problem for their specific implementation, but give no solution to the curse of dimensionality problem [17].

A summary to sensor fusion theory is that several sensors are connected and the data from them are combined. If the sensors are distributed the resulting system can be robust and produce a good result even when sensors are blocked or damaged. The problems with sensor fusion are related to the software algorithms that are affected by two problems: stablity-plasticity and curse of dimensionality.

4.2 Quality of Sensor Fusion

According to Van Laerhoven et al. regardless of the algorithms used there are some parameters that can affect the outcome of the context awareness [4]. These are the *quality of the sensors*, the *number of sensors*, the *complexity of contexts* and the *number of contexts*. Van Laerhoven et al. claims that a more precise sensor can register data at a higher resolution and give a more reliable result, but at the same time adding more sensors can have the same effect, without increasing the quality of the individual sensors [4]. Depending on what the system should do, some tasks are harder than others. If the task is recognising contexts, some contexts are also harder to characterize than others. Van Laerhoven et al. also claims that if the numbers of contexts the system is presented to is large it makes the task much harder and produces more errors [4]. To test those claims, an experiment was made and the results implicated that adding more sensors improved the result but if the contexts were complicated and there existed many similar contexts, simply adding more sensors did not improve the results [4].

In another study made by Van Laerhoven and Gellersen which used both accelerometers and binary sensors it was possible to compare the differences between the sensors [5]. The system with binary sensors required less resources during the computation compared to the system with accelerometers, but the study could not show that binary sensors was better than accelerometers [5].

5 Context Aware Clothing

With the help of the three concepts: context awareness, sensors and sensor fusion it is possible to construct a context aware system. Context aware clothing can be described as clothing that is integrated with this context aware system. An example of a context aware system from a paper by Van Laerhoven et al. is a lab coat that is fitted with accelerometers and a packet PC [4]. The accelerometers are used to detect when the lab coat is put on, and when it is, the pocket PC will enter valid mode. When it is in valid mode the user can then use it for further identifications. Another example is according to Martin et al. a wearable motherboard that can be used to monitor infants or soldiers and detect for example Sudden Infant Death Syndrome or the location of a bullet wound [1].

There are several reasons for using clothing when designing systems for context awareness. Textiles are an integrated part of human life and are used more or less everywhere. Lehn et al. claim that clothing make a good platform for computing, communication and sensing in a durable and reliable fashion since they are so integrated in our environment [2]. Textiles can also be integrated with electronics to form *electronic textiles* and Edmison et al. claim that electronic textiles can be worn in situations where other electronic devices cannot [3]. Martin et al. also claim that electronic textiles can be worn in situation where other computer systems would hinder the user [1]. Examples of electronic textiles can be found in Lehn et al., Edmison et al. and Martin et al. [2, 3, 1].

The use of electronics in textiles also have disadvantages. Electronics or cables might break when the fabric is stretched or destroyed if they come into contact with water. This limits the use of electronic textiles as it might get destroyed in situations where clothing is normally used.

5.1 Body Sensor Networks

Context aware clothing can be constructed through *body sensor networks* that are networks of sensors and computational devices, which are integrated into the fabric. Lo and Yang describe an example of how this can be used in health care for pervasive monitoring [12]. There exist several ways of integrating wires and sensors into clothing. One is to attach large electronic devices in pockets and utility belts, another is to connect electronics to wires in the textile and yet another is to add removable electronic modules [2]. This paper does however not focus on external electronic devices, but on electronics integrated into clothing.

Lehn et al. claims that sensor modules on printed circuit boards (PCB) is a good technology since it is a mature manufacturing process with a reasonable cost [2]. Lehn et al. also claims that PCB can be reused on different textiles and can be integrated with different sensor technologies [2]. A long term goal is to have flexible, movable, *hot-plug able* and reprogrammable devices. A device that is hot-plug able can be removed from the system during operation, without damage to the device or disruption to the system. Lehn et al. mention that this might be done with specialised PCB and sensor modules from different manufacturers [2]. By using removable sensor modules it is possible to remove them if the garment should be washed or used in an environment that would provide hostile to the sensors. Lehn et al. also mention that the communication between the sensors should be made digital to reduce the amount of noise [2].

There are also several techniques to connect the sensors to the wires, as well as provide power to them. Lehn et al. mentions soldering, snaps, ribbon wire connectors, raised wires in the fabric, bus connections and cross-seam connections [2]. According to Lehn et al. soldering would produce the most comfortable connections but the manufactoring process would be both complex and expensive [2]. Two methods that according to Lehn et al. would be simple and cheap with existing manufactoring processes are snaps and raised wires in the fabric [2]. As a summary, all methods have their advantages and disadvantages and according to Lehn et al. the connections across seams are a problem as the connections there must be able to withstand the pulling and stretching of the fabric [2].

If the individual sensors do not have batteries or other means of getting power, then the power needs to be distributed. According to Lehn et al. there exist two ways of doing that, *global power distribution* and *local power distribution* [2]. They claim that global power distribution leads to lower complexity but can introduce problems with power fluctuations, noise and low reliability [2]. According to them local power distribution avoids these problems but introduces further hardware requirements and raises the complexity of the system [2]. Van Laerhoven et al. describes a pin&play architecture that can be classified as global power distribution [4]. It has a layered surface on the textile that acts as a bus for communication and power distribution. This on the other hand requires that all sensors supports the architecture and are of a similar type [4]. If the sensors are not connected in this way they need their own individual power source. Lo and Yang claims that the two main power sources for sensors are batteries and fuel cells [12]. Another technique that is mentioned by Lo and Yang is power scavenging from for example motion, vibration or temperature differences [12].

A summary to the subject of body sensor networks are that they can be integrated into the fabric by different methods. All methods for integrating the sensors and their connections have their drawbacks, but snaps and and raised wires are the cheapest. Power can be distributed to the sensors through a global or local distribution, in a local distribution all sensors needs their own power source.

5.2 More Aspects of Clothing

There are more to context aware clothing than sensor networks and how they are connected. Since clothing is an important part of human culture, there are other aspects that needs to be considered. Depending on the current fashion, clothes have different sizes, material or colour. Martin et al. mention the problem that clothes have different sizes and how this affects the readings from the sensors in the garment as they move around or are affected by the size of the garment [1]. Different types of clothes also create different amount of errors. In a study by Van Laerhoven et al. the difference between accelerometers placed on a pair of pants and a lab coat compared with accelerometers strapped to the users body was measured [20]. One of the results was that the closer to the body the accelerometers were, the less error they produced [20]. Sensors that are used to detect the users location or position could be affected by the same errors, but sensors that rely on information from an external source should be less effected.

In an article by Edmison et al. several problems related to context aware clothing are presented [3]. They can be divided into three categories: *accuracy*, *functionality* and *evaluation*. According to Edmison et al. accuracy is related to the trade-off between accuracy, cost and ease of use, but also to how the accuracy of the sensors change when persons of different sizes wear the garment [3]. Functionality is according to Edmison et al. related to how the woven bolts of clothes should be designed to allow for placement of sensors, actuators and computing devices on different types of garments [3]. Other problems they mention that are related to functionality are the problems with recharging batteries and how the e-textile shall remain functional even with tear and wear of the fabric [3]. A last problem of evaluation according to Edmison et al. how to measure the difference between e-textiles and a traditional approach with separate electronics [3].

5.3 Recomendations for Context Aware Clothing

We propose that the information presented in this paper can be summarized into a list of nine recomendations that can be used when constructing context aware clothing. The list contains recomendations and approaches to consider when designing context aware clothing based on the work by the authors that are refered to in this paper but we have chosen to make a list of these specific recomendations.

5.3.1 Hierarchial System

In both the TEA project by Gellersen et al. [6] presented in Sect. 2.1 and from the work by Van Laerhoven et al. [4] presented in Sect. 3.2 and Fig. 1 the designed system is based on a hierarchial structure. By using such a system, it is easier to go from sensor data, to concepts to contexts as proposed by Van Laerhoven [8] and presented in Sect. 2 and Sect. 2.1.

5.3.2 Machine Learning

A way of having a system that is flexible, can respond to changes in the environment and recognise new contexts is to use machine learning. Machine learning is proposed for example in the work by Van Laerhoven [8] presented in Sect. 2.1 and Van Laerhoven et al. [17] in Sect. 4.1.

5.3.3 Design for the Unexpected

Since humans according to Bellotti and Edwards make unpredictable judgements and improvise [9], as presented in Sect. 2, it is good to try to design for the unexpected. Bellotti and Edwards claims that to provide context aware alternatives to the users and let them do the decision themselves is better than having the system deciding what to do [9].

5.3.4 Connect Sensors in Networks

When using many sensors, there are according to Van Laerhoven et al. two ways of connecting sensors which are presented in Sect. 3.2: the centralised and the distributed approach [4]. When the aim is to simplify the implementation or use a single processor, the centralised approach can be used [4]. When the aim is to make a more robust system where sensors and network components can fail without taking down the whole network, the distributed approach can be used instead [4].

5.3.5 Combine many Different Sensors

According to Van Laerhoven and Gellersen et al. and presented in Sect. 4.1, when combining data from several sensors, the combined result is better than any of the individual sensors can produce on their own [8, 6]. In a study by Van Laerhoven et al. presented in Sect. 4.2 it was shown that adding more sensors had a similar effect as having sensors of higher resolution [4]. By combining sensors of different types, the result can be improved further. Hightower and Borrellio claim that combining sensors that have different error distribution can improve the result and that the more different the sensors are, the more effective the combined system will be [14]. Their claim can be found in Sect. 4.

5.3.6 Electronic Textiles

According to the work by Martin et al. and presented in Sect. 5, electronic textiles can be worn in situations where other wearable computer systems would hinder the user [1]. According to both Lehn et al. and Edmison et al. and presented in Sect. 5, electronic textiles are good since they blend into the environment and textiles are a part of everyday life [2, 3].

5.3.7 Sensor Modules on Printed Circuit Boards

According to Lehn et al. and presented in Sect. 5.1 printed circuit boards (PCB) are useful for sensors integrated into electronic textiles as they are cheap, can contain a lot of different sensors and can be reused for different purposes [2]. When sensors can be reused and removed from an electronic textile, the purpose of the system can change over time.

5.3.8 Power Distribution

According to Lehn et al. and presented in Sect. 5.1 there are two ways to distribute power to sensors in an electronic textile: global and local power distribution [2]. When all sensors have the same voltage requirements and the goal is to reduce the complexity, Lehn et al. claim that global power distribution should be used [2]. If the sensors have different voltage requirements or higher fault tolerance is needed, they claims that local power distribution should be used instead [2].

5.3.9 Sensors Close to the Body

According to Martin et al. and presented in Sect. 5.2 there is a problem of having sensors in loose-fitting garments as the sensors move around as the wearer moves [1]. In Sect. 5.2 the same problem is mentioned in a paper by Van Laerhoven et al. and the conclusion is that if the sensors are placed closer to the body, the data from them contain less errors [20].

6 Discussion

Context aware clothing is an interesting subject and there are several approaches that have been taken towards it. The studies by Van Laerhoven and Van Laerhoven et al. are one [17, 4, 20, 5, 8]. Another one is the studies by Lehn et al., Edmison et al. and Martin et al. [2, 3, 1]. The studies by Van Laerhoven and Van Laerhoven et al. mostly focus on sensor fusion and sensor strapped to the body. The studies by Lehn et al. Edmison et al. and Martin et al. focuses more on electronic textiles but refer to some of the studies by Van Laerhoven et al.

This paper has not focused so much on the so called traditional approach with a single sensor and how this is different from sensor fusion. A question that therefore remain unanswered is if sensor fusion is better then the traditional approach at all. From the studies mentioned above, it seems that the sensor fusion approach is applied to different types of problem than for example computer vision and image processing tries to solve, or approach the problem from another angle. If a more in-depth study was undertaken, it would be interesting to see if there are any previous studies made on the difference between the two approaches on a specific problem, or to do such a study. A related issue that is not covered is conflicts between sensors and how sensor fusion handles this.

Other things that are not discussed in this paper are the issue of privacy and how it can be protected. When a user wears clothing that can register data about the users context and actions, it is important that this data is only used by the application that it was meant to be used by. There are also other technologies, techniques and methods that were not covered by this paper. If a larger study on this subject was made, these issues could be covered in more detail as well.

The recomendations are based on the work by the authors that are refered to in this paper. The recomendations only take these articles in account and there are other views on how to create context aware clothing. It is hard to construct recomendations or guidelines that can be applied in all situation and to all cases of context aware clothing. Therefore these guidelines should be seen as something that could be considered since they take up issues found in previous studies, but each case is different and they will not be useful in all possible situations.

To improve the result from sensor fusion it seems to be good to increase the number of sensors, but it might be better to find the optimal number of sensors instead. It also seems to be good to combine different types of sensors. A conclusion from this is that it should be good to include many sensors that complement each other. It could be a network of sensors using sensor fusion that detects how a system changes position that is combined with computer vision or a system that detects changes in the environment and combines this with information from an existing infrastructure, for example a wireless network.

For both network connections between sensors and power distribution, there are a centralised and a distributed approach. The distributed approach is more robust to errors but is more complicated than the centralised approach. Further studies on the distributed approach would be good since clothing is subject to wear and tear, and if the sensors are depending on a single wire for network connection or power the system will not be very reliable.

Further improvements to the area of sensor fusion might also come when sensors and other hardware get cheaper and shrink in size. The problem of connecting sensors and distribute power in a reliable, simple and cheap fashion also needs to be solved.

6.1 Future Work

There are aspects of context aware clothing that could not be covered in this paper and might provide opportunities for future work. They include the social aspects of clothing and how it can be combined with electronic textiles, how context aware clothing can be used in different situations and how different types of clothes change the way sensors are spread out in the garment. Since clothing is not only used for covering the body but also affected by fashion, trends and culture the integration of sensors into clothing might not be straight forward all the time. A garment on a person, in a wardrobe or in store could also be made to behave differently. On a person it might try to adapt to the persons surroundings, in a wardrobe it might reload its batteries and in a store it might advertise itself. Adding sensors to for example a skirt or a diving suit also provides different challenges and the requirements change depending on the type of clothing used.

7 Summary and Conclusion

The aim of this paper was to to provide an answer to how a set of distributed sensors can provide context awareness in clothing with sensor fusion. This have been done by breaking down the problem in four parts: context awareness, sensors, sensor fusion and context aware clothing. The first three parts provided the background to the fourth part. A context aware system depends on sensors to get data on its context and the sensors can be distributed and connected in different ways. By using sensor fusion to compute the incoming data the system will get a view of the world that is better than any of the individual sensors can provide on their own. The sensors can then be spread out over a body in a body sensor network that might be implemented in an electronic textile. The information presented in the paper has been summarized into nine recomendations for context aware clothing.

References

[1] Martin, T., Jones, M., Edmison, J., Shenoy, R.: Towards a design framework for wearable electronic textiles. In: Proceedings Seventh IEEE International Symposium on Wearable Computers. (2003) 190–199

[2] Lehn, D.I., Neely, C.W., Schoonover, K., Martin, T.L., Jones, M.T.: e-tags: e-textile attached gadgets. In: Proceedings of Communication Networks and Distributed Systems: Modeling and Simulation. (2004)

[3] Edmison, J., Lockhart, T., Martin, T.: An e-textile system for motion analysis. Studies in Health Technology and Informatics **108** (2004) 292–301

[4] Van Laerhoven, K., Schmidt, A., Gellersen, H.W.: Multi-sensor context aware clothing. In: Proceedings Sixth International Symposium on Wearable Computers. (2002) 49–56

[5] Van Laerhoven, K., Gellersen, H.W.: Spine versus porcupine: A study in distributed wearable activity recognition. In: Eighth IEEE International Symposium on Wearable Computers (ISWC'04). (2004) 142–149

[6] Gellersen, H., Schmidt, A., Beigl, M.: Multi-sensor context-awareness in mobile devices and smart artefacts. Mobile Networks and Applications **7**(5) (2002) 341–351

[7] Dey, A.K., Abowd, G.D.: Towards a better understanding of context and context-awareness. Technical Report GIT-GVU-99-22, Georgia Institute of Technology (1999)

[8] Van Laerhoven, K.: The pervasive sensor. Lecture Notes in Computer Science **3598** (2005) 1–9

[9] Bellotti, V., Edwards, K.: Intelligibility and accountability: Human considerations in context-aware systems. Human-Computer Interaction **16**(2) (2001) 193–212

[10] Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Pearson Education, Upper Saddle River, New Jersey 07458 (2003)

[11] Pfeifer, T.: Redundant positioning architecture. Computer Communications **28**(13) (2005) 1575–1585

[12] Lo, B., Yang, G.Z.: Key technical challenges and current implementations of body sensor networks. In: Proc. 2nd International Workshop on Body Sensor Networks (BSN 2005). (2005)

[13] Hanson, M.A., Powell Jr, H.C., Barth, A.T., Ringgenberg, K., Calhoun, B.H., Aylor, J.H., Lach, J.: Body area sensor networks: challenges and opportunities. Computer **42**(1) (2009) 58–65

[14] Hightower, J., Borriello, G.: A survey and taxonomy of location systems for ubiquitous computing. Computer **34**(8) (2001) 57–66

[15] Brooks, R.R.: Highly redundant sensing in robotics - anologies from biology: Distributed sensing and learning. In: Proceedings of the NATO Advanced Research Workshop on HIghly Redundant Sensing in Robotic Systems. (1988)

[16] Murphy, R.R.: The introduction to AI Robotics. The MIT Press, Cambridge, Massachusetts 02142 (2000)

[17] Van Laerhoven, K., Aidoo, K., Lowette, S.: Real-time analysis of data from many sensors with neural networks. In: Fifth International Symposium on Wearable Computers. (2001) 115–122

[18] Klein, L.A.: Sensor and data fusion: a tool for information assessment and decision making. SPIE Press (2004)

[19] Mitchell, T.M.: Machine Learning. McGraw-Hill (1997)

[20] Van Laerhoven, K., Kern, N., Gellersen, H.W., Schiele, B.: Towards a wearable inertial sensor network. In: Eurowearable. (2003) 125–130

Investigating ways to perform usability testing on multi-touch applications

Erik Carlsson

Abstract: This article has its main focus on investigating earlier work that has been done in usability testing on multi-touch applications. The article contains some overview of what usability is and different usability methods, it also describes the multi-touch technique and challenges when designing a multi-touch interface. From the information and the case studies a conclusion has been drawn of how to design for a multi-touch interface in terms of usability and what usability methods that should be used.

1 Introduction

Multi-touch applications are using one of the most recent techniques on the market to support interaction through a touch interface. Since multi-touch is a fairly immature way to interact with applications it has some limitations and flaws that are yet to be resolved [6].

Multi-touch contains of a touch sensitive interface, the interface can be navigated through the users touch or by different gestures. As the name multi-touch suggests the interface can be navigated with multiple touches at the same time, it is in these cases the gestures set of the application comes in play. Different gestures have different outcomes, it is up to the designer to decide the outcome of each gesture. Gesture sets and multi-touch techniques are going to be investigated more in Section 2.2, 2.3 and 4.

The most common way of interacting with computers and mobile devices is to have physical representations for input and objects representing the virtual pointing device e.g the mouse and keyboard [10]. The purpose of this article is to highlight pros and cons of multi-touch interaction technique. From the conclusions try to investigate how to perform usability testing that minimises the negative- and enhances the positive-side of multi-touch. Usability testing is a well known approach to investigate how well a application performs and how the users interact with it, this can be established in many different ways but the most common way are creating scenarios and making studies of users interacting with the system. In short, usability testing is a way to involve endusers in the design process for finding usability problems with the application or the system [5]. Many studies have been made on how to construct and carry out usability tests, some of these studies will be covered in Section

3. The multi-touch technique is evolving and many wearable applications now have touch interfaces. A problem is that not much has been made concerning usability testing of multi-touch applications, this is the main reasons why this article focuses on investigating this further.

This article is meant to be a guideline for people who is about to perform a usability test on a multi-touch application.The article has main focus on investigating earlier work in usability testing and multi-touch techniques and from this the conclusion of how to use this result to suggest the best usability method or methods for developing multi-touch applications will be drawn.

The article will cover what usability testing is (in Section 2.1), what methods that are common and the function of them (in Section 3), what multi-touch is (in Section 2.2) and different problems when using multi-touch (in Section 2.3). There are also some studies of related usability investigations on multi-touch applications (in Section 4 and 5).

2 Background

This section introduces some background information about multi-touch and usability, it also describes previous work and definitions in the subject.

2.1 Usability testing

Usability testing is a way to investigate an application or a system from the end-users point of view, when you involve users in this part of the design-phase you get a clearer picture of how the user interact with the system and what the users attitude towards the system is [9]. This is a very important step because the designer gets a better understanding of how to design and what to make more efficient in the system.

The most common definition of usability is the ease of use and acceptability of a system for a particular class of users carrying out specific tasks in a specific environment [1]. In this definition the ease of use is how well the system satisfies the users expectations and needs, acceptability in the definition is whether the user accept the usability flaws and still uses the system. Usability testing has been performed and was set as a standard when evaluating graphical user interfaces from the beginning of 1990 [11], but even as early as in the beginning of 1980 some companies were using usability testing. There are many ways to carryout a usability test and the methods can have different directions depending on what the test instructor want to investigate e.g if the cognitive flow in the system is of main focus there are no idea to make a standard usability test, some of the existing methods are going to be investigated and straighten out later on in the article.

As said usability testing is a way to involve users in the design process, one thing that is similar to usability testing but not the same is usability inspection. Usability inspection is based on the same thoughts as usability testing but when using usability inspection methods you do not involve users, instead usability experts investigates the system through different inspection methods[11].

2.2 Multi-touch

The focus of this article is multi-touch applications that support a single user, not several simultaneous users. This does not mean that all of the problems and advantages that are covered in the article just apply to multi-touch systems with single users, they can still be the same and solved in similar ways for multiple users.

The basic idea of multi-touch interfaces is that the device can recognise simultaneous touch from users, and through various gestures the device can interpret what to do with the information that the touches provide. For example the two most known gestures for interacting with a multi-touch application are the rotate and scale gestures, as can be seen in Figure 1. The scale gesture contains of two touch points, when the user bring them closer or wider to each other the object that are manipulated scales up or down. The rotate gesture is similar to the scale gesture, but instead of moving the two points closer or wider to each other the user move them in a circle, always keeping the two touch points on opposite sides, this results in a rotation of the object.

Multi-touch interaction is a way for the user to perform continuously interaction with the application in a high degree of freedom. This is the main advantage of multi-touch interaction compared to single point interaction (like a mouse pointer) that works in a more discrete way in matters of degree of freedom [6]. There are many different techniques to be used when developing multi-touch devices, the one that is most common is the technique Apple has implemented in Iphone and Ipod-touch. In this technique the sensing-system is embedded into the overlay of the screen and the users touch triggers sensors through electrical fields that wires information to the LCD-screen [3]. This way the user can interact with the application in terms of clicking, dragging, rotating and scaling (see Figure 1), all of this to allow the user to get a richer and more intuitive interaction possibility [8].

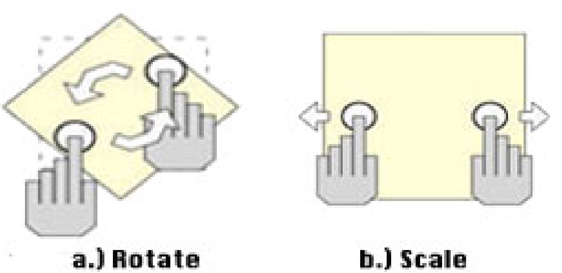

Figure 1: The pictures describes the standard gestures of how to rotate(a) and scale(b) in a multi-touch application [8].

Developers work to make the transition between the virtual world and the physical world as smooth as possible for the user. In terms of this the multi-touch technique is a step in the right direction [15].

2.3 Challenges with multi-touch applications

Many early multi-touch applications had problems with bad user performance and the users experienced difficulties understanding how to interact with the interface. The main reason for this was that the designers of the multi-touch interaction technique had no principles and studies to rely on when designing, instead he or she only guessed based on their own intuition [7]. Now a days designers know better how to design for this type of interaction and many of the multi-touch applications on the market are easy and intuitive to navigate.

One disadvantage with multi-touch applications is that they can be hard to navigate and interact with for users that have some sort of physical limit or disorder (like rheumatism) and users that have reduced visual perception [15]. To avoid this it is important to design alternative settings so that the user can pick the settings that satisfies the personal needs, for example persons with reduced eyesight can pick a setting that shows larger interface icons. To avoid and work around the limitations of multi-touch interaction technique the designers have to make some important design decisions.

2.3.1 Important design decisions when developing multi-touch applications

Physical constrains

The physical constrains that become a problem when designing a multi-touch interaction technique is that a user's finger is often larger than the standard icons and widgets, which are making accurate pointing difficult [7]. Another constraint that the users hand and fingers are likely to obscure the objects that are being manipulated. A designer also have to keep in mind as earlier mentioned the psychical attributes of the user, e.g eye perception and ergonomic position for the user when interacting with a interface [15].

When designing the multi-touch interaction technique it is also important to consider if the device is going to be handheld or wearable. When designing for example a mobile-phone interface the designer has to think about how to design so that the device can be held in the right or the left hand and still have a efficient interaction possibility independent of what hand the user chooses to interact with [2].

The solution for the physical problems in multi-touch applications can be, as Moscovich et al. [7] describes, by doing a more complex and indirect mapping between the display space and the control space.

Mapping

Mapping is as Norman et al. describes it, a technical term that means the relationship between two different things [12]. This relationship can be either a natural mapping (which means that the user immediately can understand what an action will lead to, often through well known or easy understandable metaphors) or a learned mapping (the user have to learn what the action leads to). In graphical user interfaces mapping is the relationship between actions (e.g gestures) and the possibility to predict the result of these actions. Example of a good mapping is when it is possible for the user to determine the relationship of actions and there effects. When it comes to mapping of the interface the designer can not in a satisfactory way predict how effective the chosen mapping is going to be, this is because the mapping that each user experiences is a physiological and cognitive factor [7] .

The solution to this problem can be, as Moscovich et al. [7] suggests, by examining the relationship between the degree-of-freedom of the hands and the control task.

Another way to increase the users understand-ability for the mapping is to provide the user with feedback while performing interaction with the system, this will result in easier and faster understanding of the systems mapping.

Gestures

The third thing that is a challenge for the designer of multi-touch interaction are what gestures to implement. Here the designers should ask them selves the questions; What to use? Were to use it? and Why should it be used? The problem with gestures in multi-touch is that with only touch interaction possibilities there can be several ways to perform a gesture that feels natural for the user, a gesture that feels natural for one user may not feel natural for another. Therefore the designer should carefully investigate the questions what, where and why before implementing any sort of gesture interaction [14].

One easy way to solve this problem is to make use of standard gestures that are already known, another way is to perform usability testing and from that draw the conclusion on what gestures that are appropriate to use [14].

3 Usability engineering

Bevan et al. propose three views for usability measurement [9].

- **The product-oriented view** that usability can be measured in terms of the ergonomic attributes of the product;

- **The user-oriented view** that usability can be measured in terms of the mental effort and attitude of the user;

- **The user performance view** that usability can be measured by examining how the user interacts with the product. The designer should concentrate on either - ease-of-use: how easy the product is to use, or - acceptability: whether the product will be used in the real world.

This article concentrates on the last one of the views namely the user performance view. In Nielsens book "Usability Engineering" this point of view has been divided into five usability attributes [10].

- **Learnability**

  The system should be easy to learn so that the user can rapidly start getting some work done with the system.

- **Efficiency**

  The system should be efficient to use, so that once the user has learned the system, a high level of productivity is possible.

- **Memorability**

  The system should be easy to remember, so that the causal user is able to return to the system after some period of not having used it, without having to learn everything all over again.

- **Errors**

  The system should have a low error rate, so that users make few errors during the use of the system, and so that if they do make errors they can easily recover from them. Further, catastrophic errors must not occur.

- **Satisfaction**

  The system should be pleasant to use, so that users are subjectively satisfied when using it; they like it.

Usability methods

There are several methods and ways to carry out a usability test, but the most common ways are thinking aloud, field observations and questionnaires.

Thinking aloud is a method were the test subject performs a scenario with different tasks and at the same time speaks out loud of whatever he or she is thinking about during the interaction. From this method a designer can collect not just data from the interaction, but also the thoughts and feelings from the participant. Thinking aloud is used in the design phase and requires a high amount of time to conduct.

Field observations is a method that is applicable in the final testing phase. The method works as it's name foretells, the designer observe how users interact with the system in the real environment. This usability method takes medium amount of time and have high demands on the designer in terms of knowing what to observe and how to interpret the observation.

Questionnaires is an usability method can be applied all through out the design process, it takes low amount of time and is easy to conduct. The designer hands out different questions that he wants information about, after the questions are answered the designer compiles the different answers. This is a good and easy way for the designer to get knowledge of what the user thinks about the application or system [5].

Some comparisons of these methods have been put together in Figure 2 [5]. One important thing that has to be done before testing the system on users is to construct a usage scenario that covers the important parts of the system. This scenario contains different tasks that the participants are going to execute in the usability test.

Important to have in mind when planing to perform a usability test on an application is to try to schedule the test early in the design phase and make usability testing throughout the whole process. Finding problems early is crucial because of making changes to the interface when it is already implemented can be of high cost and steals time from the process [5]. With this said, one should not underestimate the power of performing usability testing methods like thinking aloud and field observations on a finished application or system, this maybe the most accurate way for the developer and designer to get a understanding of how the users of the system interact with it and what feeling they have for it [4]. The three usability methods that are described have different

		Usability testing methods	
	Thinking Aloud	Field Observation	Questionnaires
Applicably in Phase	design	final testing	all
Required Time	high	medium	low
Needed Users	3+	20+	30+
Required Evaluators	1	1+	1
Required Equipment	high	medium	low
Required Expertise	medium	high	low
intrusive	yes	yes	no

Figure 2: Comparison of three commonly used usability testing methods (adopted from [5]).

advantages and disadvantages, the main advantages for Thinking aloud is that it pinpoints the users misconceptions and it is a cheap test. Main advantage for Field observations is that the designer gets a good ecological validity, can reveal the users real tasks and get some suggestions of functions and features. Disadvantage for Thinking aloud is that it is unnatural for the users to interact and at the same time talking about what they are doing. Disadvantage for Field observation is that it can be hard to be allowed performing such a study in terms of project schedule [10]. The disadvantage with questionnaires is that the participants can not feel free to express them selves, instead they are bound to answer the questions on the questionnaire. The advantage of questionnaires is that it is a fast and cheep way to get some pinpointed information about the application or the system.

Vital when choosing usability method to work with is to know that they all have areas where they perform well, but also areas were they perform less well. Since different usability testing methods addresses different parts of the design process it is a good idea to not rely on a single method but instead try to combine them so that they can complement each other in terms of advantages and disadvantages [10].

4 Related work

In this section related work of how usability testing has been performed on multi-touch systems are investigated.

4.1 Case one, Physical

This is an evaluation conducted by Potter et al. in interest of investigating how errors when pointing on a multi touch screen (in terms of accurate pointing) can be solved through measuring performance and using usability testing [13]. They made a comparison between three different touch techniques to bee able to tell which one of the techniques performed best in terms of fewest errors. The different techniques that was analysed were Land-On (were cursor is direct under the finger and the screen register the first touch), First-Contact (cursor is direct under the finger but supports dragging) and Take-Off (cursor is 1/2 inch above finger and click is made through lifting finger, supports dragging).

Twenty-four participants participated in the experiment, the subjects was told what they were going to do and how to do it. Each of the twenty-four subjects were individually measured in the experimental session. Every one of the participants were tested on the three different techniques (Land-On, First-Contact and Take-Off). The test was constructed so that the subjects had five practice turns that were followed by fifteen real test for each technique. If an error was made the participant had to try again until the task was solved. Potter and the rest of the supervisors counted numbers of errors that occurred during the tests and analysed the variance of performance of the different techniques. After all the trails were done, the participants were asked if they could rate the different strategies in terms of how satisfactory, the ease of learning and grade of awkwardness they experienced during the testing session.

The result that Potter got was that he was able to identify two kinds of errors, one was when the participant selected the wrong target and the second error was when the subject touched a blank part on the screen. He also found out that it was a significant connection between type of error and type of technique. The First-contact and land-On technique had similar amount of number of wrong target errors. The Land-On technique had significant more blank touch errors. The technique that performed best in the experiment in terms of errors were the Take-Off technique, it had least wrong target errors.

The analyse of the questionnaires showed that the Take-Off technique had higher rating of satisfaction than the other two techniques.

In this example Richard L. Potter used two different techniques, one to analyse which type of touch technique performed best, and also made a sort of questionnary to collect the participants thoughts and feelings of the different techniques. The second method are described in section 3 in this article but the first one is more of an usability inspection method where experts analyse the data from the experiment.

4.2 Case two, Mapping

This experiment is performed by Moscovich et al. [7], they examine mapping for an object transportation and orientation task. The reason for this experiment is to get a understanding of the mapping between the users finger movements and the visual feedback that the user receives. The technique that were used in the experiment for manipulating and interacting with the multi-touch application were a touch technique that alow the user to translate, rotate and scale an object simultaneously, the technique is called "stretchies".

Six female and six men all right handed participated in this experiment. No one of the twelve subjects have had any previous contact or experience using a multi-touch application. Each participant spent twenty-five minutes filling out questionnaires and performing the test. Before proceeding to the actual test each subject had to complete different trails, each trail took approximately thirty seconds. After completing the trails the actual tests were performed, the test took about 8 minutes to complete. After all participants had completed there tests the tracking errors were collected and analysed through variance and mean values by experts. The participants also had to fill in questionnaires that contained grading of what they felt during the test in terms of usability and efficiency.

The result of the experiment showed that unimanual interaction (controlling the object with fingers of one hand) performed better than the bimanual interaction (using both hands to control the object) in terms of less errors when performing the experiment. When rotating the object the bimanual interaction had an increasing error rate of 75%, while the unimanual interaction only increased with 28%.

The feedback that came from the questionnaires were that most of the participants experienced rotating through bimanual interaction to be very difficult. One more interesting thing was that one of the subjects noticed the physical problem when fingers shadows the screen only when rotation with both hands.

What Tomer Moscovich did to investigate the mapping for a object transportation and orientation task was to measure the performance of the participants and after completed test applying the usability method questionnaires to get a better understanding of the systems usability.

4.3 Case three, Gestures

This usability evaluation is performed by Wu et al. on a multi-touch table to get a better knowledge of how effective the usability on there gesture set is, that are used as an instrument to interact with the system [14]. The gestures that Mike Wu investigated were Annotate (writing on the touch surface while relaxing the hand on the surface), Wipe (removing text from the touch surface), Cut/Copy-N-Paste (copy, cut and paste different images) and Pile-N-Browse (sort the images in a pile).

Five female and five male in the ages 19-30 years not associated with the laboratory participated during an hour-long session each. No one of the subjects had have any earlier experience with touch or gesture interaction on tabletops. The participants were each given instructions on how to perform the gestures that were implemented, this before start of the session. After the instructions the subjects were given tasks to complete in a fix time. The subjects were told to during the experiment talk out loud on what they were thinking and why they did as they did. At the end of the session, participants were given a questionnaire asking their agreement with a collection of statements. They were also asked to list the best three things and worst three things about the interface.

After all participants had finished the test there were an obvious difference in how difficult the gestures were for the subjects to carry out. The gesture with most errors and hardest to use were the Cony-N-Paste gesture. The easiest gesture to perform were the Annotate gesture. During the experiment many errors arise, but thanks to visual feedback and the opportunity to cancel a operation the participants could correct themselves and was then able to carry out the task.

From the questionnaire important information of how the participants experienced the different tasks appeared. The thing that they felt was the easiest to implement was cancelling a mistake, the hardest thing they experienced were adjusting the size of the selection box when copying.

What Mike Wu did as mentioned in section 3 and as illustrated in Figure 2, was taking two of the most common usability evaluation methods(Thinking aloud and questionnaire) and combining them to get a good overview of how well their gesture set performed when handled of inexperienced users.

5 Discussion and Future work

As presented in this article multi-touch applications have some disadvantages, it is also a rather unknown way for users to interact. Usability testing on the other hand is a mature and well documented method to satisfy the users needs and expectations on a system or application. Therefore involving usability testing when designing for a multi-touch interface is not unusual. Problem occurs because of the wide variance of usability methods that are available for the designer to choose from. Therefore in section 4 earlier work of usability testing on multi-touch has been investigated with the intention to get a better understanding of the problems that multi-touch possesses and how they have been handled. Three main problems the interaction designer have to solve when developing the multi-touch interaction technique are as mentioned in section 2.3 physical, gestures and mapping.

So how to avoid these flaws? If we take a closer look at the case study from Potter et al. [13] in section 4, we can se that they have through the usability testing method questionnaire and through evaluation of the result from the experiment, come to the conclusion that to avoid clicking errors that comes from the user "shadowing" the screen with there hand while clicking it is preferable to use a Lift-Off clicking method that have the clicking area 1/2 inch above the finger. Performing studies like this, given that the goal is to investigate what clicking technique results in least clicking errors, it is important to not just only collect the data from the experiments, but also to involve at least two usability methods to really understand if the technique that performs best are the one that users feels most satisfaction over. Potter performed a questionnaire, but in my meaning this was not enough, to really get a understanding of what the user thinks during the test, I suggest they should have used a Thinking aloud testing method combined with the questionnaires, this in order to collect all the thoughts and feelings from the participants and not "lock" them in the questions that were on the questionnaire.

The second case that was investigated is the one that Moscovich et al. [7] performed. This experiment was conducted to get a better understanding for the mapping between the hand interaction and the visual feedback. There experiment showed that manipulating objects in a unimanual way resulted in less error rate than in a bimanual way. The result from the experiment was collected through questionnaires and data collection. This is the same usability methods as Potter used and as mentioned I do not believe this is enough to in a satisfactory way being able to tell what the users are thinking of the system. One important thing to consider is however the result of this study, namely that the unimanual interaction technique performed best.

Looking at the third case from Wu et al. [14] where they investigate a gesture set, I believe that this is the best performed test of the three in terms of usability. Wu make use of two different usability methods, Thinking aloud and questionnaires, and from this he gets a wider perspective of how the gestures are embraced from the users. The result is clear on what the participants think is the hardest thing to do and what the easiest thing to do is. Another vital observation that was mad were that the participants appreciated the cancelling possibility in the application.

To summarise the observations that I think is important from the case studies are that when evaluation a multi-touch interface it is central to combine usability methods to get accurate result of the users experience. The combination can wary but in these cases thinking aloud and questionnaires are to prefer. Other observations that have been made are that to avoid the problems with multi-touch interaction it is good to use the Lift-Off clicking technique, try to design for unimanual interaction and when implementing different gestures it is important that the user has the alternative to cancel an operation. Provide the user with feedback when interacting with the system is important to facilitate the mapping of the interface and to make the system easier to learn and remember. As mentioned in section 3 different usability methods focus on different parts of the design process, therefore it is easier to combine methods that are at the same stage and plan when to use different methods.

In future work it would be interesting to investigate how different usability methods performs on different multi-touch applications, to investigate if there can be a standard way for evaluate specific applications.

References

[1] John Brooke, Nigel Bevan, Fred Brigham, Susan Harker, and David Youmans. Usability statements and standardisation: Work in progress in iso. In *INTERACT '90: Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction*, pages 357–361, Amsterdam, The Netherlands, The Netherlands, 1990. North-Holland Publishing Co.

[2] Bill Buxton. Multi-Touch Systems that I Have Known and Loved. *Microsoft Research*, 2007.

[3] A. Crevoisier and G. Kellum. Transforming ordinary surfaces into multi-touch controllers. In *NIME 2008*, pages 113–116, Genova, Italy, 04/06/2008 2008.

[4] Joseph S. Dumas and Janice C. Redish. *A Practical Guide to Usability Testing*. Intellect Books, Exeter, UK, UK, 1999.

[5] Andreas Holzinger. Usability engineering methods for software developers. *Commun. ACM*, 48(1):71–74, 2005.

[6] Tomer Moscovich. *Principles and applications of multi-touch interaction*. PhD thesis, Providence, RI, USA, 2007. Adviser-Hughes,, John F.

[7] Tomer Moscovich and John F. Hughes. Indirect mappings of multi-touch input using one and two hands. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1275–1284, New York, NY, USA, 2008. ACM.

[8] LYL Muller. *Multi-touch displays: design, applications and performance evaluation*. PhD thesis, University of Amsterdam, June 2008.

[9] J Maissel N Bevan, J Kirakowski. What is usability. In *Proceedings of the 4th International Conference on HCI*, pages 1–5, 1991.

[10] J. Nielsen. *Usability engineering*. Morgan Kaufmann Publishers, NY, 1994.

[11] Jakob Nielsen. Usability inspection methods. In *CHI '94: Conference companion on Human factors in computing systems*, pages 413–414, New York, NY, USA, 1994. ACM.

[12] D.A. Norman and B. Collyer. *The design of everyday things.* Basic Books New York, 2002.

[13] RL Potter, LJ Weldon, and B. Shneiderman. Improving the accuracy of touch screens: an experimental evaluation of three strategies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 27–32. ACM New York, NY, USA, 1988.

[14] M. Wu, C. Shen, K. Ryall, C. Forlines, and R. Balakrishnan. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Proceedings of IEEE TableTop-the International Workshop on Horizontal Interactive Human Computer Systems*, pages 183–190, 2006.

[15] www.educause.edu. 7 things you should know about. multi-touch interfaces. 2008.

A Comparison between Web Applications and Desktop Applications

Martin Hannerfors

Abstract: Internet and the World Wide Web have since the mid-nineties been part of a transformation. The Web has gone from distributing static page-based information to serve as a medium for advanced and rich Web application access. The applications on the Web that earlier only were used for information gathering through forms are now fully interactive with client-sided processing similar to the traditional client-server applications. The line between desktop applications and Web applications are getting thinner and it is now appropriate to evaluate if it still is necessary to separate the development into categories with two different types of applications as result. This paper studies the differences between Web applications and desktop applications by looking into commonly used Web technologies and quality attributes of Web applications. The result shows that Web applications are almost as capable as desktop applications when it comes to processing and interactive user interfaces. But that there are other user related factors separating their development.

1 Introduction

The Internet has over the last ten years been part of a transformation and evolution. Both the use of the Internet and its content have changed and matured since its advancement in mid-nineties. The World Wide Web has gone from serving as a communication channel and displaying information to providing fully functional e-commerce and a medium for rich interactive application access [20]. During the mid-nineties HTML form-based and browser dependent Web applications, called thin clients, caused a paradigm shift. The application trend until then had been client-server applications with increasingly responsible clients. The thin Web application clients pushed the control and processing back to the server side and only took responsibility for simple interaction and rendering of presentation [11]. This first generation of Web applications were both easy to use and easy to deploy, despite the limited user interface. The benefits out weighted the downsides and the thin Web applications were quickly adopted [21]. Since then the technologies for Web applications have matured to fully functional developing environments. The Web technologies used today produces rather adequate Web applications with rich user interfaces and processing abilities similar to traditional desktop applications. Two commonly used technologies are Ajax technology [12] and the Flex framework [9, 14], both supporting rich user interfaces and client processing.

The new generation of Web applications is blurring out the line between Web applications and desktop applications as the Web applications are reaching the criteria's of an ideal Web application stated by Duhl et al. [9, 8]. The discipline of Web Engineering has been introduced by Murugesan et al. (1999) trying to establish a systematic approach for successful development and deployment of high quality Web based systems and applications [17]. Today the field of Web development is fully mature and is almost equally established as regular software application development, there seems no longer to be a gap between the two different application forms. The goal of this paper is to investigate if it still exist any differences between developing a rich Web application and a desktop application. In order to be able to do a comparison between them the definition of a Web application and a desktop application will be constrained and stated, e.g. constrain the types of Web applications to thin and rich clients and only look at Web applications that run within the browser. The research question this paper will answer is: Is there enough difference between Web applications and desktop applications to make a distinction between their development?

The paper is organized as follows: after this introduction Section 2 will explain and define what a desktop application is referred to in this paper. In Section 3 Web applications will be explained in detail, different Web applications will also be placed in a continuum from thin clients to rich clients. Section 4 and Section 5 will define thin and rich clients and explain Ajax technology and Flash/Flex, two technologies used for developing rich clients. In Section 6 quality attributes of Web applications are discussed and in Section 7 this paper ends with discussion and conclusions.

2 Desktop Applications

A desktop application is an application running in a desktop or laptop computer designed for solving certain tasks. It has high processing power capable of performing complex tasks with a high user interaction [19, 5]. The desktop application is running on top of a windowing operative system and uses a rich graphical user interfaces with behaviors such as drag-and-drop and direct manipulation [9, 3, 13, 30]. The typical desktop application is installed on a computer and is executed from a local stored hard drive. It can either work offline or online [9, 19]. If the application is running offline it executes both the front- and the back-end of the application [32]. When the application is functioning in an online state, e.g. communicating with a server using a centralized database, the application it most often only the front-end part of the application. The application is then called a client-server application and has less responsibility, e.g. only responsible for rendering the graphical user interface and supply the user with interaction possibilities.

A desktop application can communicate almost directly to the computer's

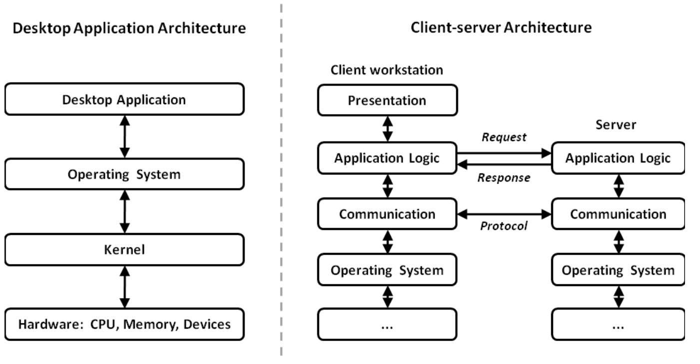

Figure 1: An overview of a generic desktop application architecture and a generic client-server architecture based on Stallings et al. [26, 29, 28].

hardware through the operating system and the kernel, see Figure 1, giving it its processing power. The application is responding almost instantaneously to user actions; the delay is minimal and is only affected by reading and writing to the physical memory or permanent storage and when computations result in delays [9, 5]. Several guidelines and heuristics concerning usability [18] and interface have been produced for supporting high quality desktop applications.

In this paper the term desktop application refers to an application installed on a client computer. It has a rich and responsive user interface supporting methods such as drag-and-drop and direct manipulation and gives the user good input/output feedback. The referred desktop application operates in an offline state making it solely responsible for all processing and rendering.

3 Web Applications

A Web application is accessed from the Web and is often used through a Web browser. Web applications can vary widely in appearance, from small-scale and short-lived read-only services to large-scale and complex enterprise applications [7]. There are some similarities between Web applications and regular content-based websites. Just as a web-page a Web application uses Hypertext Transfer Protocol (HTTP) to transfer information to and from a server and the Hypertext Markup Language (HTML) is commonly used for rendering and display [32, 6, 24]. HTTP is the most important and popular transport protocol on the Web and defines how messages are formatted and transmitted between a client and a server. The protocol is stateless, meaning every command is executed independently without knowledge of previous commands [13, 6, 10]. The stateless aspect of HTTP makes it suitable for its original intended purpose, transporting page-based documents, but is a rather ill fit for Web applications that prefer a conversation with the server. The difference between a Web application and a regular content-based websites is the Web application's task-orientation. The user has specific goals, tasks and expectations when arriving to the Web application in contrast to when arriving to a web-page where the purpose is browsing for general information. There are also more interactions between the application and the user that require user action and reaction in a Web application compared to when the user browse a web-page [34].

The term Web application is a rather ambiguous expression. Web applications are a heterogeneous group built on many different technologies, stretching from the early simple form based applications to the present Rich Internet Applications (RIA). The more present and interactive RIAs can, according Bozzon et al. [5], roughly be divided into four broad categories: 1) Scripting-based applications using Ajax technology; 2) Plugin-based applications with advanced rendering functionality such as Flash and Flex; 3) Browser-based (XUL) applications where rich interaction is natively supported in the browser; 4) Web-based desktop technologies where the application is accessed from the Web but executed locally outside the Web browser. A Web application continuum have been established by Wroblewski and Ramirez from thin clients to rich clients [33] where any application can be placed, see Figure 2. With this continuum the diverse Web applications can be gathered and compared as groups for easier overview. In this paper Web applications are divided into two types. The Web application is either a thin client or a rich client instead of using the continuum and a floating degree of client richness. These two types, a thin client and a rich client, are discussed more in Section 4 and Section 5.

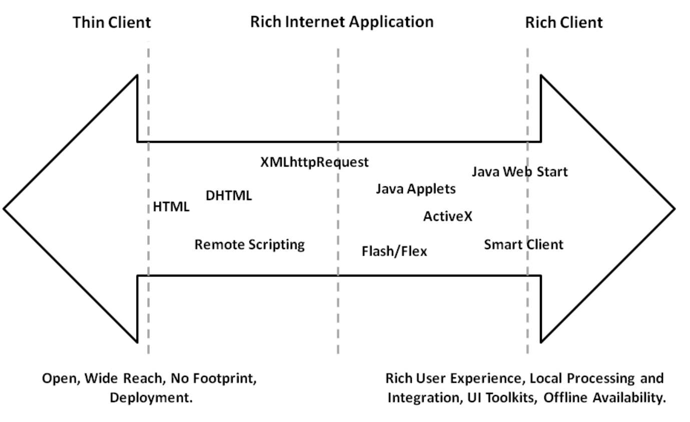

Figure 2: Illustration of the Web application continuum redrawn after Wroblewski et al. [33].

4 A thin client

The typical thin client is the traditional form based HTML Web application. The thin client runs on top of the Web browser and the functionality and logic is limited to browser scripting and the HTML element's functionality [25, 6]. The result is an application that relies solely on a remote server's business logic and with restricted user interaction. The primarily usage is information presentation and form-based information gathering. In order to process user actions or update the current state, the client must send a request to the server and then wait for a response [23]. The benefits of a thin client is its high browser and multiple platform accessibility [8]. No external plug-ins or installations are needed other than the Web browser. The architecture and coding is simple when only using HTML and it is easy to develop and deploy. The ease of use is the main reason for the thin clients' high popularity despite the low interaction [25, 21].

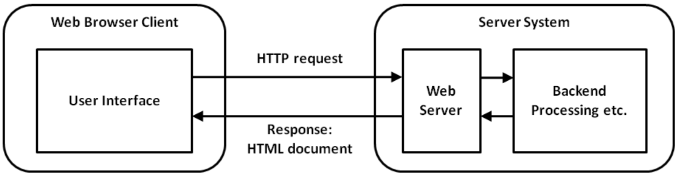

Figure 3: The thin client requesting a HTML document with HTTP through a page reload, redrawn after Garrett (2005) [12].

5 A rich client

A rich client is a Web application with a more desktop-like behavior compared to a thin client [33]. The rich client supports high user interaction and an interactivity previously only found in desktop applications, e.g. informative animations and drag-and-drop functionality but with the accessibility of a traditional Web application [8]. Even if the rich client's interactivity and processing power leans towards a desktop application the rich client still falls short compared to a desktop application. The main difference between a desktop application and a rich client is in the rich client's hardware accessibility [26, 29, 28]. A rich client can only communicate with the computer hardware through either a dedicate technology environment or through the Web browser which results in processing restrictions [21, 12, 9, 14]. The applications with the highest rich client value in Wroblewski's and Ramirez' continuum (Figure 2), e.g. applications using the web-based desktop technologies such as Java Web Start and Smart Clients are able to run outside the Web browser and in an offline state. Technically these rich clients are similar to traditional client-server applications. The other applications that are considered being rich clients are Web applications that run within a Web browser. These applications uses techniques such as Ajax, Flash or Java and are closer in behavior to the RIAs defined by Allaire et al. [3, 9]. Rich clients, especially the browser dependent RIAs, are characteristically loaded by the Web browser at page load. The application then handles data rendering and event processing separate from the browser making these applications less dependent on the Web browser. The rich clients communicate with the server only when information is required or at data submission, resulting in fewer round trips to the server compared to the traditional thin clients. [5, 21]. Today there are a several technologies possible of producing high quality RIAs, e.g. Adobe Flash/Flex, JavaFX, Microsoft Silverlight and a number of different Ajax frameworks. The following two to subsections will explain two of the more frequently used rich client alternatives, Ajax technology and the Flash/Flex framework.

5.1 Ajax

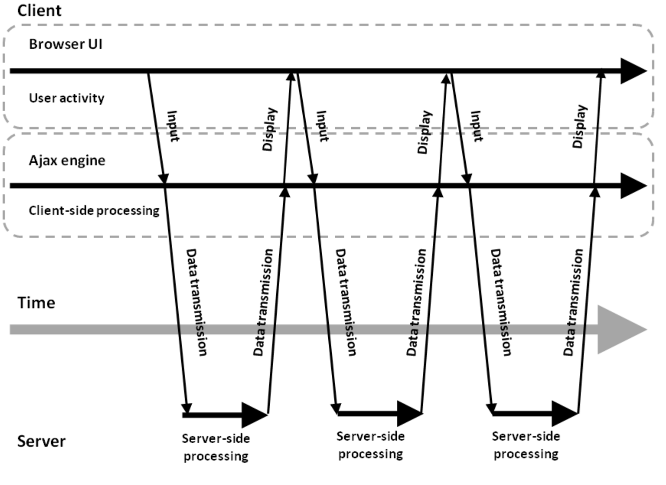

Figure 4: Asynchronous communication using Ajax. The Web browser sends an HTTP requests to the server using XMLhttpRequest. The response is either an XML document or in plain text and the web-page content can be manipulate without reloading the page using the XML data. Redrawn after Garrett (2005) [12].

Ajax is an acronym for Asynchronous JavaScript and XML and is a constellation of technologies that has been developed and standardized past the past 10 years. These technologies improves together the functionality and user interaction in Web applications [12, 22]. This combination of techniques are one of the most widely known and used techniques for building rich clients. One of the advantages with Ajax over other techniques such as Flash is that Ajax does not need any additional tools, plug-ins or runtimes during application development or before the application can execute in the Web browser [19]. The technologies required for Ajax are; XHTML and CSS for presentation, Document Object Model for dynamic interaction with data, XML and XSLT for data manipulation, XMLHttpRequest for asynchronous communication (Figure 4) and JavaScript (or equivalent) to bind the technologies together [12]. The main benefit of using Ajax technology is the possibility of asynchronous server communication in the otherwise start-and-stop characteristic of page-based websites. With Ajax small amounts of data can be retrieved from the server when needed without having to refresh the full Web page [31]. The traditional model of a Web application is by its nature synchronous, i.e. the technology platform does not support individual HTTP requests without reloading the page. With Ajax the behavior of Web application moves towards that of a desktop application [31].

5.2 Flash and Flex

Flash and Flex are rich media platforms for developing RIAs, viewable and available consistently across many different platforms, browser and devices [9, 1]. Both Flash applications and Flex applications are written in the script language ActionScript which is a well-defined object-oriented language [14]. Flash is a timeline based environment suited for designers creating animations, advertisements and other Web graphics. Flex is an offspring of Flash built for making rich client applications without the use of a timeline. Instead Flex uses a XML-based language for describing and implementing user interfaces [19]. Both Flash and Flex produces a compressed SWF files consisting of multimedia content and binary ActionScript logic [1, 2]. The SWF file is executed in the Web browser using Flash Player, which is one of the most used front-end technologies [21] used by 99 % of Internet-enabled desktops in mature markets [27] during 2008. The Flash Player is a separate plug-in to the Web browser and must be installed before a ActionScripted application is executable [4, 16]. The Flash Player has almost the same processing and interface possibilities as traditional client-server technology but with very little access to the operating system, i.e. almost no access to the file system or to the computer hardware. Because of the client-sided processing these applications allow immediate interaction with user and can at the same time reduce the traffic and load on the server and network [23].

6 Quality Attributes of Web Applications

Developing for the Web is different from other software development such as traditional desktop application development [7] and Web application development fall somewhere in between traditional software development and website development when it comes to guidelines [34]. Also when it comes to quality

attributes, measures of non-functional attributes in a system, web applications and desktop applications are different [20].
Offutt (2002) has listed seven important quality criteria for a successful Web application [20]: 1) Reliability; 2) Usability; 3) Security; 4) Availability; 5) Scalability; 6) Maintainability; 7) Time-to-market. Offutt states that even if each and every one of the criteria listed also are important for traditional software development, they are weighted differently.

In traditional software development the criteria time-to-market is rated far more important than other quality criteria, being the first to market a product or software is an important goal in the traditional software industry. In the software market it is often economically more lucrative to release a product early, even if the product itself lacks the desired quality [20]. Updates and new releases are common and accepted and have little negative effect in the long run for the application. Often the user bought a copy of the application and he or she is then more likely to adjust to a bad user interface or bad application experience compared to if the application was accessed free of charge through the Web [15]. If a user of a Web application is unhappy he or she will quickly switch to another Web application if an alternative exists, the user do not feel the need to adjust to the application to same extent as with a bought copy of a desktop application. For Web applications it is more profitable to release a better product later than being the first on the market with a poor quality application since the users tend to quickly switch application if they find anything better [20].

7 Discussion and Conclusions

In this paper the structure and base of desktop applications and Web applications have been investigated in order to study if there still exist differences between the two application platforms. The key features of the desktop application are the rich interface capabilities and its processing power making it far superior over the first generation Web applications. These early Web applications that came through in the mid-nineties could not compete against the desktop application in terms of computing power or in interface possibilities but had the advantage of simple deployment and accessibility. Today several different Rich Internet Application techniques have evolved and Ajax technology and the Flash/Flex framework are two of the most commonly used. These rich clients have both rich user interfaces and client-sided processing making them comparable to the traditional desktop application as long as the executed task only demands a moderate level of processing, for example an online word processing application. Desktop applications still have better processing capabilities due to their low-level access to the computer hardware compared to Web applications who must rely on either the Web browser's processing power or the dedicate technology environment, e.g. the Flash Player environment.

The main difference between developing Web applications compared to developing desktop applications is not in performance, but in their user's behavior [20, 15]. Desktop applications and Web applications are comparable as seen from a technical perspective. They have both client-sided processing capabilities and are able to display impressive graphical user interfaces. There are performance differences but not to the extreme extent, depending on the application's task a Web application can perform similar result as the equivalent desktop application. Both developing forms have mature developing environments and are capable of achieving high user experiences. The difference between Web application development and desktop development is in their different domains that result in different user prerequisites, e.g. on the Web it is not the first released application that will succeed but the best application. The result of this paper indicates that Web applications are almost as capable as desktop applications when it comes to processing and interactive user interfaces. But that there are other user related factors separating their development such as the users' expectations and demands on the application. In order to get a more reliable comparison between Web applications and desktop applications research should be conducted with application case studies and user test groups.

References

[1] Inc Adobe Systems. Adobe flash player 10 security. White paper, Adobe Systems, Inc, November 2008. `http://www.adobe.com/devnet/flashplayer/articles/flash_player_10_security.pdf`, accessed 2009-05-06.

[2] Inc Adobe Systems. Swf file format specification version 10. Specification, Adobe Systems, Inc, November 2008. `http://www.adobe.com/devnet/swf/pdf/swf_file_format_spec_v10.pdf`, accessed 2009-05-07.

[3] Jeremy Allaire. Macromedia flash mx - a next-generation rich client. White paper, Macromedia, March 2002. Available online (14 pages).

[4] Erik Botsford. Developing rich web interfaces with flash. White paper, SBI, march 2003. `http://download.macromedia.com/pub/solutions/downloads/business/sbi_perspective.pdf`, accessed 2009-04-27.

[5] Alessandro Bozzon, Sara Comai, Piero Fraternali, and Giovanni Toffetti Carughi. Conceptual modeling and code generation for rich internet applications. In *ICWE '06: Proceedings of the 6th international conference on Web engineering*, pages 353–360, New York, NY, USA, 2006. ACM.

[6] Jim Conallen. *Building Web applications with UML*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2000.

[7] Yogesh Deshpande, San Murugesan, Athula Ginige, Steve Hansen, Daniel Schwabe, Martin Gaedke, and Bebo White. Web engineering. *J. Web Eng.*, 1(1):3–17, 2002.

[8] Mark Driver, Ray Valdes, and Gene Phifer. Applications are the next evolution of the web. Gartner, May 2005.

[9] Joshua Duhl. Rich internet applications. White paper, IDC, November 2003. Available online (33 pages).

[10] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext transfer protocol – http/1.1, June 1999. `http://tools.ietf.org/html/rfc2616`, accessed 2009-05-03.

[11] Piero Fraternali. Tools and approaches for developing data-intensive web applications: a survey. *ACM Comput. Surv.*, 31(3):227–263, 1999.

[12] Jesse James Garrett. Ajax: A new approach to web applications, 2005. `Http://www.adaptivepath.com/ideas/essays/archives/000385.php`, accessed 2009-02-25.

[13] Gerti Kappel, Birgit Pröll, Siegfried Reich, and Werner Retschitzegger, editors. *Web Engineering - The Discipline of Systematic Development of Web Applications*. John Wiley & Sons Ltd., England, 2006.

[14] Joey Lott, Darron Schall, and Keith Peters. *ActionScript 3.0 Cookbook*. O'Reilly Media, Inc., 2006.

[15] C. Mariage, J. Vanderdonckt, and C. Pribeanu. State of the art of web usability guidelines, 2004.

[16] Colin Moock. *Essential actionscript 3.0*. O'Reilly, 2007.

[17] San Murugesan, Yogesh Deshpande, Steve Hansen, and Athula Ginige. Web engineering: A new discipline for development of web-based systems. In *Proceedings of the First ICSE Workshop on Web Engineering*, 1999.

[18] Jakob Nielsen. Enhancing the explanatory power of usability heuristics. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 152–158, New York, NY, USA, 1994. ACM.

[19] Tom Noda and Shawn Helwig. Rich internet applications - technical comparison and case studies of ajax, flash, and java based ria. Technical Report UW E-Business Consortium, UW E-Business Institute, University of Wisconsin-Madison, 2005.

[20] Jeff Offutt. Quality attributes of web software applications. *IEEE Softw.*, 19(2):25–32, 2002.

[21] Cameron O'Rourke. A look at rich internet applications. *Oracle Magazine*, 18(4):59–60, 2004.

[22] Linda Dailey Paulson. Building rich web applications with ajax. *Computer*, 38(10):14–17, 2005.

[23] Christine Perfetti and Jared M. Spool. Macromedia flash: A new hope for web applications. White paper, User Interface Engineering, 2002. `http://www.uie.com/publications/whitepapers/FlashApplications.pdf`, accessed 2009-04-22.

[24] Herwig Rollett, Mathias Lux, Markus Strohmaier, Gisela Dosinger, and Klaus Tochtermann. The web 2.0 way of learning with technologies. *Int. J. Learn. Technol.*, 3(1):87–107, 2007.

[25] Mark S. Silver and Sidne G. Ward. Browser-based applications: positive transference or interference? In *WWW Alt. '04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 426–427, New York, NY, USA, 2004. ACM.

[26] William Stallings. *Operating Systems: Internals and Design Principles.* Prentice Hall, 2005.

[27] Millward Brown survey. Flash player penetration, December 2008. `http://www.adobe.com/products/player_census/flashplayer/`, accessed 2009-05-10.

[28] Michael M. Swift, Brian N. Bershad, and Henry M. Levy. Improving the reliability of commodity operating systems. In *SOSP '03: Proceedings of the nineteenth ACM symposium on Operating systems principles*, pages 207–222, New York, NY, USA, 2003. ACM.

[29] Andrew S. Tanenbaum and James R. Goodman. *Structured Computer Organization.* Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.

[30] Matias Urbieta, Gustavo Rossi, Jeronimo Ginzburg, and Daniel Schwabe. Designing the interface of rich internet applications. In *LA-WEB '07: Proceedings of the 2007 Latin American Web Conference*, pages 144–153, Washington, DC, USA, 2007. IEEE Computer Society.

[31] Carlos A Velasco, Dimitar Denev, Dirk Stegemann, and Yehya Mohamad. A web compliance engineering framework to support the development of accessible rich internet applications. In *W4A '08: Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, pages 45–49, New York, NY, USA, 2008. ACM.

[32] Aaron Weiss. Webos: say goodbye to desktop applications. *netWorker*, 9(4):18–26, 2005.

[33] Luke Wroblewski and Frank Ramirez. Web application solutions: A designer's guide. `http://www.lukew.com/resources/WebApplicationSolutions.pdf`, accessed 2009-05-02, April 2005.

[34] Luke Wroblewski and Esa M. Rantanen. Design considerations for web-based applications. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*, 2001.

A study and comparison between iTunes and Winamp

Jacob Jansson

Abstract: Winamp and iTunes are two Mp3playing softwares among many out on the market today. But they are two of the biggest one with millions of users [1][2]. With this many users usability in the interface design should play a vital role. Usability can be tested in many various ways but this report focus only on one. Heuristic evaluation which is a method to test the usability in an interface. Three heuristic evaluators have tested both softwares and given their evaluations and conclusions. iTunes did not have many usability issues as expected, only few minor problems. Winamp did have some problems in their interface which was full of information and buttons to the point of confusing. My conclusions is that iTunes has a big focus on new users while Winamp on the other hand focuses on more experienced users and if that is actually their goal they are doing a good job.

1 Introduction

iTunes and Winamp are two of the biggest audiosoftwares on the market. This study will test the usability in both softwares and maybe will give some insights to why they are the biggest on the market with millions of users[1][2]. The goal is to analyze the two softwares Winamp, iTunes and their Interfaces. The focus will lie in the usability of both programs. Usability in software interface design is very important when designing programs. Many times you have many different programs designed for doing the same exact thing. Winamp, iTunes, Windows mediaplayer and many more are different programs that have the same basic functions. The only thing that separates them is their interface design and some special functions specific for that program. Their graphic or their mapping are very important factors to what differs one program from the next when they have the same basic functions. Winamp and iTunes are two very successful Mp3playing softwares which mean they should have a very good usability and interface design. An analyze and a comparison should give a interesting insight into why they are successful and why some people chose one software over the other.

2 Background and history

Winamp was one of the worlds largest mp3playing software for computers. In recent time iTunes has come into the market very strong taking more and more of the market probably due to clever marketing and designing by apple.

2.1 Winamp

Winamp was first released in 1997 and was a simple software for playing music on a computer. The program basically included a menubar with only the most basic functions, play, open, stop, pause and unpause. Later that year Winamp 1 was released which featured more of a graphical interface. Buttons for various functions, a graphical spectrum analyzer and a volume slider that changed color depending on how high the volume was set. Winamp 2 was released in 1998 and at that time was the most downloaded software for windows. This version had a better looking graphical interface and had more functions like an equalizer and the possibility to add and install plug-in programs. In 2000 Winamp had more then 25 million registered users. Winamp 3 which was a big change to Winamp 2 had main problems which caused that version to fail completly. Many users did not like that the new version had no backward compability to previous skins and plug-ins and decided to go back to using Winamp 2. Winamp 5 was quickly made and had the best features from both Winamp 2 and Winamp 3. Winamp 5 came with more functions then just listening to music. With this new software you can rip music and even burn cd's and encode music to different formats. In 2006 Winamp had more then 57 million users and the software are constantly updating and getting better [1].

2.2 iTunes

iTunes is a digital media player introduced by Apple in 2001 and from the beginning only worked on Mac computers. It featured a playlist, burning of cd's and lots of other features. In 2003 Apple introduced iTunes store, and from there you could buy songs and albums online and download them directly home to your computer legally. Later that year Apple built in support for Microsoft windows 2000 and Windows XP for iTunes. Apple releases the ipod-series and later the iPhone and designs iTunes to be the the platform those devices need to communicate with a computer. In 2007 apple released a newsletter hinting they had 110 million active users for iTunes [3].

3 Usability

Usability is not easy to define. What good usability is and what to measure as usability differs from product to product. And what good usability is even differs from user to user. One user might find a product very easy to use while another user find it incomprehensible [6] [4]. ISO (The International Standards organization) define usability as the effectiveness, efficiency and satisfaction in which a user achieve a specific goal [2]. This definition implies that usability will change depending on task and depending on different users. A five component framework is given by Jordan to further explain usability [6].

- **Guessability**A measure of the cost for a user to perform a new task he has never performed before. The less time spent on the task and on errors the higher the guessability is. Two good examples is door handles or fire extinguishers which are products that need very hight guessability.

- **Learnability** This is more a measure of the cost for a user to reach a certain level of competence on a task rather then the time it takes for him to do the task for the first time. This component is more important when there is more time for training and the ability to understand a task the first time is less important. An extreme example could be a pilot learning to fly a plane.

- **Experianced User Performance** Refers to more experienced users performing tasks. The effectiveness for these users will probably not improve much so they will have a certain performance and efficiency for a certain task. Important for specialist in any area in society from driving a car to managing a powerplant.

- **System Potential** Is a measurement of the maximal performance of the system. Unlike the above system potential is a interface quality that has nothing to do with the interaction between system and user. Thus the system quality doesn't vary either over time or among different users. Dos prompt which we used before we had Windows can serve as an example. you have to write long system lines even though you have no problem remembering them the task is still tedious and time consuming.

- **Re-Usability** This component address the possible decline in user performance after a longer break away from a task. Forgetting how things were done or decline of efficiency. Softwares which you don't use regularly like excel for some users. If long enough time has passed since you last used the program you almost need to relearn the software all from scratch.

There are a number of ways to determine if a software has good Usability. In Jakob Nielsens article Usability inspection methods he gives examples on methods to inspect if a program has good usability [8].

- **Heuristic evaluation** Involves usability specialists evaluating elements with established usability principles.

- **Cognitive walkthroughs** Simulates a user´s problem solving process at each step trying to see if the user will use the correct action.

- **Formal usability inspections** A six step procedure with strict defined roles to combine a heuristic evaluation.

- **pluralistic walkthroughs** Meetings with people who goes through a task step by step and discuss every choice.

- **Feature inspections** Different lists of features that are usually used to complete different tasks. Then it´s checked for long sequences or difficult sequences that require more experienced users.

- **Consistency inspections** If a designer has made multiple projects he can analyze an interface to so if it is consistent with his other projects by comparison.

- **Standards inspections** An expert in a certain standard area inspects an interface design to determine its usability.

These are some of the various different tests that exists to test usability but of course there are more then these. To get the best result it is better to use several of these instead of just using one of them [8]. In this report I will only use one of these evaluation methods mostly because of the time limitation we have for this course.

3.1 Heuristic evaluation

Heuristic evaluation with three heuristic evaluators. All evaluators are regular computer users but are not experts in usability. One of the evaluators have been using iTunes for some time but are first time user of winamp. The second evaluator is the other way around a user of winamp but not iTunes. The third user have been using both programs for some time. Jakob Nielsens describes the best number of evaluators to be between three and six. Heuristic evaluation can be performed by a rather small number of evaluators and is quite time effective and not expensive which makes it a powerful evaluation tool. The heuristic evaluator is given some time with the software to get familiar with the most important functions. Then the evaluator shows the heuristic evaluator ten usability principles for interface design and explains what they mean. When the heuristic evaluator has familiarized himself with the principles he goes through the program and all its important functions again. He pauses from time to time and analyze the functions and the program against the given usability principles. Whenever a usability problem is discovered the evaluator makes notes about the problem. Heuristic evaluation method is a easy test to do because it is so open. All test subjects do not have to do exactly the same thing with the program because basically everyone is just looking for usability issues and can search for them in however way suits them best [5]. The ten usability guidelines [7] for interface design given by Jakob Nielsen is:

- **Visability of system status** The system shows the user and keeps him updated about the system status and what is going on with the system. Another word to describe this principle is good feedback.

- **Match between the system and the real world** The system should talk the right language to the user. A language that is easy to understand and are close to dialogues and sentences in the real world and not more like command lines which only a programmer would understand.

- **User control and freedom** The user should have the control and the freedom to do what he wants when he wants to do it. Not be angered because certain functions doesn't exist which he would have liked there.

- **Consistency and standards** The system should be consistent in its interface and known standards should be used. Do not have similar buttons that do different things and icons should follow international standards when they exist.

- **Error prevention** Design software so that errors as much as possible is avoided. Make it impossible for the user to cause errors himself and when errors do occur have smart error messages that inform the user what is wrong and help the user to correct the problem.

- **Recognition rather the recall** Minimize the amount of information the user needs to store in the memory and make that information visually available on the screen. Only to a certain degree of course because to much information works in the other direction and confuses the user. A user should not need to remember information from a former dialogue which is not visual in the current dialogue.

- **Flexibility and efficiency of use** The system must feel flexible and efficient to use for the user otherwise the user might get frustrated and think the system is to slow and inefficient.

- **Aesthetic and minimalist design** Information that is irrelevant or seldom used should not be visible. Accessible but not visible.

- **Help users recognize, diagnose, and recover from errors** If an error occurs the system should tell the user what is wrong and why and help the user to solve or get around the problem.

- **Help and documentation** If the help function is needed in a system it should be designed as good as possible. Easy to search for the problem you have and easy answers that list what you need to do to step by step to solve your problem.

4 Evaluating Winamp and iTunes

An heuristic evaluation was performed according to Jakob Nielsens guidelines. Three heuristic evaluators were involved in the testing and each tested both Winamp and iTunes. They were each given some time to interacting with the program and then given a list of usability principles that they tried to identify. Whenever a usability problem was found it was documented using paper and pencil writing down everything.

4.1 Winamp evaluation

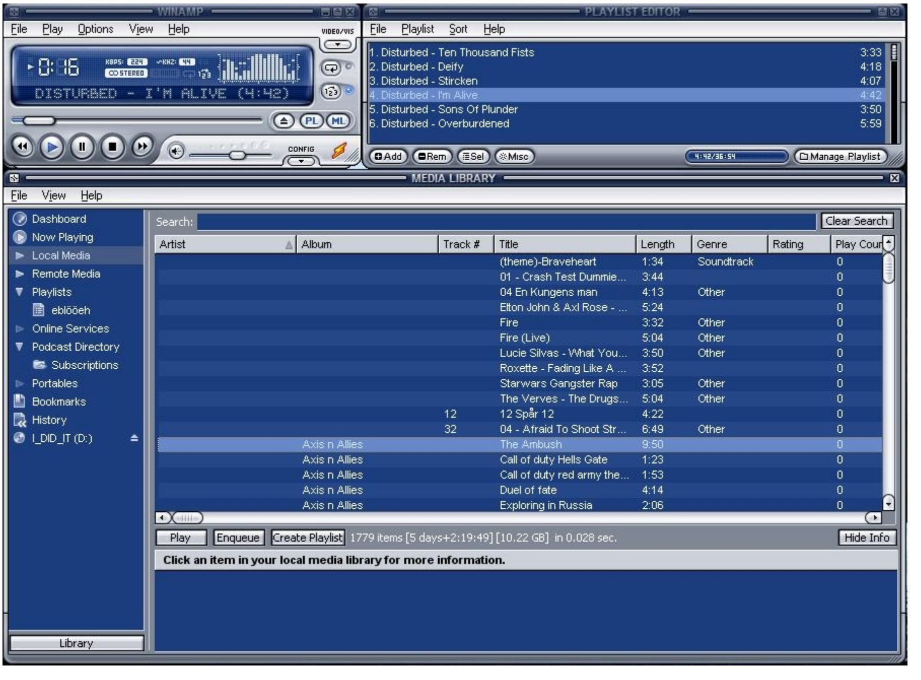

Figure 1: Print screen of winamp with all windows

The latest version of winamp was used which at that time was version 5.55 and the usability issues is listed below.

- **minimalistic design and standards** All three heuristic evaluators found this usability issue. Winamp has many windows which contains lots of information when you have just installed the program and the standard settings are set. Hard to know where to look or focus attention and hard to find specific buttons when needed to be found. The buttons don't always follow international standards so the button itself can be hard to understand. But two of the heuristic evaluators found out that winamp is very customizable and many of the windows can be shut down or customized to look just like you want. After some work with Winamp it was possible to minimize the software interface to a much better level.

- **Error prevention** Two of the heuristic evaluators found this problem. Most of the time in Winamp you have a temporary playlist where your music is listed and it is only saved when you manually go in and save the playlist. The problem is that the standard way of opening a new file will erase the temporary playlist and put only the newly opened file in a new temporary playlist. You can add new songs to your temporary playlist whithout erasing it but that requires you to find a small button almost hidden away at the bottom of your playlist. This problem caused some grief with both users when they from time to time erased all the songs they had carefully searched for and added to the playlist.

4.2 iTunes evaluation

Figure 2: Print screen of iTunes

The latest version of iTunes was also used which at that time was version 8.1 and the usability issues that was found is listed below.

- **Minimalistic design** Two of the heuristic evaluators found that in iTunes the media library had pre set playlists which were unnecessary, unwanted and the user were not given a choice to remove them.
- **User control and freedom** From now on there was only one of the heuristic evaluators who found the following usability issues. One found out that iTunes didn't have that good support for controlling the media with global hotkeys. He wanted to be able to fast forward in songs by only using the keys on the keyboard.
- **Minimalistic design and usercontrol/freedom** The interface in iTunes is very locked and there are limited possibilities to customize the interface to your liking.
- **User control and freedom** One user could not find the possibility to show only the albums instead of showing every single song and file you had in your library.
- **Help users recognize, diagnose, and recover from errors** When a song is moved on the computer the link to the media library is broken. When that song is attempted to be played a error message comes up saying that the song could not be found and advice the user to search for it. One user would have liked the software to help in this search. perhaps the program could search through the whole computer for the missing file if the user wanted that.

5 Comparison between Winamp and iTunes

From a usability point of view both these softwares have problems and issues. It will be hard to rate the usability problems and say which one is worse then the others. One way to rate them can be how many of the heuristic evaluators found that particular problem. If all heuristic evaluators found the same problem in one of the programs that is probably a bigger usability issue then when only one of the heuristic evaluators found a particular problem. Winamp didn't have as many usability problems as iTunes but the ones they did have seemed to be bigger and caused more grief. Winamp had so much info and so many windows and buttons that it is more suited for experienced users. Even an experienced computer user who just have never used that particular program but is used to many different programs and interfaces will find Winamp confusing when opening it for the first time. The program is not consistent, it has several menus in the different windows where all of them has a bar called file. There is two different play buttons located on different places one that is rectangular says play and the other is round and have the international standard play-icon.

Information like how long a certain song is can be found at three different places at the same time and the overall feeling is just confusing. iTunes did have usability issues also but the general feeling the heuristic evaluators had was that iTunes was the simpler program to use. Easier to navigate and understand, easier to get started with and more consistent.

Figure 3: Winamp without the library. The button which is marked by ML opens the media library again

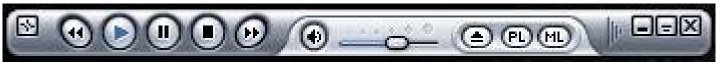

Figure 4: Winamp minimized to just one bar where only the most basic music controls is accessible. (This picture is enhanced, in its original scale it's as wide as Fig.3)

The biggest issue Winamp had could also be solved with all the built in customization functions. Theses are just two examples of how Winamp could look on the computer screen. The smaller the format gets the more functions also disappears until you just have the basic functions for playing music. This makes Winamp suitable for more experienced users and users who has the time to learn all the functions of the program and want to be able to customize the program to look like you want it to.

6 Discussion

I expected both of these programs to have very few usability problems. Both of them have had many versions before in which problems were found and corrected into the later versions. This is the way of many softwares out today. They are constantly upgraded and improved and then distributed easily through the Internet. a company can either have a department that sits and find these problems in their software or they let the public be the heuristic evaluators and then just correct the problems that the users find. I was surprised to find that Winamp got so poor grades from the Heuristic evaluators when they judged the usability of the software. If you look back on previous versions of Winamp it seems the software has just gotten more and more advanced but at the same time poorer and poorer in usability. Winamp was good in many other ways but usability was not one of the strongest aspects of the program.

Figure 5: Early version of Winamp

Winamp has gotten more advanced and has many more functions but I think the software developers in this case have forgotten about usability in their software design or chosen not to focus on that. I think the consequence of that will be that they will keep their users, who already are accustomed to Winamp, satisfied but they will not get many more users then they already have. If a new user install the latest Winamp, opens it and find all those windows and all those buttons I could imagine he would rather just unistall the program and get another media software then sit down and try to figure out how everything works.

Figure 6: Latest version of Winamp with another skin then Fig.1.

iTunes was easier to use but had less special functions. I think Apple have really focused on usability and want everyone to be able to use their software. The usability problems iTunes had was mostly specific to experienced users in nature. The ability to customize the interface to your liking or the ability to use hotkeys to fast forward in songs is not problems for the typical user but rather specific problems for more advanced users. The overall feeling i got when listening to the Heuristic evaluators was that there were no usability problems in iTunes that would deter a new user from starting to use iTunes. From a usability point of view I would say that iTunes is the better software.

References

[1] Aol unveils new winamp version with enhanced dashboard and remote features, 2006. [online] `http://www.timewarner.com/corp/newsroom/pr/0,20812,1539885,00.html`.

[2] 9241-11. Ergonomic requirements for office work with visual display terminals (vdts) - part 11 : Guidance on usability. 1998.

[3] apple news/media reports. Analyst on apple tv in digital living room, 2005. [online] `http://www.ipodnn.com/articles/07/03/16/piper.on.apple.tv`.

[4] Nigel Bevan, Jurek Kirakowski, and Jonathan Maissel. What is usability? 1991.

[5] Rolf Molich Jakob Nielsen. How to conduct a heuristic evaluation. 1990.

[6] Patrick W Jordan. *Contemporary Ergonomics*. 1994.

[7] Jakob Nielsen. Ten usability heuristics. 1994.

[8] Jakob Nielsen. Usability inspection methods. 1994.

Can tangible user interface reduce distractions while driving a car?

Christoffer Kopp

Abstract: Is new technology safer and better than old technology? In this paper the evolution of car dashboards will be discussed from a safety point of view. Car accidents are number nine on the list of what causes the most deaths throughout the world. So it is very important to continue to work for a safer car environment. One large cause of car accidents is driver inattention, such as focusing on the dashboard rather than on the road. Dashboards from the nineties and backward consisted of sliders and pinions for controlling the in-car temperature and fan-speed. This has almost disappeared in the dashboards of today, and been replaced by buttons and a display that provides feedback. The old type of dashboard that was more of a Tangible User Interface gives safer feedback and instant information of what kind of interaction is needed with the dashboard. This is safer than the new dashboards where the driver has to separate all the buttons that look and feel the same by focusing at the dashboard and thereby gets distracted.

1 Introduction

The cars of the 21st century have a dashboard that looks quite different from the dashboards from the nineties and back. A dashboard from the nineties and back only consisted of sliders and pinions. For example, when changing the in-car temperature, fan-speed or where to focus the fan, a pinion or a slider was used. The old type of dashboards did not have any displays that gave feedback, instead the slider or the pinion itself with some kind of indications around the slider or pinion provided the feedback (as seen in Fig. 2 a and 2 b). When comparing these older types of dashboards with the ones that many cars have today a big difference is noticeable. Many of the sliders and pinions from the old type of dashboard have almost disappeared (as seen in Fig. 3 a and 3 b). The sliders and pinions have been replaced with buttons that gives feedback on a small screen next to it. These differences will be addressed and discussed in section 3 “Dashboards; yesterday versus today”. The impact that car accidents has on the people around the world is hard to realize. Each year around 1.27 millions are killed in a car accident and up to 50 millions are injured each year [8, 3]. With every one of these accidents follows personal tragedies; someone might have lost their parents or been put into a wheel-chair for the rest of his/hers life. This high number of accidents also has a large impact on the world economy. According to the report of World Health Organization (WHO) the total cost of car accidents worldwide is up to USD 518 billion per year [8]. These numbers are predicted to rise with around 65 percent between 2000 and 2020 if nothing dramatically happens [8]. A major reason for car accidents is driver inattention [6]. According to a study made by the National Highway Traffic Safety Administration in the USA, at least 25 percent of all highway crashes involved some type of driver inattention [5]. These numbers shows that it is of great importance to reduce the distractions from the driver. This statistic is addressed in section 2 of this paper. Section 2 will also address different kinds of distractions that can occur when driving a car. Section 4 presents the Tangible User Interface along with its merits and demerits. Is the trend of adding more digital information and removing the sliders and pinions from the modern car dashboard a way in the wrong direction? Will this development increase the number of distractions and thereby the number of accidents? These questions will be discussed in section 5.

2 Different distractions

At least 25 percent of all accidents that occur in the United States are due to some type of driver inattention [5]. What different types of distractions are there and how common are they? This will be addressed in this section of the article.

2.1 Different distractions while driving a car

According to the National Highway Traffic Safety Administration, USA, there are four different kinds of driver distractions [10]:

1. Visual distraction: Visual distraction can be dived into three types. One where the visual field is blocked, a second where the driver is focusing on the dashboard and in-car techniques (for example adjusting the radio or adjusting the temperature [6]) and the last type is when a driver sees something on the road or next to it but does not understand it [10].

2. Auditory distraction: Auditory distraction is when a driver is focusing on sounds rather than driving. Like when listening to a mobile phone or the radio [10]. Salvucci [9], testing different types of phone dialing interfaces, proved that even if there is a voice interface the time of executing a phone call is longer when driving than standing still. This shows that the driver is distracted even when he only uses his voice to interact with the phone [9].

3. Biomechanical distraction: This kind of distraction happens when the driver is moving something in the car with one or two hands and thereby removes his/her hands from the steering wheel [10, 8, 3].

4. Cognitive distraction: This distraction occurs when the driver is thinking of something else which distracts the driver from the main goal, which is driving [10, 3]. The kind of distractions can be talking to someone in the car, talking on a cell phone or just thinking on something else [10, 3].

There have been numerous tests to see which of these distractions that has the greatest impact on the driver. Tasks where the drivers need visually attend (like looking at a display to see the temperature in the car) are not proven to give more distraction than auditory distraction, biomechanical distraction or cognitive distraction as one may think [3]. It is rather a combination of the different distractions that has the greatest impact on the driving [15, 3].

2.2 Statistic about car-crashes

The WHO (World Health Organization) during 2004 did a study to see what caused the most deaths throughout the world. At first place came coronary heart diseases (7.20 million deaths), at sixth place came HIV/AIDS (2.04 million deaths) and already at ninth place came road traffic accidents (1.27 million deaths) [8]. These numbers show the importance of improving the traffic environment. The most common reasons for car accidents in the USA that has a deadly outcome, is that the driver is under the influence of alcohol [3]. From the total numbers of accidents that had a deadly outcome at 2004, 39 percent of these accidents were from drunk-driving [3]. For Sweden these numbers where 11.9 percent [3].

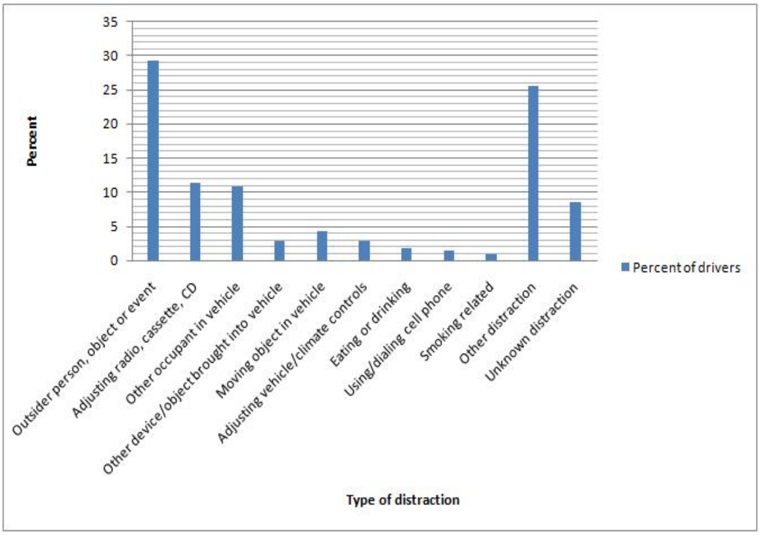

Figure 1: Type of distractions. Diagram made by numbers from [6].

Another big reason for accidents is driver distractions. The National Highway Traffic Safety Administration released in 2001 a report based on highway accidents in the United States of America. According to this report at least 25 percent of all car accidents involved some form of driver inattention [6]. The same report highlights the type of distraction the driver had at the time of the accident. The main reason for the inattention was "outside person, object or event", 29 percent. Some other reasons for the inattention are listed in Fig. 1. These numbers are from 5000 police reports in the United States from the years 1995-1999. Unfortunately there has not been another data gathering this broad since this period which could make the data a bit misleading, but it still gives a good hint of which distractions that are most common. According to a new study about cell phone use (in 2006) made by the AAA Foundation for Safety Traffic (in the United States) the share of accidents involving using a cell phone has increased to around 3.5 percent [1]. This study did not look at other distractions. Since the four kinds of driver distractions often overlap it is not possible to state which kind of distraction that causes the most accidents.

3 Dashboards; yesterday versus today

Dashboards from different decades and different models look quite different. What are the differences when it comes to temperature setting, fan speed and fan focus? Four different cars from three decades and in the same price-class (250.000 Swedish kroner and above) are compared according to the stated focus points. The cars in this analysis are an Alfa Romeo Spider -84 (Fig. 2 a), Range Rover -93 (Fig. 2 b), Saab 9-5 -02 (Fig. 3 a) and a Honda Accord -09 (Fig. 3 b). There is no idea to go further back in the history to find differences between dashboards since the technology for dashboards has not changed that much before the eighties. The relevant parts in each picture are highlighted with a frame. The older the cars get the more basic the dashboard looks like. The Alfa Rome Spider (Fig. 2 a) is limited to just two different sliders and a button for controlling the temperature settings and fan settings. When looking at a slightly younger car, the Range Rover (Fig. 2 b), it is noticeable that more sliders are added that gives the driver more options for controlling the heat and fan. There are in these two decades (the eighties and the nineties) sliders only for controlling the in-car climate. When looking at the next decade (the 21th century) (Fig. 3) the buttons have made their entrance on the dashboards and thereby replaced the sliders. The Saab 9-5 (Fig. 3 a) which is slightly older than the Honda (Fig. 3 b) has the feedback from the buttons for controlling the heat and the fan on a small display right above it. The newest car in the comparison, the Honda Accord (Fig. 3 b), just like the Saab (Fig. 3 a) has buttons for controlling the in-car climate. The difference is that in the Honda the feedback from the buttons is on a display quite far from it (The top square in Fig. 3 b). No one of the cars in Fig. 3 has any kind of physical feedback, like not being able to push the heat button any more, when it is not possible to increase the temperature after reaching the end of the temperature scale.

The only way to notice this is by looking at the temperature display. The older cars in Fig. 2 both have sliders for controlling the temperature and when reaching the end of the temperature scale they give a physical feedback. It is not possible to slide the slider any further at that direction. The summary from this comparison between four different cars in the same price-class gives a hint of the trend in the car industry. Cars from a higher price-class have almost all of them removed the sliders and replaced them with buttons. How the feedback from the buttons is presented is slightly different between different manufactures but the trend is that all of the feedback is presented on a mutual display.

(a) Alfa Romeo Spider (b) Range Rover

Figure 2: The dashboards of Alfa Romeo Spider -84 and RangeRover -93

(a) Saab 9-5 (b) Honda Accord

Figure 3: The dashboards of Saab 9-5 -02 and Honda Accord -09

4 Tangible user interface

The Tangible user interface (TUI) is a way to grasp and manipulate physical objects to change digital information. The physical object is also often a representation of the digital information manipulated [12]. The dashboards in the cars from the nineties and backward consisted of sliders and pinions, which are kinds of TUIs. Are these sliders and pinions a safer way to modify different settings than just by pressing a button like the modern dashboards? To answer this question it is important to understand the different merits and demerits of a tangible user interface. This and some background to what a TUI is will be discussed in the following sub-sections.

4.1 Background to Tangible user interface

For a very long time mankind have grasped and manipulated different physical object to build a rich picture about objects that is of great value in haptic interaction [5]. Much of this haptic perception has been lost when introducing digital technologies that only are maneuvered by a mouse or a button [5], so with the TUI Ishii and Ullmer wanted to rejoin the richness of the physical world in human computer interaction (HCI) [5]. Djajadiningrat et al. [2] states that when designing a TUI interface it is very important to think in the term of affordance. The word affordance was introduced to the interaction design process by Norman's book "The design of everyday things" [10] and the overall meaning of this word is that it should be possible to understand how to interact with the design by just looking at it [10]. Djajadiningrat et al. [2] say that when using Tangible user interface there is of great importance that the coupling between action and feedback is really strong. They list factors that are of great importance;

1. "Unity of location: the action of the user and the feedback of the product occur in the same location"[2].
2. "Unity of direction: the direction of the product's feedback is the same as the action of the user"[2].
3. "Unity of modality: the modality of the product's feedback is the same as the modality of the user's action"[2].
4. "Unity of time: the product's feedback and the user's action coincide in time"[2].

So, it is not only important to design buttons that are easy to understand but it is also of great importance to present feedback in a correct way.

4.2 Merits and demerits of Tangible User Interfaces

Like all types of different types of interfaces the TUI interface has several merits and demerits. The largest benefits can be summarized into one word;

affordance. A good TUI interface makes it possible for the user to know which kind of interaction is needed for executing a specific action, such as buttons need to be pushed, sliders need to be slided and so on [2]. A great TUI interface does not need to have written instructions next to it since it is obvious how to interact with the product [2]. By not having to read or think about how to interact with the interface the user can stay focused on more important things, like on the road. Another benefit with the TUI interface is that it can use the knowledge and experience of the user to create interaction devices that are very easy to understand [2]. Unfortunately there are also some drawbacks with the TUI interface; if it is implemented poorly the physical object may not indicate the right functionality which makes the action difficult to understand and execute [2]. Another issue in the TUI is that it relies on natural mapping for creating meaningful couplings between form and function [2]. This is a problem since not all virtual data has a physical counterpart [2]s. Users preconceptions about an interface may cause problems when interacting with the TUI interface. This is because people are very familiar with the normal GUI interfaces and when interacting with the TUI they try to do it as if it was a normal GUI interface and not as they would do if it was not a computer [5]. If the TUI interface has a lot of buttons and sliders that look the same it may be hard to know which one that does what without reading the instructions, which may take away the focus on the task performed [2]. That is why the designer should design the buttons and sliders so that they communicate their purpose by its form and actions [7]. To summarize the merits and demerits of the TUI interface it is of great importance to design the interface in such a way that it is really easy to understand how each slider and button works, and what digital object they affect. If the interface contains of many buttons and sliders they must differ in the way they look so no misunderstandings can occur.

4.3 Token + constraints

In 2005 Ullmer and Ishii introduced a new kind of TUI interface that could reduce the drivers' risk of inattention when interacting with the dashboard. This interface is called token and constraints [13]. The token and constraint interface is built by different tokens that can be manipulated in ways controlled by a constrainer. Each token is a physical object that can be moved in different ways in order to access or manipulate digital information [13]. Figure 6 shows an example of a token and a constrainer, the black square is the token and the grey slide surface is the constrainer. The constrainer indicates by its form that the token can be moved either right or left, and not up or down.

This kind of interface, token and constraints, has numerous benefits as compared to a regular interface. Ullmer and Ishii list these benefits as follows in their article "Token+Constraint Systems for Tangible Interaction with Digital Information" [13];

1. "Increased passive haptic feedback";

2. “Increased prospects for active force feedback”;
3. “Decreased demands for visual attention”;
4. “Increased prospects for embedded uses”;
5. “Flexible, widely accessible sensing technologies”;

These are some of the benefits that make tokens and constrains really good when added to an in-car dashboard. They will reduce some of the visual distractions that a driver can experience while driving since the driver does not need to look at the object he/she is manipulating. The interface also provides haptic feedback which is good since the driver can feel when he/her reaches the end of the constrainer, for example when adjusting the heat.

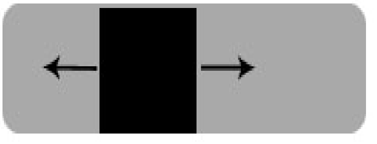

Figure 4: Token + Constraint

5 Discussion and Conclusion

According to the NHTSA report 25 percent of all car accident in USA where due to distracted drivers [6]. If these numbers are general for all countries they cost the world economy around USD 130 billion each year [8]. To prevent these accidents is it important to reduce the type of distraction objects in the in-car environment. The trend that more buttons are added to the dashboard with feedback at a different place, and that sliders and pinions is removed gives a worse distraction situation. Without the constraints that a slider provides the driver must focus on the button and display to know which temperature, fan speed or fan focus is chosen. For example in Fig. 2 a the in-car temperature is controlled by a slider that has a red mark that indicates warm at one end and a blue mark that indicates cold at the other end of the constraint. These marks are already well known to everyone and when moving the slider to one end the driver knows and feels that it is not possible to get further on the scale, a perfect example of a token and a constraint. When comparing this to the new car, Fig. 3 b, the driver must look at the display to realize when he/her has reached the end of the scale which decreases the drives attention on the main focus, the road. Since the last gathering of statistic regarding driver distractions is from the years 1995-1999 it is hard to say whether the new kind of dashboard is worse than the older ones, from a distraction point of view. But analyzing the merits and demerits from section four (“Tangible User Interface”) the new kind of dashboard with much more buttons and less sliders and pinions should increase the driver distraction. This is because the many buttons that looks similar forces the driver to look and read at the button to know which button affects what. A dashboard with many buttons that do not have any constrains that provide feedback also force the driver to look at some display to know which setting that has been chosen or when reaching the end of the scale. For example when adjusting the in-car heat, an old dashboard with a pinion is not able to turn any more when reaching the end of the temperature scale. But with the new dashboard that mostly has a button for changing the temperature there is no way to know when reaching the end of the temperature scale without looking at the display connected to the button. This is one of many examples of when an older type of interaction type is better than a new one. In new cars there are a lot more technology than in older ones. Because of this there need to be a lot more buttons on the dashboards. More buttons take more room and distract the driver even more. Is all this technology really important? If the designers remove some of the technology away from the driver there will be room for sliders and pinions. By removing some of the technology there should also be less distraction for the driver in the in-car environment, thereby reducing the number of accidents that is caused by driver distractions. From the literature study one could say with some certainly that more buttons and less sliders and pinions on a dashboard would increase the drivers distractions and thereby the number of car accidents. So to go back to the era of sliders and pinions may very well be a better way from a safety point of view.

References

[1] AAAFoundation.org. Cell phones and driving: Research update. Technical report, AAAFoundation.org, 2008.

[2] T. Djajadiningrat, S. Wenseen, J. Frens, and K. Overbeeke. Tangible products: redressing the balance between appearance and action. *Resonal Ubiquit computing 8*, 2004.

[3] Geneva Economic Commission for Europe. Statistic of road traffic accidents in europe and north america. Technical report, United Nations, 2007.

[4] C. Franzén. *Trafikskadades erfarenheter av vård och kostnadseffektivt stöd.* PhD thesis, Umeå University, 2008.

[5] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. Technical report, MIT Media Laboratory. Tangible Media Group, 1997.

[6] L.Staplin, J. C. Stutts, D. W. Reinfurt, and E. A. Rodgman. *The role of driver distraction in traffic crashes.* PhD thesis, University of North Carolina, 2001.

[7] D. Norman. *The design of everyday things.* The MIT Press, 1998.

[8] M. Peden, R.Scurfield, D. Sleet, D.Mihan, A.A Hyder, E. Jarawan, and C. Mathers. World report on road traffic injury prevention. *World Health Organization*, 2004.

[9] D. D. Salvucci. Prediction the effect of in-car interfaces on driver behavior using a cognitive architecture. *Nissan Cambridge Basic Research For Cambridge Center*, 2001.

[10] P.O. Sandberg and E. Eriksson. Advantages and challenges of introducing limited internet access in a truck environment: Focus on an interactive user manual. Master's thesis, Umeå University, 2009.

[11] O. Shaher, N. Leland, E. H. Calcillo-Game, and R. J. K. JAcob. The tac paradigm: Specifying tangible user interfaces. *Personal and Ubiquitous Computing*, 8(5):359–369, 2004.

[12] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *IBM System Journal, vol 39*, 2000.

[13] B. Ullmer, H. Ishii, and R. J. K. Jacob. Token + constraint system for tangible interaction with digital information. *ACM Transactions on Computer - Human interaction*, 12:81–118, 2005.

[14] K. Young and M. Reaga. Driver distraction: A review of the literaturel. Technical report, Australian College of Road Safety, 2007.

[15] K. L. Young, M. A. Regan, and M. Hammer. Driver distraction: A review of the literature. Technical Report 206, Holden Ltd, November 2003.

The Possibilities of Using Electronic Paper in the University Environment

Tea Meriluoto

Abstract: Every year several tons of paper are used at universities around the world. Regardless of the size of the university, there are vast amounts of documents printed out. We propose that electronic paper technology is the key to stop the world from drowning in paper waste. With the introduction of digital reader devices in educational programs and development of electronic paper displays, campuses could be turned if not totally paperless, then at least into using less paper. What are the benefits of electronic paper in university setting? How large of an environmental effect could these technologies have and how widely could they replace physical paper? This paper introduces some possibilities and examples of electronic paper usage in university environment with a literature study of the existing technologies and their applications.

1 Introduction

Paper production is a large source of CO2 emissions [2] and paper products represent one third of the waste handled yearly in the industrial countries such as the USA [3]. It has entered common knowledge that we need to find sustainable ways of operating in order to avoid the environmental crash course we are following with the recent exploitation of natural resources and the ravaging consumption habits. Universities, with their status as havens for young and knowledge-seeking minds, should take the lead in reducing the use of physical paper. They should employ the existing resources as well as be pioneers in finding and developing new solutions. Electronic books and digital readers have been on the market trying to break through for some time now, but the readability as well as usability of most of the existing displays is strongly criticized by users and media. In our opinion, a far greater potential can be seen in a book of tomorrow: electronic paper sheets that are bound together with a spine holding enough memory to transform the empty pages to any literature work in history. What are the benefits of electronic paper in university setting? How much environmental effect could these technologies have and how widely could they replace physical paper? This article studies the possibilities of electronic paper bringing us a step closer to a paperless - and therefore more environmental sound - campus.

Universities state in their environmental plans their goals of reducing the use of paper. This usually implies not giving up of paper but the use of recycled paper, which is a step into the right direction but there are more effective means [2]. With the unlimited information resources at our hands and the reading preferences of people, the paper usage is hardly diminishing. Instead of supporting the production of more paper - be it recycled kind and therefore more environmentally sound - the focus should lie on switching the media. The information in most cases is already digital and therefore instead of converting it to a physical paper, the opportunities of digital ways should be embraced and the user should be introduced to do more in this medium. This requires right tools, and electronic paper has in our opinion a great potential to be one.

Section 2 introduces the concept of electronic paper and discusses the existing techniques as well as applications of electronic paper. The affordances and functionality is also presented as well as the environmental and ergonomic aspects. In section 3 we explore how physical paper is used on campus, why people like to use paper for reading and documenting, as well as some examples about the amount of waste this usage causes in two specific campuses. The possibilities of electronic paper usage on campus are presented in section 4 with three examples that have been tried out on different campuses.

2 Electronic Paper

The electronic paper is an electronic display that mimics and exceeds the qualities of normal paper [7]. Ideally it is as thin and bendable and shares the light reflecting qualities of normal paper which makes it readable even in direct sunlight - an impossibility to almost any laptop or PDA display on the market today. Whereas normal paper has static content or requires manual manipulation with pen and eraser to change the contents, a sheet of electronic paper can change its contents several times in a fraction of a second. Also due to the plastic character of electronic paper, it does not wear as the traditional paper does. All it requires is a little electric charge and a processor to reorganize its pixels and the transformation is complete.

Mann (2001) describes the ideal electronic book looking like a regular book consisting of electronic paper sheets connected to a processor and memory in the spine [7]. This book of the future can contain a whole library and enables functions such as searching in and among documents, annotating and browsing through the library of documents. There are virtually no limitations to the possible contents of an electronic paper device because with the help of network connection the whole of Internet could be accessed.

Electronic paper is not a new concept and when mentioning it, many people tend to think of the flop of digital readers when they first were introduced to the market in the turn of the millenium. Although the idea of the electronic paper display is good and has full potential to be used as a tool to replace many unnecessary uses of traditional paper, the early solutions - the digital readers - were not nearly good enough an alternative to replace paper, as Sellen and Harper (2003) point out in their book "The myth of the paperless office" [14]. In order to succeed in designing the book of the future, it is of a great importance to identify the aspects of paper usage that people are accustomed to, as well as the advantages and shortcomings of it. Only based on these findings a "better" substitute can be created. What is needed is a solution that addresses the shortcomings of physical paper without neglecting the existing benefits.

2.1 Creation of Electronic Paper

According to Mann (2001), the inventor of the electronic paper is Nick Sheridon [7]. He is a physicist and researcher and when he joined the Xerox's Palo Alto Research Center (PARC) in 1975, he noticed the paradox in his colleagues' dream of one day replacing printed books and magazines with computer displays. He came up with the idea of doing it the other way around: replacing the display with a paper - enhanced one of course. He designed the Gyricon, a transparent and thin silicone-rubber sheet containing tiny solid bi-polar plastic spheres which with the help of electric charge can view black-and-white images. Unfortunately this technology was very expensive and as Xerox judged it to be a no-go, he was pulled away to other projects in 1977.

In the mid-90's Joseph Jacobson from MIT duplicated Sheridon's work with two of his students and they improved it by replacing the little spheres with hollow capsules that contained colored oil and tiny electrically charged chips of paint. These capsules react to electric current near the sheet and act like pixels on a monitor. As a result of this success they founded the first company to commercialize electronic paper, the E Ink. Their product in 1999, a display sign that could show dynamic text, had unfortunately a vital shortcoming in order to be the future of electronic paper: due to their technology it was not bendable. At the same time even Xerox had regained an interest to this field of research and Sheridon got to continue his work. He got closer to acquiring the affordances of paper as he managed to manufacture large quantities of flexible Gyricon sheets [11] together with the plastic product giant 3M.

By the development of electrically conducting plastic, there was finally means for the electronics that were critical to making the electronic paper. Even though it is not the best conductor, plastic is fast enough for most applications, as well as being lighter, cheaper and more flexible than silicone or copper that were used before. Lightness, inexpensiveness and flexibility being the key attributes that make physical paper so usable, this implied a new era for electronic paper techniques. In the next section we introduce three of the most used ones.

2.2 Technologies

As stated before, the research and development of electronic paper displays has existed for some decades now and in the search for the most optimal solution it has already resulted in various different techniques. The race to develop the optimal paper-like display, enhanced with the possibilities the electronic nature of it brings, is only accelerating. Electrophoretic, electro-wetted and liquid-chrystal displays are some of the most researched new techniques, and they are described in turn, as well as one interesting example application of the last one.

Electrophoretic displays, used typically for digital readers on the market, for example E Ink display in iRex Digital Reader and Amazon Kindle. This technology is based on the early design of Sheridan as well as the slightly more developed one by Jacobson (see section 2.1). According to Rogers et al (2001) it is an attractive technology due to the low-cost processing, mechanical flexibility and large area coverage characteristics. They report developing organic active matrix backplane circuits that form mechanically flexible sheets of electronic paper [13]. Electrophoretic display contains charged pigment particles that are suspended in a liquid between two parallel conductive plates. These particles act as pixels. By applying voltage to appropriate regions of the display, patterns and figures can be formed on the surface, as this makes certain sections reflecting and others absorbing light.

Another method, called Electro-wetting has been proved to be an attractive technology. Its main benefit is its possibilities of rapid manipulation of liquids on micrometrical scale, as the displays are able to switch from a white to a coloured pixel very fast. This opens up the possibilities to stream video on the electronic paper. Pixels are formed by the little droplets of colored oil in a liquid, which form a flat film between a water surface and a hydrophobic surface, that depending on the voltage applied contracts and subtracts. For example, without a voltage, a colourful homogenous oil film is presented, but when voltage is applied, the oil film is contracted and the pixel is hidden. Hayes and Feenstra (2003) show that electro-wetting enables reflective display technologies significantly faster than electrophoretic displays [5]. Electro-wetted display is superior in brightness to reflective liquid-chrystal displays, about four times, and when compared to other emerging technologies the result is twice as bright.

Liquid-chrystal display (LCD) technologies have also been emerging. This technology is common as a television display. A kind of LCD display was even created by Mary Lou Jepsen, for the XO-1 computer of the One Laptop Per Child (OLPC) project. For the purposes of the project Jepsen was to come around the huge cost of a laptop display, and therefore she decided to design the laptop from the premises of the display [17]. The OLPC organization aims to help and inspire the children in schools in the developing countries to learn more. The organization has built the OLPC XO-1 laptop with power-saving, networking and tolerance in mind. The display of a XO-1 laptop actually accomplishes many features wanted for electronic paper, the cutting-edge display technology that is designed in the spirit of digital reader displays and could be future technology of electronic paper sheets. In the interview with Stanik (2007), Jepsen explains that for the OLPC XO-1display she has used a transreflective process in which she has added a tiny color filter over a part of the pixel [17]. In this way she has changed the bulk layout of the pixel, resulting in a one-third resolution of colour. As human visual system is analog and biological instead of digital, in this pixel layout human eye can perceive a high resolution of about 800 by 600 in color, which is on the maximum limit to what an eye can perceive. According to Jepsen it is a low-cost and low power consuming technology that enables great reflection capabilities and therefore the display is easy to read even in direct sunlight.

2.3 Affordances of Electronic Paper

Ideally the electronic paper would posses all the positive features of paper but also embrace the features enabled by it being electronic. In a presentation of their Gyricon project, PARC formulates the possibilities of electronic paper as following: "[an] electronic reusable paper display could be very thin and flexible. A collection of these displays could be bound into an electronic book. With the appropriate electronics stored in the spine of the book, pages could be updated at will to display different content" [11]. In our opinion this is a quite exact description of acquiring the affordances of physical paper in an electronic paper device. In order to be a success, electronic paper has to be all of the above. Most importantly it has to be portable, which implies the weight of the device has to be very low, the size not exceeding much that of a sheet of paper. The power consumption must be low in order to be able to keep using it for a long time without having to carry along a charger. It has to be readable anywhere and anytime, even in direct sunlight, just like traditional paper.

Even though electronic paper has to overcome the challenge of power consumption and long lasting battery time, it has possibilities beyond the physical paper, such as possibly being backlit in order to enable reading in the dark. It has to afford for spreading out and organising just like traditional paper, for example the pages of the future book could be removable. It has to afford for browsing through various pages, bending and rolling. Annotating and highlighting has to be easy and effortless, just like marking pages of paper with a pen.

Traditional paper documents and books being not interactive or integratabtle with each other, leads to a new affordance problem that arises first with electronic paper: the compatibility of the data formats. This is not a new problem for computer field, as it has been a challenge with nearly any kind of digital data software, and it definitely has been a problem with the past electronic readers. The different brands producing readers were offering formats compatible with their signature software and no standard was established. In order to be successful, either a standard must be found or the electronic paper has to be compatible with all document formats.

Aliakseyeu et al (2004) have developed virtual electronic paper that is used in their Visual Interaction Platform (VIP). Their electronic paper prototype is able to "help in managing, storing and annotating images; managing, creating and editing sketches and can assist in re-drawing and over-drawing" [1]. Doing this is something that electronic paper should afford. The challenge comes from the fact that Aliakseyeu and his group do not have physical electronic paper sheets but virtual sheets that act just as electronic paper in a large tablet display. This solution is neither portable nor does it address the other aspects that the physical paper does, but in our opinion it does show an interesting vision of digitalizing paper as well as the manipulation of the resulting virtual paper.

So far much of the criticism faced by the existing electronic readers has been based on the fact that they do not afford browsing through pages as can be done with a physical book [14]. Obviously the above mentioned VIP by Aliakseyeu et al has the same shortcoming. Different ways to read a book depend on the type of the book: a novel is read from a cover to cover but a textbook might be scanned through or sought for facts only in no particular linear order. Therefore a successful application of electronic paper must contain several sheets of electronic paper that can be browsed through just like the sheets of a physical book. The optimal amount of these sheets in an electronic paper device is to be studied in order to find the best solution.

An affordance of the electronic paper device is that it can contain a whole library of books and documents, which forms a challenge. Comparing information between several books and other documents is a common method of studying and researching. In order to work efficiently with these information sources, they are often spread on a table. Therefore electronic paper application should contain removable sheets of electronic paper in order to claim this affordance. In our opinion it is a vital one for the usability of the electronic paper.

2.4 Ergonomics of Electronic Paper

Wyatt et al (2006) propose that mobile devices can increase chances to injury due to the very portability of them. When it comes to laptops, they claim that the constraints, such as the inseparability of keyboard and display, increase the ergonomic problems because the portability of the device encourages the user to skip the traditional work settings [18]. According to Wyatt et al these non-ergonomic use positions of portable computers are related to Cumulative Trauma Disorders (CTD's). What they imply is that the work on computers is not conducted anymore on the ergonomic office desk set, but instead the mobility of the device lets the users to work anywhere, in this way forgetting about the ergonomy and proper working position, which might lead to physical aches and pain. The good thing they point out about the new and emerging display technologies on the other hand is that they are minimising the occurrence of Computer Vision Syndrome (CVS), as the displays are more eye-friendly.

A study by Shieh and Lee (2007) suggest that the optimal reading distance for the electronic paper display is 500 mm, which is a slightly longer than the recommended distance of 360 mm of traditional paper but is shorter than the 562 mm distance recommended for cathode ray tube (CRT) display [15]. Shieh and Lee even suggested the optimal reading angle to be from 120 to 125 degrees horizontally, or about 30 degrees below the line of sight. According to them the most important factor in defining the reading distance was the size of the characters and not so much the light conditions or the type of the electronic paper. They do suggest that light conditions did, on the other hand, significantly affect the angle of reading.

2.5 Environmental Aspects of Electronic Paper

While electronic paper can help tackle the problem of drowning the world in paper, this technology does have its drawbacks on nature as well. Being electronic the name itself suggests there must be some energy source involved and obviously producing such a device does as well require energy. These consumption figures can anyhow be reduced to minimal as the new technologies such as the electro-wetting enables electronic paper to display its contents without any additional energy and draw energy only while changing the image, which happens in just some milliseconds. An environmental study by Counsell and Allwood (2007) suggest anyhow that switching into electronic paper could reduce both the amount of waste and carbon dioxide emissions significally [2] even taking into consideration the grid electricity to power the use of electronic paper as well as the cost of producing the device (more in section 3.2). An aspect less studied is the waste caused by the electronic paper devices after the use, or the possibilities of recycling the materials. This is an environmental challenge in the design of electronic paper devices.

2.6 Existing Applications

So far electronic paper has been considered mostly as a type of display technique applied on different kind of electronic reader devices for electronic magazines and books, such as Amazon's Kindle and iRex Digital Reader. There are no devices on the market, that would use electronic paper sheets bound together to mimic a book in the sense that is idealized in this paper. Other applications of the technology are digital photo frames, dynamic price tags and and even some watches, for example by Epson Seiko. The electronic paper display technology is also attractive for dynamic signs and advertisements, especially because it is economical even for large sizes.

In our opinion the reason behind the small amount of - and the lack of a break through of - the products imitating the physical books, is the lack of a usable electronic paper display that is paper- and book-like. Another important factor is that the technology is commercially quite new and so far all the companies trying to use it are making their own specialized products, which leads to a lack of standards and common protocols in the field.

3 Physical Paper on Campus

Physical paper is met everywhere on the campus settings: posters on the walls, piles of papers in the offices, lecture handouts, books, journals, exam papers, registration and other administrative documents, there are printers and photocopiers around, napkins in the cafeteria, toilet paper in the bathrooms, flyers, even the paper reduction policy is printed on a paper. Although it is plain that digitalization of the usage of toilet paper as well as paper napkins is impossible and that a remarkable reduction in the consumption in this kind of use of paper is hard to achieve, the reduction in uses of print paper is possible and in everyones best interest. In order to find the ways to shift the usage of print paper to another media, we firstly need to understand the reasons why paper is used in the first place. Secondly, we will investigate some of the environmental aspects of the physical paper and the amounts of it used at two Swedish campuses.

3.1 Affordances of Physical Paper

Why is paper so widely used and why do we prefer to work with traditional paper? Sellen and Harper(2003), as well as Liu and Stork (2000), say it is due to its affordances [14, 6]. First of all traditional paper is very readable and easy on eyes, no matter in which light conditions it is used - as long as there is light. Secondly, paper affords bending and folding and sheets of paper can be separated, spread and organized. Paper is lightweight - at least as long as we are not talking about great amounts of it, and it can be bound together in various ways. Paper affords annotating and marking although it might be hard or impossible to remove the contents from its surface. When printed or drawn on, paper does not require any batteries or electrical power to keep displaying the contents but stay usable until it is damaged or demolished in other ways. The print can fade if exposed for sunlight, but in the dark of museums and libraries ancient parchments dating thousands of years back can be found in readable although often very fragile shape.

The downsides of paper should not be neglected either: paper can be torn, the print on paper might be hard or impossible to remove or manipulate, paper can only have one content and if erased it wears. Searching contents of paper has to be done manually and in large volumes it can be a burdensome task. Merging various paper documents together will involve clipping and glueing or scanning and reprinting (if available) but in this case the document turns electronic in the process. Paper is very sensitive to moist and is flammable, and in this way very fragile. Even though one sheet of paper is very light, a book consisting of great many sheets of paper can be very heavy - not to mention the weight of all the books in a library. This is due to the fact that a sheet of paper can only have a limited amount of contents, which leads to that the possible contents on physical paper are proportional to the volume of the paper(s). We call this the volume/contents ratio of physical paper. It is very hard to define the proportions of this ratio due to the fact that the quality of the content can vary greatly due to text size, information contents of the text and other similar factors.

3.2 Environmental Aspects of Physical Paper on Campus

Let us start with some examples. Luleå University with 12'250 students and 1400 staff members reported the amount of print paper used annually to be 30 million sheets or 150 tons [4] whereas the annual figure for the University of Gothenburg with 50'000 students and 2050 staff members was 396 tons [10]. Both universities claim to use printing paper that is at least 30 percent recycled, but according to the study by Counsell and Allwood (2007), using recycled paper can only reduce the energy consumption related to paper production by 35-49 percent depending on to what extent the paper is recycled as well as if the recycled pulp has been transported [2]. The same study proposes that by switching to electronic paper alternatives, the energy savings can be up to 65 percent with for example a display from E Ink. These numbers apply per a tonne of paper and therefore could make a huge impact with the amounts, 150 and 396 tons, annually.

When considering the potential reductions in climate change gases emitted per tonne of office paper, a difference can be seen between the two strategies. Whereas recycling can cut the CO2 emissions with 76 percent, the figure for emission savings in electronic paper usage is as high as 85 percent. These numbers are based on the standard emission factor for the UK which is 46 kg CO2e/GJ [2]. If these two strategies - printing on recycled paper when printing is assessed to be sensible and switching the use and manipulation of paper to electronic devices in all the other cases - are combined, then the universities boosting with green policies will get a whole new credibility.

4 Possibilities for Educational Use of Electronic Paper

The applications of digital documents and electronic paper on campus are many. In this section three cases are presented: a prototype of an electronic textbook reader, an electronic campus project where these readers could be used and a case of cheating prevention measures with electronic exams.

4.1 eTextReader - an electronic textbook of the future?

McFall (2005) suggests that there are great potential in electronic textbooks, probably the highest potential in all of electronic books, just as long as these textbooks are designed to be electronic and interactive, instead of just copying paper books [8]. In this paper McFall presents his prototype of electronic textbook, the eTextReader, which he has designed with the shortcomings of other textbook reader devices in mind. The eTextReader makes the learning experience highly interactive and adresses much of the potential of an electronic textbook. The functions include possibility not to only annotate and highlight, but also to un-highlight, to highlight in different colors and to group and organize annotations of an electronic document. It uses the possibility of viewing layers where the user can view only the original document or one of the customized layers with certain or all annotations. According to McFall this was a very attractive feature to students. The biggest impact on classroom learning according to him is the possibility of collaborative reading: the software enables students to share annotations, notes and diagrams they have made with classmates.

The eTextReader brings a whole new feature to classroom sessions: individual and anonymous interaction with the lecturer as well as with rest of the class. McFall points out the importance of students asking questions and the lecturer being aware of the shortcomings in understanding the study contents by the students. Because asking and pointing out problem issues of understanding in front of the whole class can be an unescapable object for some students, with the help of eTextReader these questions can be pointed anonymously to the teacher in the way that he can see them directly on his reader device and therefore can then address the issues better [8]. McFall points out that the probability of a student taking the step and asking the questions or pointing the weaknesses of understanding is greater when he can do it while reading the text and therefore instantly when finding such an issue. According to the results of evaluation of their performance and the usability issues of the reader device, students however pointed out that many of the features were hard to use and therefore did not consider the device to better their study experience.

With the findings of the first test as well as the potential of his software in mind, McFall developed eTextReader further and a new study was conducted just tree years later. In this new study [9] McFall et al (2006) presents some better user gradings. Due to improved technology for the tablet computer used to run the software, most of the usability issues were addressed. This time students found eTextReader useful and gave positive feedback on the collaboration features as well as the possibility of bookmarking and highlighting features. A highly used feature by the students was the text search. Another aspect a student pointed out was that although he never before did highlight or annotate on his expensive textbooks because he wanted to be able to sell them forwards after the course, he now had started using both of the features in the digital textbook and found it very helpful. McFall pointed out some advantages in the teachers aspect to students handing in assignments via the reader. First of all no paper was involved. The second advantage was learning related: the assignments were given to the students in the text in places where the issues were appropriate. Thirdly, instead of giving a separate text sheet of questions, the student had to find the questions in the textbook, which was thought to increase the students probability of reading the whole text.

4.2 Northwestern - a Case Study of Electronic Campus

Not only is there room for specific design for textbooks, but even attitudes and overall operations of a campus can be moulded to welcome the electronic change. In United States of America, the Northwest Missouri State University, in order to comply with the emerging information-driven society, began its Electronic Campus project as early as in 1987. According to Rickman et al (2003), the project did not only aim to welcome the technology and prepare the facilities and infrastructure to the technical implementations, but "has evolved from an academic system with high capacity communication structure to a complex educational delivery system" [12]. They state that the electronic means are part of daily life at campus and enhance the learning experience. What Northwest did was to provide the students as well as faculty and staff early on access to computing and they digitalized much of the information channels on campus. By the beginning of the 90's the 6000 students of the university had access to more than 3000 computers. Also a project to test laptop usage in a course work was launched, but it fell on economical grounds as there was too little intrest from the students to buy the equipment.

None the less, quite soon most of the coursework that was digitally enhanced was completely online, and Rickman et al (2003) note that the university offered 42 online courses as well as four online degrees [12]. In order to make sure that the students could make the most out of the electronic resources at the campus, the university provided a compulsory course for the first year students to introduce them to the system and give them insight to microcomputing and software packages. Unfortunately although this course was highly appreciated because it prepared the new students to better meet the electronic campus, it was removed in 2002 in order to reduce the general requirements of graduating in the effort to attract more transfer students [12].

It would have been interesting to investigate the amount of print paper consumed at what is called an electronic campus environment and to compare it to some typical universities with no similar projects. Unfortunately, the paper consumption figures for Northwestern were not available. Nevertheless we think that the they have set a good example for other universities in the work for a study environment with less paper.

4.3 Electronic Exams

When exams are conducted electronically, some serious consideration has to be directed to issues such as integrity of the students, security of the exam results and measures against possibility of cheating. If the exam is conducted on-site in a class room or similar with electronic readers provided by the examinator, possibilities of cheating will be very similar to cheating in a traditional pen-and-paper exam and therefore will not be a specific issue. On the other hand electronic examinations allow for distance exams and seem like a brilliant tool for distance education. When the teacher is not able to supervise the students, there is increasing risk for cheating.

Simon (2005) has been studying this and has tried electronic watermarks in exam papers on computer science to enable catching possible cases of cheating with electronic distance examinations. Simon says that the "additional data facilitates the detection of files which have a common ancestry, regardless of any subsequent editing of those files, and thereby significantly increases the amount of effort required for one student to plagiarise another's code" [16]. He implies that if students co-operate and for example copy and past code from one to another, the examinator will detect this afterwards as well as the unique identification tags will point out the collaborators. His anti-plagiarism strategies are applicable when a text editor or development environment is used to create program files. The students download a source code that they are expected to develop and which includes authentication and identification data invisible for the student. An example of identification data Simon used was using a different shade of ink for each paper.

Simon concludes that meanwhile the task of stopping students from cheating and detecting all the cheats is impossible, the knowledge that there are means for calling the cheats and the threat of getting caught is enough to scare most of the students to be honest. Also he thinks that when cheating is made complicated, most of the students will consider it easier to do a honest work.

5 Discussion

Even if the reality is against a totally paperless environment, less paper would be a step in the right direction. Some papers will and should stay, but in our opinion there is a lot of paper material that is unnecessary and especially paper printed only for reading purposes is such. By developing the display technology no paper would need to be printed just to be read and then thrown away.

While writing this article we tested an iRex digital reader. Even though in the beginning we were sceptical to digital readers and their affordances, the experience was positively surprising. What did not surprise was that updating the pages is slow, the response time of the input pen is long, the pen is not very accurate and sometimes stops working, the battery time is only a couple of hours and the display is of no use in direct sunlight or in dark. The device is also about the size and weight of a mini-laptop and a lot more expensive. But when comparing the reading experience (indoors) holding this display instead of a laptop, being able to make markings on the documents with the pen instead of having to navigate first the cursor to the correct place and then use the keyboard to type things, being able to draw on the document with the pen and having a matte display to read from instead of the backlit laptopdisplay, it was easy to prefer the device over a laptop. No articles were printed during the writing of this research paper and most of them were downloaded and read on the reader.

We consider the benefits of electronic display techniques in university environment to be remarkable, not only in monetary and environmental aspects, but especially for students. The first issue we thought it would adress was the heavy book loads students have to carry and the back aches it causes to carry them around in bags. Therefore the fact that it no articles could be found about the back problems of students was a very unfortunate one, as it would have been a good point to reflect upon and list as an argument for electronic paper devices. As soon as electronic paper devices adress all the preferred affordances of physical paper, it can also allow for enhanced study techniques for students. Especially if more work like the one by McFall is done and electronic learning is taken to its full potential.

Reducing the use of paper does have an environmental effect that can be decreased with alternative means. Therefore it is very important that resources are allocated in finding the most optimal alternatives to paper, and we suggest electronic paper as a strong candidate, because its production and usage is shown to use less natural resources in the long run. Not all physical paper can be replaced by the electronic variant, but majority of campus related paper products could as well be digital, as long as there are sufficient devices to manipulate and use them. What happens to the electronic paper when it is used or disfunctional, that is an issue that requires further studies. Its total effect to environment can not be defined until this is also added to the equation. Therefore already in the design phase the environmental aspects of the materials and the possibility to recycle as much of the device is important.

A digital reader is no e-paper, but it is the closest a consumer can get today. Taking the best of the digital readers and adding the affordabilities of real paper, we believe an electronic paper device designed with right usability parameters will be able to largely replace paper as we use it today. We believe that when the evolution of electronic paper reaches the state that has been described in this paper as an "ideal" solution, the future book, then there is no excuse for students to stick to their traditional heavy books and folders full of paper. Instead, with less cost, they can buy electronic books in their electronic paper device that can contain literature and study material for a life-time need, have less back-problems and be able to study more efficiently.

6 Summary and Conclusions

Not all use of paper can or should be replaced with electronic means, but by prioritizing the printing and using the technology available, the scarce resources as well as significant amounts of money can be saved in the ever tightening budgets of universities. The same applies for students, as the electronic handling of information reduces the costs of storing, transporting and distributing, which reduces the price of a book for the end user. Most of the documentation as well as teaching and learning material can be kept electronic and never should be printed. Even today it is possible to secure the confidentiality and take measures to stop students from cheating in electronically conducted exams. On the other hand it will take some time for people to give up information posters and commercials, handing out flyers and printing out diplomas and the like, until we live in a world where office walls are interactive white-boards. Nevertheless, any university should take measures to study the possibilities which enable them not only to reduce the paper usage, but to try alternate methods, such as electronic paper devices, and make the necessary changes to accommodate to them. Although future would not be paperless, there must definitely be less paper.

References

[1] Dzmitry Aliakseyeu and Jean-Bernard Martens. The electronic paper prototype with visual interaction enriched windows. In *EUSAI '04: Proceedings of the 2nd European Union symposium on Ambient intelligence*, pages 11–14, New York, NY, USA, 2004. ACM.

[2] Thomas A.M. Counsell and Julian M. Allwood. Reducing climate change gas emissions by cutting out stages in the life cycle of office paper. *Resources, Conservation and Recycling*, 49(4):340 – 352, 2007.

[3] EPA. Municipal solid waste in the united states: 2007 facts and figures, November 2008.

[4] Andreas Hallgren. Pappersförbrukning, April 2006. Luleå Tekniska Universitet, `http://www.ltu.se/d856/d1172/d1653/d1734/1.6048`, accessed 2009-05-09.

[5] Robert A. Hayes and B. J. Feenstra. Video-speed electronic paper based on electro-wetting. *NATURE*, 425(6956):383–385, 2003.

[6] Ziming Liu and David G. Stork. Is paperless really more? *Commun. ACM*, 43(11):94–97, 2000.

[7] Charles C. Mann. Electronic paper turns the page. *Technology Review*, March, 2001. `Http://www.technologyreview.com/web/12295/`, accessed 2009-02-22.

[8] Ryan McFall. Electronic textbooks that transform how textbooks are used. *The Electronic Library*, 23(1):72–81, 2005.

[9] Ryan McFall, Herbert Dershem, and Darcy Davis. Experiences using a collaborative electronic textbook: bringing the "guide on the side" home with you. In *SIGCSE '06: Proceedings of the 37th SIGCSE technical symposium on Computer science education*, pages 339–343, New York, NY, USA, 2006. ACM.

[10] Lena Olsen. Universitetet i siffror, December 2008. Göteborgs Universitet, `http://www.gu.se/omuniversitetet/presentation/universitetetisiffror/`, accessed 2009-05-09.

[11] PARC. Gyricon, 2009. `http://www2.parc.com/hsl/projects/gyricon/`, accessed 2009-04-19.

[12] Jon Rickman, Kim Todd, Tabatha Verbick, and Merlin Miller. The evolution of the electronic campus: from a communication system to an educational delivery system. In *SIGUCCS '03: Proceedings of the 31st annual ACM SIGUCCS conference on User services*, pages 65–69, New York, NY, USA, 2003. ACM.

[13] John A. Rogers, Zhenan Bao, Kirk Baldwin, Ananth Dodabalapur, Brian Crone, V. R. Raju, Valerie Kuck, Howard Katz, Karl Amundson, Jay Ewing, and Paul Drzaic. Paper-like electronic displays: Large-area rubber-stamped plastic sheets of electronics and microencapsulated electrophoretic inks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9):4835–4840, 2001.

[14] Abigail J. Sellen and Richard H. R. Harper. *The myth of the paperless office*. The MIT Press, 2003.

[15] Kong-King Shieh and Der-Song Lee. Preferred viewing distance and screen angle of electronic paper displays. *Applied Ergonomics*, 38(5):601 – 608, 2007.

[16] Simon. Electronic watermarks to help authenticate soft-copy exams. In *ACE '05: Proceedings of the 7th Australasian conference on Computing education*, pages 7–13, Darlinghurst, Australia, Australia, 2005. Australian Computer Society, Inc.

[17] John Stanik. A conversation with mary lou jepsen. *Queue*, 5(7):9–15, 2007.

[18] Pat Wyatt, Kim Todd, and Tabatha Verbick. Oh, my aching laptop: expanding the boundaries of campus computing ergonomics. In *SIGUCCS '06: Proceedings of the 34th annual ACM SIGUCCS conference on User services*, pages 431–439, New York, NY, USA, 2006. ACM.

Calm technology: Design guidelines

Jenny Peterson

Abstract: This paper aims to draw some guidelines on how to design technology that is calm. This was done by first gathering some information on the topic of calm technology, which is an idea that was first coined by Mark Weiser. The overall goal of calm technology is to be informative without being obtrusive in a world where ubiquitous computing will be more and more common. It is described as information that does not require full attention when interpreted. More literature has been gathered in order to find out how the theory calm technology can be obtained in practice. The result was found in different research fields and applications namely Ambient Information visualization, slow technology and natural interaction. They all contribute to the conclusion that calm technology should be designed to be perceived at a quick glance. Displays and computers should also be integrated in the environment in a way that makes the technology behind it disappear. This can be done by integrating it in decoration, aesthetics or by letting the user interact with the content in a natural way.

1 Introduction

The way that we use computers is gradually changing. From being stationary computers in our homes and work space they are now moving out into the environment, becoming portable and wireless, but also taking new forms and becoming invisible to us. Devices will be embedded and located everywhere: in homes, in public spaces and even on the people themselves. These different devices will be connected with each other to form networks where they can cooperate and exchange information so that they can fully support the human [20]. This is a new computer era called ubiquitous computing and along with this, calm technology follows [19]. Both of these concepts were originally coined by Mark Weiser [17].

The invisibility of computers can be achieved in two ways. They can be invisible to us because they are small and more integrated into our environment. Ubiquity can also be a mental disappearance, where the computers can be large and visible but not perceived as actual computers, but rather as other interactive objects [19]. Ubiquitous was originally a term to describe what role computers would have in future, but that future is very much here now.

With ubiquitous computing some design issues arise [19]. When computers are everywhere they should stay in the background, adopt to our needs and to the changes that we make and only call for our attention when crucial. The computers should be a part of our life without being obtrusive. They should support us in our everyday and working life and help us perform our tasks, but still let us be in control. We should be able to be in different situations, for example searching information, communicating and collaborating with other people, without being interrupted and pulled out of the context by computers [10].

This is where calm technology comes in. It deals with one of the big challenges of ubiquitous computing: designing technology that is calm and non-obtrusive [19]. In this report I want to find out how technology should be designed in order to be calm. I want to outline some guidelines for this.

To be able to do this, literature has been gathered on the subject of calm technology which is presented in Section 2. The following Section 3 describes some design fields that apply the theories of calm technology. The fields have been classified by the specific approach that is used to obtain calm technology. In Section 4 the concept of calm technology and the practical fields are discussed and evaluated to find some features that can outline design guidelines for calm technology. The conclusions from this are presented in Section 5.

2 Calm technology

To create calm technology is one of goals of the ubiquity era [7]. Information technology should be designed to both be calming and informative, something that is described as opposites [19]. If computers were designed the same way they are now, but placed in the environment, users might be overwhelmed by the fact that they are all around us, but also that all of them have information that reach out to us that we have to grasp [7].

A calm technology should neither cause stress nor interrupt [7]. It should let people live their lives as usual, performing tasks and activities, but still support them. This support should work as an enhancement or extension to the person's unconscious and intuitive mind, by presenting information in an easy and accessible way [7]. The technique behind it should be as invisible as possible while the content is important and emphasized. Further more, a calm technology lets people be in the control and not the other way around.

Information is non-obtrusive if it is chosen to be consumed, and obtrusive if it interrupts and pulls you out of the context you are in and the thoughts you where thinking [2]. It is also believed that the reason that some technology is obtrusive and some is not has to do with how they get a hold of our attention. Calm technology is described as something that engage both the center and periphery of our attention and moves back and forth between them [19].

We use the center of our attention, or central processing, when we gather information trough explicit data [11]. Explicit data is something that is precise and clear but that need attention to be understood. This is described by an example where we want to decide whether to take an umbrella with us or not [11]. We can make the decision by looking for a weather forecast on the Internet which will have all data presented that we need. To be able to do so, we must pay attention to it and interpret the text that we read, which require some cognitive work. This is the data that most displays have [11].

Implicit data however, is more discrete and is received through the periphery processing of our attention [11]. This is a more unconscious and instinctive way of gathering information which require little cognitive thought. To check the weather forecast this way would be to look out through the window and make a quick decision based on what is seen. Full attention is not needed and the window can be in the periphery all the time. The comparison between a window and our peripheral attention is further described by Alessandro Valli [18]:

> *"When a room has a window, it provides a continuous flow of information about the external environment - the cycles of day and night, the movement of sunlight and shadows, the succession of bright and cloudy moments, and the alternation of dry and rainy patches. You rarely pay explicit attention to all this, but you are peripherally aware of it, and you feel uncomfortably isolated if you are cut off from it."*

By periphery, it is not meant what we see in the corner of our eye, but everything that we are aware of that is happening around us, but not have our complete attention. The relationship between the periphery and the center of attention is important in calm technology and is exemplified when driving a car [19]. The sound of the engine is always there but the attention is payed on the driving and listening to the radio and the passenger. But if the engine would make a sound that is unordinary the driver would quickly attend to it, which means that it was in the periphery all the time.

What is in the periphery is in no way unimportant, it is just not important for the moment. If it gets important it will be put to the center again. Weiser states that this transition is encalming for two reasons [19]. For one, if we have data in our periphery, we can be adjusted to much more than if it was in the center of our attention, since no or little cognitive processing is needed. If we would receive as much data in the center as in the periphery it would cause an overload since we would need to pay attention to it all. The periphery is informative but yet not causing cognitive overload. To have much information in the periphery is also better. For example, to talk to a person face to face require less attention than talking in phone, since additional information can be interpreted trough implicit data like body language and facial expressions [11]. Secondly, it is encalming to decide for our selves what will be in the center at the moment [19]. This gives control and thereby calmness. We choose what we want to pay attention to and when.

Calm technology reduces the cognitive load by placing information in the periphery. This cognitive load can lead to information overload which is an individual's physical response to an information intense environment [3]. This environment can be intensive for some reasons. It can be an ordinary decision-making in basic everyday tasks where there are too much options, or searching something relevant amongst a load of irrelevant information. It can cause frustration to nor have time nor capacity to read it all [3].

Calm technology gives new perspectives to interaction design but is not suitable for all technology and situations. It is needed when technology is located in public space and in our everyday life [11]. In those situations we want to do something else and pay attention to other activities and not be bothered by computers. Calm technology should give us time and support to do so by offering implicit data instead of explicit. It is therefore not suited for expert or personal use where a full attention is needed. Other application and functionalities that are not suitable are those that need immediate attention, for example video games or fire alarms [19].

3 Examples

Calm technology deals with the issues and thoughts that needs to be considered when implementing ubiquitous computing. It is more of a vision and a way of thinking, than a fact. Because of that, the concept has been interpreted in different ways by different researchers and developers. This has led to a number of new fields that has their core in Calm technology but with some additional features. These following directions are all referring to Mark Weiser or Calm technology in some way and many of these show practically how calm technology can be obtained. I have chosen to interpret these directions into how they are connected to calm technology: calm technology through interaction, aesthetics or mental rest. They are here described with examples.

3.1 Calm technology through aesthetics

The fields that are most connected to calm technology are the fields Ambient Media and Ambient information visualization. These systems are located around us and convey some sort of data through abstract visualization. They can be put in environments where it is not suited for computers and convey information without the need of immediate attention [13]. They are calm technology in practice, since they are designed to present information that do not require full attention and reduce the cognitive load [12].

The visualizations are to be lived with rather than used, and therefore needs to be calm and and non-obtrusive [14].The system can be physical artifacts that express some data, for example a lamp that changes the intensity of the light depending on different data input [14]. It can also be as displays that looks like posters or paintings [14]. The displays must be carefully designed to invoke a person's peripheral attention so that they can be understood with just a glance and even from a distance. This can be done by using basic perception rules like involving shape, color, sound, grouping, size to visualize the state of the content [18]. Animation distracts when constantly calling for the viewer's attention. The message should be clear and easy to comprehend, in opposite to devices that continuously take up the center of attention, where the user constantly must make decisions, assume things and choose right information source [18]. Further on, the systems should be designed with an appealing appearance. They should fit in in the environment and look good [14].

Figure 1: a) Energy Aware clock. The energy consumption of a house hold is visualized through a clock. b) InVis-a-Wall. Calender visualized through a wall paper.

The term Information decoration is said to be a sub category to ambient information visualization and is a way of dealing with the amount of information that is flowing towards us [1]. The concept is a reaction against displaying information on displays and suggest that information could be visualized through the environment instead. The author Van Mensvoort explains this by comparing how we seek and receive information with how humans used their surroundings to seek information about 30 000 years ago. Foot steps in the sand were a sign of an animal nearby, birds singing in a certain way could be a sign for storm and so on. In our society now we seldom look in our environment for information that we need. He argues that we in the modern society are not using our full potential to convey information. He also supports the ideas of calm technology and react on the fact that information designers too often design devices that are suited to reach a user's center of attention. In another article Josien Pieters writes that the current way of providing information through focus-needed screens is obtrusive and does not match the human perception. Humans through evolution have sought information at the edges of our field of attention and decide for ourselves if we want to attend to it. Van Mensvoort has a solution for this and writes:

"Look around you, wherever you are. Try to recognize all of the

forms and patterns in the space. The flowered wallpaper, the humming of the air-conditioning, the fish in the aquarium, a shadow on the wall. Do you realize how few of the patterns in our environment are being used as information carriers? Information overload? What information overload? The so-called information society has barely scratched the surface of our human bandwidth!" [1]

Van Mensvoort suggest that we should present information through the environment around us. The patterns around us could be information carriers as a form of information decoration. Decoration is natural for us and something we have around us. Besides, if we do not want to seek information in the decoration, it is good enough to just be decoration. It will not be disturbing when people see it as decoration instead of information [1].

Examples of information decoration are the Energy AWARE Clock and the wall paper InVis-a-Wall (1). The first one is a clock that visualizes the energy consumption in a household. The clock is wirelessly connected to an energy meter and shows how much that is being used at the moment and how much that have been used earlier. It is developed by a project team on Energy Design in Eskilstuna which is a part of the Interactive Institute [4]. The second application is a final bachelor project in University of Eindhover made by interaction designer Josien Pieters [12]. The pattern on the wall paper visualize a person's schedule or activities.

3.2 Calm technology through mental rest

Another term that is connected to calm technology is slow technology [6]. This concept is also focused on the aspects of the integration of computers in our lives, the ubiquity era. It involves new ideas to improve the relation between humans and computers, as they become a bigger part of our lives.

To be able to define Slow technology, a good start is too define its opposite, Fast technology [6]. This is the shape of the information technology that we have today that just like many other phenomena in society are quickly provided, like communication or fast food. People use information technology to perform tasks in fast and efficient ways. The computers or devices are designed specifically to be used as tools in specific situations to get quick answers or accomplish something. In the article where the expression first was coined the author argues for a transformation into slow technology [6].

This is information technology that fulfill other demands than just calmness [6]. It is designed for reflection and to give time for mental rest rather than being effective tools. The technology surround us and should therefore be used in activities that last for a longer time and that opens up for reflection. If fast communication is picking up the phone and calling someone, the slow version would be to travel to the person and talk with the person face to face. It takes longer and demands more effort but is also more valuable to us. The environmental integration of slow technology is compared to a chair [6]. A chair is not only designed for the specific purpose to be sit on, it is also designed to be a part of the environment for a long time and therefore needs to be aesthetically acceptable and well-integrated into the rest of the environment.

Slow technology go against the current when it comes to interaction design. It is not about designing a system that is the easiest to understand, learn and use. It is up to the user to discover the meaning of the artifact, what it means and how it works [11]. This might take longer time but due to letting the person reflect on it, the object gets more interesting. It differs from calm technology in a way that it encourage mental activity and demand more attention when using it [6]. On the other hand it is a concept that do focus on retrieving calmness, but in a slighter different way. I therefor see it as a sub category to calm technology.

An example of slow technology is also developed by Josien Pieters with co-worker Lizette Reitsma as a project during an internship at the Visual Experience Department at Philips Research in Eindhover [11]. They made a prototype of an installation that would be placed in a hotel lobby and that changes according to the sound level in the room. The system is called SISSy (2) and is implemented with flip dots that retreat when there are many persons in the room that makes a lot of noise and grows when the room is more silent. If a person gets curious and walks up to it it disappears. Another example is the informative art developed by the Future Applications Lab on Viktoria Institute in Sweden. The application shows bus departure times, visualized through an art work inspired by the dutch artist Piet Mondorian (2). Each square represent a bus and the location of the square shows in what direction the bus drive towards. Color and size describe how much time that is left to be able to catch the bus [14].

3.3 Calm technology through interaction

Another field that claims to have the theories of calm technology as foundation is Natural interaction. The concept is created by interactive systems engineer and experience designer Alessandro Valli [17].

This concept goes beyond using interfaces with icons, menus and pointing devices to interact with an object and suggests a more natural way [17]. The goal is to improve the relationship between computers and humans into a more natural manner. If we could interact with physical devices in our surroundings the same way we would do with real physical objects the mediated technology between us and the information would disappear. This kind of interaction would be a more natural way of interacting, as described:

> *"people naturally communicate through gestures, expressions, movements, and discover the world by looking around and manipulating physical stuff. The key assumption here is that people are meant to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do." [17]*

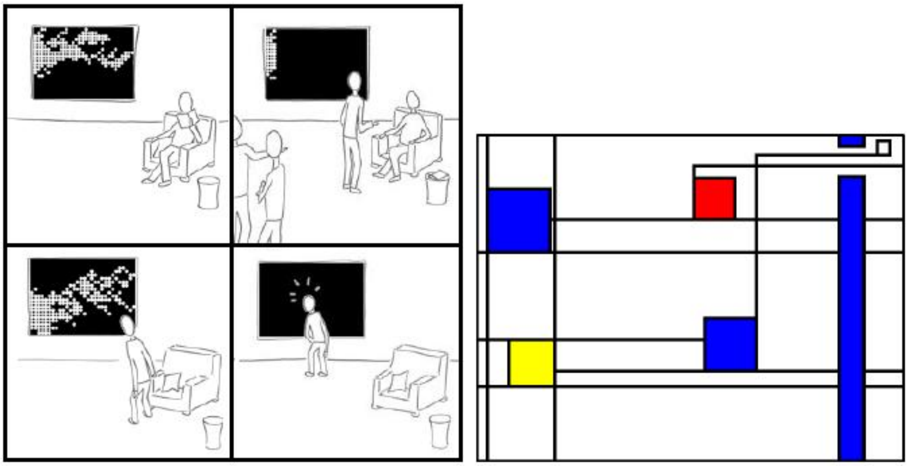

Figure 2: a) SISSy. When the room in noisy or when someone walks up to it the flip dots in the installation retreats. b) Informative art. Bus departure times visualized through art.

By interacting with computers the same way as someone would interact with a physical object the author believes technology could be more adopted to human needs. In history the human-machine-interaction has always been determined by the machines language and the user has always had to adjust their language in order to suit the machine's constraints and capability [17]. This relationship has improved over the years and has come to the point where it is possible to turn this around and instead make the machine adopt after the human needs and their way of communicating.

This human communication is described as, in contrast to learning an interaction scheme to be able to communicate with computers, a spontaneous way of exploring what is found interesting and attractive at the moment, by reaching out and touching it [17]. Designing natural interaction is about presenting content in an appealing way and let the users express what they find interesting spontaneously.

For a person to interact with digital information in a spontaneous way it can not be presented with computers or things that look like computers. It must look and behave like real objects [17]. It is suggested that it is only then the interaction can become natural. The difference in behavior between digital and physical objects is that digital objects are not affected by gravity and other laws of physics, which means that they can be modified, altered and behave in ways that are impossible for real objects [17]. To solve this when designing natural interaction these constraints are added to the digital information or objects. When the digital representation look and behave naturally, the technique behind it will become invisible and disappear from our minds. By removing this wall or tool between the us and the content, our behavior towards it will

Figure 3: a)SensitiveTable Media Interaction. The pictures can be moved, enlarged and minimized by hand movements. b) sensisitiveFloor Water. Digital screen where water ripples are formed when touched.

become more natural[17].

An example of natural interaction is the installation sensisitiveFloor Water where water ripples (3) are formed when a person walks over it or touches it [16]. Another example is the application SensitiveTable Media Interaction (3) where pictures are manipulated through hand movements. They can be enlarged and moved around through touch [16].

4 Discussion

The paper has dealt with what calm technology means originally and how it has been interpreted by different researches since then. Designs that encalm have been done with some different strategies, the closest and largest field among the concepts of ambient media and ambient information visualization. Visualization can be done through integrating information in physical objects or in art work like paintings. There are still some constraints in providing high quality displays but in future this might not be a problem.

These ambient objects should be designed to match perception rules that apply to our periphery. This means designing the parameters color, grouping, size and shape in a way that makes it easy to see the content. Too much sound or movement in the visualization can cause the center of attention to be too involved. When interacting with computers that obtain the center of attention a user often has to draw conclusions and make decisions for themselves. This is not the case in the field of ambient applications. They should have a clear message that is easy to understand with just a glance. This is something that Mark Weiser defined in his first attempts to outline calm technology and I therefore think that it should be seen as a guideline. Another requirement in this field was that the systems should be appropriately designed to fit the environment, through aesthetics. I also think this should be added as a guideline since it has been mentioned as a must have in all of the fields that I have researched in this report. This quality is obvious in the fields of informative art and information decoration where art and decoration are central.

I have chosen to define calm technology in a wide approach. That means that I have not only defined calm technology as in what Mark Weiser did, I have also tried to find the qualities of calm technology in related fields. Slow technology and natural interaction are two interesting fields that add some additional ideas on how to retrieve calmness. Slow technology are technical artifacts that are designed with the aim to make the viewer reflect over the content. It should be designed both to what it is used for, but also to fit in the environment when it is not used. New design aspects to this is that the artifact should not be designed with the goal to be easy to understand, learn or use and not be time-efficient. Overall, the concept of slow technology and informative art is to let the user reflect on the content, what it is and how it works which will contribute to a mental rest and calmness. This is something that i want to add to the guidelines, it may not be what Mark Weiser had in mind but it is a concept that lead to calmness in a different way and is therefor interesting.In the case with the visualization of bus departure times opposite reactions have been made concerning the art and the context [9]: How are we supposed to know that the art work carry some information in it, and is not just art? Problems with visualizing information in art is that a people will not understand how to read it as anything else than art. They are also somewhat dependent on the context they are in, it might be easier to grasp what the art is about if it is placed in the same context as it has information about.

The concept of natural interaction suggests another strategy to obtain calm technology by making technology invisible. This is done by adding physical constraints to the digital content and presenting it in an appealing way. In this way, the ordinary interaction schemes that is used when interacting with computers disappears, and is replaced by a natural interaction. Thus, the technology behind it becomes invisible. One might wonder in what way natural interaction is calm since calmness has a lot to do with receiving and understanding information at a glance, while natural interaction is about interacting with them. I do agree with the author and developer that it contributes to calm technology to interact with something in a natural way. Implicit data in one of the demands for calm technology and to interact with natural expressions like gestures and what ever comes to mind, is in deed implicit. No cognitive effort has to be done. They are unconscious and instinctive movements and thereby peripheral. I therefor want to embrace the qualities of natural interaction and add it to one of the guideline in order to obtain calm technology.

Literature describe calm technology as everything from being the same thing as ubiquitous computing to only being a theory about having information presented in the periphery in stead of the center. It can be questioned if ubiquity is a need in order to have calm technology or the other way around, that calm technology is a need to be able to have ubiquity. I believe the latter, that calm technology should be seen as an important design aspect when implementing for ubiquity. If computers will be places anywhere, on walls, in the home, in public areas and also be many, not all can be designed to be sit in front of, causing cognitive load when presenting explicit data. They must be designed to be easily perceived by us through the periphery. Secondly I think it is very important that they are designed to blend in with the environment when not used, an environment filled with screens might not be so pleasant. Informative art is a sub category to slow technology but may seen similar to information decoration. The difference is vague and I suggest that they only carry different names since they are developed by different authors. The work of informative art or information decoration can be very similar, as in the case of In-Vis-A-Wall and bus departure times. They can also be very different when it comes to SISSy and the Energy consumption clock.

Calm technology brings some solutions for usability issues that arise with the ubiquity era. These issues will no longer be restricted at designing interfaces between one computer and one user. They are broadened to involve usability issues of users interacting with many devices that are all around. The subject brings many positive approaches to design but opens up for questions on the term "calm". It is claimed that calmness is obtained through peripheral information. But according to me, the definition of "periphery" must be better understood in order to be able to make satisfying solutions. The work of "In-Vis-A-Wall" is interesting but somehow I believe a calender on the wall would be more in the center of my attention than if I keep it in my bag and look at it when I need to. A wall is so big that it brings attention to it. Another issue concerning the wall paper is that it might be stressful to see upcoming activities visualized in a room that you are in and see all the time. So perhaps this wall is not in my periphery after all. I suggest that a smaller piece would be better in order to perceive the peripheral awareness of a user. Slow technology is a good thought and could work for some ideas, like SISSy. Although, to display crucial information like bus departure times might have negative effect, causing frustration when not understood. A piece of art that is more mapped to the context of bus departure and looks less like a painting could work better.

5 Conclusion

The guidelines were discussed and motivated in the discussion Section. The most general and frequent qualities are stated in the first two guidelines and guideline three and four are found in the related fields of calm technology. These are the design guidelines needed for technology in order to be calm:

1. It should be easily perceived at a quick glance.
2. It should be integrated in the environment and be aesthetic.
3. Letting the user reflect on the content, what it is and how it works, will contribute to a mental rest and calmness.

4. A natural interaction with the content, where the content looks and function like real objects, will contribute to calmness.

Future research is necessary to do concerning possibilities in integrating information in environment. Organic User Interfaces is a research area that look into the options of creating displays that are non-planar, which means they can change their shape through bending or twisting. They can also be integrated in fabric which opens up for new interesting ambient artifacts[8]. Another important research field is psychology, regarding information technology and its effect on people and also cognition and perception. If this is studied, more requirement for technology that suits us people can be found.

References

[1] *Information Decoration: Our environment as an Information Carrier.* Idea Books, ISBN, `http://www.nextnature.net/?p=1565`, artvertising: the million dollar building edition, 2007.

[2] Katy Barnett. Information overload and how to avoid it. Technical report, University of Hull, 2007.

[3] Jan Buse. För mycket information? - om begreppet information overload. Technical Report ISSN: 1401-5358, Bibliotekshögskolan, Borås Högskola, Borås, Sweden, 1997.

[4] Energy Design. Energy aware clock, 2006-2008. `http://www.tii.se/node/3400`, accessed 2009-05-08.

[5] Hallnäs and Redström. Designing for reflection. Technical report, Viktoria Institute, 2000.

[6] Lars Hallnäs and Johan Redström. Slow technology - designing for reflection. *Personal and Ubuiquitous Computing*, 5(3):201–212, 2001.

[7] Björn Hermes. Calm technology (ct): Increasing supply without increasing demand. `http://www.hermans.org/agents2/ch4_3.htm`, accessed 2009-05-08, 2000.

[8] David Holman and Roel Vertegaal. Organic user interfaces: Designing computers in any way, shape or form. *Communications of the ACM*, 2008-05-30.

[9] Robert Kosara. When informative art isn't. `http://eagereyes.org/VisCrit/InformativeArt.html`, accessed 2009-05-08, 2006.

[10] D. Ley. Ubuiquitous computing. *Emerging technologies for learning*, 2(6):64–66, 2007. `http://www.canterbury.ac.uk/education/protected/ppss/docs/emerg-tech07-chap6.pdf`, accessed 2009-05-08.

[11] Josien Pieters. Internship report, philips research: visual experience department. http://www.student.tue.nl/r/j.j.m.pieters/SE3.1/internship%20report%20reading%20mode.pdf, accessed 2009-05-08, 2008.

[12] Josien Pieters. Between you, me and the wall. Technical Report Final Bachelor Project, Industrial Design, University of Technology, Eindhover, 2009.

[13] Zachary Pousman, John Stasko, and Michael Mateas. Casual information visualization: Depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1145–1152, 2007. http://www.cc.gatech.edu/~john.stasko/papers/infovis07-casual.pdf, accessed 2009-05-08.

[14] Holmquist Skog, Ljungblad. Between aesthetics and utility: Designing ambient information visualizations. http://www.viktoria.se/fal/projects/infoart/pub/ambient-infovis.pdf, accessed 2009-05-08, 2003.

[15] Norbert Streitz and Paddy Nixon. The disappearing computer. *Communications of the ACM*, 48(3):32–35, 2005.

[16] Alessandro Valli. The examples of sensisitivefloor water and sensitivetable media interaction, 2006. http://naturalinteraction.org, accessed 2009-05-08.

[17] Alessandro Valli. Natural interaction white paper. http://www.naturalinteraction.org/images/whitepaper.pdf, accessed 2009-05-08, 2007.

[18] Alessandro Valli. Notes on natural interaction. http://www.idemployee.id.tue.nl/g.w.m.rauterberg/Movies/NotesOnNaturalInteraction.pdf, accessed 2009-05-08, 2007.

[19] Mark Weiser and John Seely Brown. *The Coming age of Calm Technology.* http://www.johnseelybrown.com/calmtech.pdf, 1997.

[20] BG Wennersten. Ubuiquitous computing - en vision som kan bli verklighet. http://www.datainspektionen.se/Documents/rapport-ubiq-computing.pdf, accessed 2009-05-08, 2007.

Using table-top displays to enhance creativity and collaboration in design projects

Anders Salander

Abstract: A result of the technical advancement is that new products and interaction techniques has to be developed. Typically, products and systems developed today are complex in nature and they often require a large design-team working in parallel on different parts in order to create the final product. This puts high constraints on efficient collaboration and knowledge exchange between the design-team participants. Interactive table-top displays are a technological tool that might support the kind of collaborative activities found in design projects and this paper aims to investigate if this is in fact true. Results show that table-top displays come with a set of important advantages over traditional systems in order to support collaboration and shared understanding.

1 Introduction

Design is important for most companies today. As development in technology evolves, new products and services needs to be implemented. However, this often leads to complex problems that are impossible for a single individual to handle instead collaborative design-teams is often used. Complex design projects often depend on knowledge in a wide variety of domains in order to be successfully solved. This often implicates that team-members from different areas have to cooperate and collaborate to solve the mutual problem.

A resulting problem with this kind of group-constellation is lack of shared understanding between the different participants. However shared understanding can be an important factor for the success of a design project. One possible solution for creating shared understanding and enhancing the collaboration in design projects is the use of interactive table-top displays in the collaborative part of the project. Several advantages with table-top displays compared to ordinary vertical displays [21] have been identified and Scott et al. [24] have presented a set of important guidelines regarding the development of interactive table-top display systems.

The motivation of this paper is to investigate the impact that interactive table-top displays have on complex co-located collaborative activities such as design projects. In order to discuss the advantages of this kind of system the paper first presents some important background information regarding a designer's role, the work-flow in design projects as well as some characteristics of collaboration. An introduction to interactive table-top displays is then be presented highlighting advantages and disadvantages, interaction issues, design guidelines as well as current usage areas.

2 Designers and Design projects

The motivation of this section is to describe important features for a designer and the role a designer play both in design projects and organizations as well as characteristics of design projects and the process of design.

The demand for increased product performance, lower prices and reduced time-to-market on today's market have increased the importance of designers. Nowadays the design department plays a central role in almost any company since the designers have much influence over the company products [19, p. 7-8]. The definition of a designer in this section follows Pahl and Beitz [19, p. 1] definition, that is the term designer is used synonymously for design and development engineers.

Beside the fact that designers determine product features such as safety, ergonomics, operation and recycling designers also affect the products production and operating costs, quality and the production lead times [19, p. 7-8]. Design is in a psychological respect a highly creative process which forces a designer to have basic knowledge in many areas. Topics such as mathematics, material technology, engineering and design theory should be possessed by a designer [19, p. 1] since they have a great responsibility. The designers' ideas, knowledge and skills may have a large impact on a product's technical, economical and ecological properties [19, p. 7-8].

Designers often face situations where they work on assignments or problems they have little or no previous knowledge of [7, p. 85]. This requires designers to possess certain problem solving characteristics [19, p. 50-52] as well as the ability to communicate and understand the users and their situation [7, p. 22]. Having a holistic view is important for designers and they should understand many areas, as mentioned above, including marketing. A reason for this is that designers should be able to combine help from experts in the various areas with their own knowledge and expertize to successfully combine form and function as well as production issues during the design process [13].

Design projects are complex in nature. The impact of modern technology on design projects is that a single designer seldom have all the knowledge needed to complete a design project single-handedly. Instead large design teams are often needed in order to complete the project successfully [19, p. xviiii]. Usually the complex design tasks are broken down into smaller parts [4, p. 228] that allow different groups to work in parallel towards the solution [6, p. 7].

The increasing importance of design in organizations as well as the increasing technological complexity entails that many projects require skills in several different domains and typically people with different expertize have to collaborate in order to solve the problem neither of them could solve individually [29]. A successful product development project depends on design, manufacturing knowledge and marketing [6, p. 7]. An example of this is that a software development project typically contains designers, programmers, human-computer interaction experts, marketers and end-users [8].

The involvement of the actual users of the product early in the design process is important in design projects [20, p. 280]. It offers a company qualitative information directly from the end-users which in many cases leads to a more successful product with higher usability in contrast to design projects with no end-user participation [12].

One common issue faced in design projects is that there rarely exists one single optimal solution, instead there may exist many possible solutions each with its own advantages and disadvantages [29].

Design projects can be executed over several years and the original designers might not participate during the whole project [8]. Therefore documentation is a crucial part of the design project, a designer or design team should be able to review previous design decisions in order to avoid problems encountered in the past as well as reusing effective ideas [4, 8, p. 441].

The organizational structure and the top management's view on design is an important factor in the success of design projects. Many companies have a strict hierarchical structure in their organization. According to Johansson and Wodilla [13] designers often feel misplaced and powerless in this kind of structure.

The design process is typically performed iteratively and considered more 'chaotic' and unstructured than other project types. Designers often use a variant of trial-and-error using breakdowns as a tool for new solutions and knowledge creation [2]. Breakdowns is a unique opportunity for a designer to reflect and learn even though it can be painful, costly and time-consuming at times [8]. Johansson and Wodilla [13] highlights the importance of giving designers appropriate space and responsibility in the organization in order to work efficient toward the solution of the complex project.

The increasing industrial globalization also implies that companies world wide have to find new ways to increase their efficiency and use their employees' skills effectively in product development in parallel with the demand of cutting both production cost and lead times. Larsson [15] describes one of the Volvo Car Cooperation approaches to meet this market demand, that is, the development of a global virtual design teams. Although this method creates many challenges, both technological and cultural, it also allows the company to take advantage of knowledge within the company distributed all over the world in order to create new innovative solutions and products.

3 Characteristics of Collaboration

Section 2 described a designer's role as well as the complexity of design projects regarding the distribution of competencies between individuals and their area of expertise. Collaboration plays an increasingly important role in complex projects [8]. The following section aims to highlight some important characteristics of collaboration often found in design projects.

3.1 Face-to-face interaction

In design projects it is crucial for the designers to communicate with each other continually in order to solve problems, discuss alternatives and question assumptions as soon as they appear. Face-to-face collaboration is very efficient in the sense that the discussions can be performed in "real-time" and kept in phase with the actual design activity [16]. The efficiency of the discussions in design projects often depends on the participants ability to interpret non-verbal modalities such as gestures, deictic references and eye contact [17].

Observations by Bekker et al. [3] indicate that design teams on average use nine gestures per minute during a meeting session. The study highlight the importance of gestures during collaboration as well as what kind of activities that involve common gestures. Three basic activities were identified in design projects, these were design activities (e.g. when design ideas are discussed), management activities (e.g. when discussing about which information has to be documented) and overall conversational regulation activities. Bekker et al. found four different gesture categories that were repeatedly used in some activity during face-to-face collaboration in design projects.

- *Kinetic.* These gestures are mainly used in design activities to simulate how a person would use and interact with the product. However kinetic gestures are also used in management activities both to describe the design process and information regarding the management of the project.

- *Spatial.* This is used both in discussions of design and managerial activities. The primary scope of use is to visualize sizes of objects or distances between objects and persons or similar tasks.

- *Point.* The main activities where pointing is used is in design and managerial discussions. It is often used to refer to objects, people, places or ideas and is triggered by pointing and referring to a part of a document or whiteboard, information on a computer-screen. Pointing gestures can refer to very specific information such as a word or sentence but it can often refer to more vague information such as a concept described on a whiteboard.

- *Other.* Gestures that do not fit any of the categories above but still are used during collaboration is included here. This kind of gestures are mainly used to complement some kind of sentences, indicating turn taking in speech, attract attention or similar activities.

3.2 Inter- and Multidisciplinary Interactions

An interesting issue in design projects is the group composition. Groups with highly qualified members with the appropriate knowledge, skills and motivations will be more efficient on average than groups consisting of less highly qualified members [4, p. 332]. The implication of bringing together many people with different background and training have led to higher idea generation, development of new methods and the production of more creative and original designs [20, p. 9].

Typically one might expect that groups with high diversity, that is multidisciplinary teams with a broad spectrum of knowledge should perform better results than more specialized interdisciplinary teams since the participants may have richer, non-redundant knowledge and different views of the problem at hand [4, p. 339-340]. However, multidisciplinary teams often comes with disadvantages as well. Design teams consisting of several participants with different backgrounds run a higher risk of experiencing communicational breakdowns since the participants might have significantly different perspectives on the problem at hand. This will in turn slow down the development progress [4, p. 9-10].

A simple example is that a computer scientist's interpretation of the term representation often differs significantly from a graphic designer or a psychologist interpretation. In practice this kind of differences will probably lead to confusion, misunderstanding and communication breakdowns in the design team [4, p. 9-10]. Beside using different vocabulary design teams often suffer from differences in standard operating procedures, values and goals. Differences in these factors make coordination more difficult and coordination is an important factor in order to archive success in multidisciplinary groups [4, p. 339].

Interaction among participants can highly influence the group outcomes and is crucial both in inter- and multidisciplinary groups [4, p. 336]. Interdisciplinary groups benefit from the fact that they all work in a certain domain and might undertake similar work even though they may have different specialization and expertize in a particular domain. Problems with interdisciplinary group composition is that boundaries based on shared histories of learning might create discontinuities between participants and nonparticipants as well as difficulties for newcomers in the group. Since all participants work in the same domain it is important that the group try to avoid being stuck with only already established wisdom in order to progress [8].

Typically traditional learning and working environments such as university departments are interdisciplinary and their development of division of labor have been proved very powerful. However, there is also evidence that "real" problems typically cannot be successfully approached by an interdisciplinary point of view because of the complexity of the problem [8].

Even though multidisciplinary teams might experience for example communication breakdowns they have been proved to successfully generate more creative solutions if managed accurately. One successful example is IDEO, a company that is highly successful in new product-design and uses multidisciplinary teams in their new product-development process. These teams have the potential to be highly creative and create new innovative products by combining old ideas in new ways [4, p. 331-332].

3.3 Shared Understanding and Knowledge Creation

The complexity of design projects often have the effect that designers lack a full understanding of how other participants' work affect their own part of the task [1], hence, shared understanding among participants is crucial in collaboration and product-development [6, p. 106]. Designers often depends on other people's expertise as well as the need to collect knowledge from textbooks, standards, legal constraints and information from prior design efforts [1].

It is important to consider several aspect of the problem-domain in design projects and much knowledge are held within different team-members as tacit knowledge. Communication between involved parties is very important in order to make use of the tacit knowledge and create a shared understanding [1].

Due to the limitations of human cognition, verbal and gestural communication alone is often not sufficient for shared understanding, hence, externalization is important as it form a basis for critique and negotiation toward a joint solution [1].

Designers working on building design often use formal or informal drawings, pictures, video, verbal and text-based specifications as external artifacts to document and communicate ideas. In these kind of projects there are design-teams working on different levels of design for the final building, typically architects are responsible for the architectural design while engineers are responsible for the structural design and construction workers are responsible for developing plans for the construction process. In order to be able to successfully construct the building all participants need to have a shared understanding [23] that for complex projects can be achieved by communication and when design project participants teach and instruct each other [1].

Multidisciplinary design-teams often encounter problems of lacked shared understanding in the beginning of the project since the participants are expert in different domains. However, the shared understanding evolves during the process as knowledge is communicated between participants [8]. A participant's area of expertise may also change during a design project as one develop new knowledge and evolve a shared understanding about other participants expertise areas and situations [14]. According to Elfving [6, p. 15] creation of new knowledge is dependent on a combination of existing knowledge and a flow of information.

3.4 Idea Generation and Creativity

Historically cognitive scientists regarded cognitive concepts such as intelligence, talent and creativity as internal processes in the brain. In this paper the definition of creativity follow the definition of Sundholm, Artman and Ramberg [31], that is, creativity is a collaborative activity with social and communicative transactions between participants who in some sense share a common goal.

It has been shown that group collaboration often perform better than a randomly selected single individuals regarding problem-solving and idea generation [4, p. 330-331]. Groups typically solve the problem more accurately and often produce more and better ideas as well. Two factors that improve the results in collaboration is aggregation and synergy. Aggregation means that each individual in a group brings unique resources to it in form of energy, knowledge, skills or attitudes that might be essential in order to accomplish the task [4, p. 330-331]. The other factor is synergy. This implies that the effectiveness increases due to joint action and cooperation. By using the collaborators individual knowledge and taking it to the next level new innovative ideas can be created. An example of synergy is the creative solution that might be created when participants with different background and area of expertise collaborate on a shared problem and where the solution goes beyond the knowledge of any individual member prior to the collaboration [4, p. 330-331].

Pahl and Beitz [19, p. 75] claims that designers often find solutions for difficult problems by intuition, meaning that solutions suddenly appear after some searching and reflection in the problem-domain. Good ideas are always evaluated subconsciously or preconsciously according to prior experience, expert knowledge and of the task at hand, making them seem like conscious thoughts which origin often is impossible to trace. There are several methods for the encouragement of intuition and the creation of new ideas. A simple and common set of methods involve discussions with colleagues. These kind of methods can be very efficient provided that the discussions do not stray too far and are kept organized. Some common idea generation methods include: Brainstorming and Synectics [19, p. 76-81].

People often think that creative persons work individually but in fact most of our intelligence and creativity comes from collaboration and interaction with other individuals [8]. Even though the individual creativity is a basis for collaborative creativity it is often highly overrated. Creative activities are performed on a daily basis in real world problem-solving tasks, not only in research labs. People typically experience situations with unpredictable conditions in their everyday life which may eventually lead to breakdowns [9]. Creativity does not happen inside a persons head, instead it is dependent on the interaction between the individual and its socio-cultural context. Creative activity evolves through the relationship of the individual and the world of his or her work as well as the relationship to other individuals [8].

4 Interactive table-top displays

Section 3 presented a few important characteristics in collaboration. This section aims to explore some characteristics of horizontal surfaces such as table-top displays and how it can support collaboration.

Tables are something familiar. Tables are used in daily activities in most cultures and today tables are found in many environmental settings such as homes, offices, control-centers, design centers, waiting areas and meeting rooms. There are many forms of tables for example coffee tables, meeting tables and desktops used to support a number of activities. Tables provide a familiar and convenient feeling and are used both in leisure and industrial work-place settings for people to meet, look over documents and a lot of other activities that requires a shared physical surface and face-to-face collaboration with verbal and gesture modalities [26, 32].

The use of a familiar physical object enhanced with the digital and computational capabilities that interactive table-top displays offers creates many interesting possibilities in social interaction. A study by Forlines et al. [10] identified that direct-touch interaction outperformed computer-mouse interaction for bimanual tasks often performed on table-top displays. The fact that users control the information displayed by direct interaction with the surface implies that large-size table-top displays may reduce the cognitive load between the user and the digital information as well as extending the human working memory capacity [27].

Several prototypes of table-top displays have been developed, for example UbiTable [28], DiamondTouch [5, 26], EnlighTable [32] and InteracTable [30]. many areas of use for these kind of systems have been studied, some proposed usage areas are in co-located collaborative environments such as educational environments [18], disaster management [25] and design projects [17].

Since the technology of multi-user table-top displays is relatively new there are many interaction issues yet to be discovered when using table-top displays in real world settings instead of in controlled research lab studies. Ryall et al. [22] performed a series of real world experiments where they observed how the users interacted with the table-top and what kind of issues they experienced during usage. They basically found three basic problem areas concerning the user interaction, some of the issues in each area is presented below:

- *Touch interactions.* One common issue regarding the user interaction was simultaneous touching. Some users tended to be hesitant to interact with the table at the same time as other users even though the table-top display used during the observation allowed multi-user interaction. When users interacted simultaneously, issues such that accidentally bumping or touching other users were found to be uncomfortable especially in setting where the users did not know one another. Another important issue were the inability of the system to distinguish accidental touches from intended touches. Some users also experienced discomfort in touching the screen

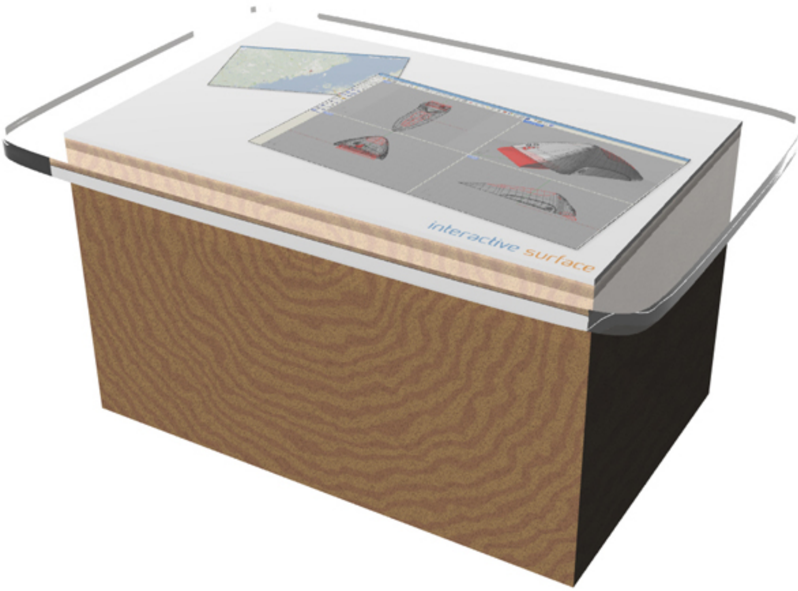

Figure 1: One example prototype of a interactive table-top display.

directly due to hygienic concerns.

- *Organization of content.* The observation showed that the physical size of the table-top display is of crucial concern. A small sized table typically led to a higher degree of crowding and cluttering and infringement of each others 'personal territory' both physically and virtually.

  Conflicts regarding a users taking for example a document that another users interacted with, users placing their own information on top of other users information and users changing the layout of the table-top display while other users interacted were some of the identified problems.

  Table-top displays also faces the challenge of information orientation among users. Some information such as long texts are highly dependent on the orientation for the users in order to be of any value. There exist several proposed solutions for this problem such as that in Lumisight-table [17] but today there exist little information about which technique is most suitable for various applications [22].

- *Physical setting.* Aside from the table-top size mentioned earlier the ergonomics of the table-top display is important. The way users interacts with the table-top display tend to depend on the physical height and viewing angle as it affects the readability and reachability of the information. According to the observations table-top displays that requires the users to sit seems to be more suitable to more casual interactions such as photo browsing while table-top displays that require the users to stand seems to be more suitable for productive tasks.

4.1 Design Guidelines for Table-top systems

In order to avoid the identified issues, Scott et al. [24] designed a set of eight guidelines for developers of collaborative table-top display systems. The main reason for this is to help developers design and create more ergonomic products supporting better interaction techniques for end-users of the system. These guidelines is summarized below:

4.1.1 Support Interpersonal interactions

Table-top systems should support the fundamental mechanism used in collaboration without interference. The technology should not be a obstacle for collaboration between the users instead it should be a tool for enhancing the collaborative activity.

Communicative gestures, deictic referencing and meeting coordination activities are crucial parts in collaborative activities therefore table-top display systems must support these kind of activities for co-located collaboration.

4.1.2 Support fluid transition between activities

Users should be able to switch smoothly between activities such as writing, drawing and manipulating artifacts on the table-top display, hence, the technology should not introduce extensive overhead in order to transition between activities. Systems that provides little or no overhead allow users to focus on the communication and task at hand instead of the actual system.

4.1.3 Support fluid transition between personal and group work

People often perform rapid and fluid transitions between individual work and group work during collaboration. When collaborating around a table they also seem to create a personalized area where they store information or artifacts for their individual use. Hence, table-top display systems should support these transitions and provide the users with some kind of personal space without interfering with the interaction and collaboration.

4.1.4 Support transition between table-top collaboration and external work

Typically the collaborative activity is only a part of a bigger operation that involves even more people. Typically the participants work individually beyond the table-top activity. The system should support transition and transferring of files and information created both individually and collaboratively between different displays and devices such as laptop computers.

4.1.5 Support the use of physical objects

Table-top systems provide a unique capability compared to vertical displays, that is the use of physical objects. The system should allow physical objects that can be used to interact with the system as well as other physical objects not directly used in the system such as coffee cups and laptop computers.

4.1.6 Provide shared access to physical and digital objects

Sharing object are an essential part in design tasks and sharing a surface can enhance the design process for collaborating designers. Tables are ideal for sharing information and it is common to see groups of people performing discussions gathered around a table. Table-top display systems should support equal access to physical objects used in the system as well as the digital information shared on the display. Each user should have the possibility to view the digital information easily so problems with orientation and occlusion has to be taken into account when designing a table-top display system.

4.1.7 Consideration for the appropriate arrangement of users

Collaboration around a table-top display implicate that users stand or sit at various locations around the table-top table. There are several factors that influence a users preferred position around a table, among them are the table size and shape. Users typically have a 'personal space' which has to be respected in order for the users to feel comfortable when interacting with the table-top display and the other users. Even though the users are located at different positions around the table they need to have the same access to the information as anyone else as mentioned in the previous guideline.

4.1.8 Support simultaneous user actions

In order to use table-top displays efficiently they have to support simultaneous multi-user interactions a feature that traditional computer systems often lack. By allowing simultaneous interaction the users can focus on the work at hand and increase the collaboration with the other participants.

4.2 Horizontal versus Vertical Displays

Rogers and Lindley [21] performed a study comparing how problem-solving collaboration was effected by horizontal or vertical orientation of the display. There was three different settings examined in the study; a large-size vertical wall display, a horizontal table-top display and a 17 inch PC-monitor. The result of the study showed some significant differences between horizontal displays and vertical displays. Test-subjects perceived the horizontal table-top setting to be more natural for collaboration while vertical settings felt more socially awkward. The main reason for this was that the test-subjects experienced a higher degree of participation in the horizontal setting. One result of this was that the test-subjects tended to exchange more ideas between each other in the horizontal setting compared to the large-size vertical setting while the PC-monitor setup generated significantly less ideas compared to the other two setups.

The horizontal table-top display encourage the collaborating participants to switch roles more frequently, resulting in a higher degree of contribution by each individual member. This could lead to a situation where more solutions are explored and externalized. In the vertical setting one test-subject typically took a leading role and acted more like a presenter while the other test-subjects had more passive observer roles as they felt that it was socially awkward to collaborate side-by-side in front of the vertical screen rather than face-to-face over the horizontal table-top.

Communication in the vertical settings was primarily based on verbal instructions while the test-subjects complemented verbal instructions with non-verbal gestures. Several test-subjects typically thought of the benefit of using horizontal table-top displays for creative and informal collaborative tasks such as design projects.

Findings by Ryall et al. [22] show that people do not tend to think of table-top displays as computers. Users experienced that a table-top display were less intimidating and more playful than an ordinary computer equipment. Ordinary displays are typically used for visual output with little or no user interaction while surfaces on the other hand encourage users to interact [26].

4.3 Areas of use

Interactive table-top displays have a wide range of potential usage areas and these kind of systems have been evaluated in several different domains. In recent years interactive tables such as table-top displays have received a lot of attention and they are attractive systems because of the range of possible applications areas they offer [11]. A few proposed usage areas are presented in short below.

4.3.1 Educational settings

Morris et al. [18] show that the use of table-top displays in educational group-work activities such as learning a foreign languages can be beneficial since it seems to encourage the students to collaborate in order to learn. Findings by Matsushita et al. [17] shows that collaboration using table-top displays also leads to more fruitful discussions.

4.3.2 Disaster management

Scotta et al. [25] identifies several important benefits with table-top systems in the response phase in disaster management. Collaboration and coordination are crucial in order to successfully manage crisis situations. Disaster management often require complex organizational structures and many entities such as fire brigades, police departments and ambulance resources have to collaborate and coordinate their work very careful. Typically time and stress are critical factors in disaster management as there may be human lives at stake.

Scotta et al. presents a prototype system based on the DiamondTouch table that is meant to be used as a tool in disaster management. Operational leaders for respective rescue team work in a control-room collaborating around a large table-top display showing for example maps of the crisis area as well as the position of the individual rescue personnel on site. This enables the operational leader to work in parallel and have real-time contact with their individual rescue-team at the same time as they maintain a good overview over the whole situation and rescue operation [25].

4.3.3 Urban planning

Arias et al. [1] presents a system for urban planning. A key part of this system is a horizontal interactive table-top surface. Urban planning projects demand several collaborating parties from different domains in order to be solved satisfactory. This kind of projects benefit from the possibility to use external object representations in order to create a shared understanding and overview over the whole project among the participants. The importance of using physical artifacts as a tool for creating understanding in design projects are highlighted by Matsusiha et al. [17].

5 Discussion

Interactive table-top displays are a very interesting product that could extend the current work-flow for complex and highly collaborative tasks such as design projects. The main advantage is that it combines the traditional and familiar setting of collaboration around a table with the digital augmentation and the power of computer technology. This allow the users to interact with the digital information at the same time as they can take advantage of non-verbal face-to-face modalities such as body language, gestures and eye contact in the collaboration. Studies by Morris et al. [18], Matsushita et al. [17] and Rogers and Lindley [21] have already shown that these kind of systems could be a factor for stimulating the collaborating participants to be more creative and exchange more ideas.

The fact that people perceive interactive table-top displays as less intimidating than ordinary computer systems can have a large number of benefits.

People that typically are afraid of using computers could be encouraged to interact with this kind of systems allowing a previously 'neglected' target-group to take part of new rich interaction possibilities and information sources. This scenario could benefit the future development of more user-centered products and services targeting for example elderly people since companies could involve people from that target-group early in the design process and hence get a better understanding of their needs and demands.

Even though this kind of systems have high potential for success they suffer from a set of crucial interaction issues that has to be resolved in order to be fully usable. Writing is a common procedure in most tasks since each project demands documentation to some extent and table-top displays is not at all suited for this kind of task. Streitz et al. [30] present a solution for this implemented in the InteracTable. An external wireless keyboard is used as a complement to the table-top display in order to allow typing input. However, this solution comes with a number of drawbacks that have to be considered when designing the system. It is important to make the keyboard feel like a natural part of the entire system instead of feeling disconnected and too externalized. The keyboard must have a natural placement on the table-top and should allow the users to get easy access to it instead of induce a set of user-actions to allow typing. The users should also have equal access to a keyboard to allow simultaneous writing.

Issues regarding the limitations of human vision accounted for in table-top display systems is another important research area. Each user should be able to view the informational content in a satisfactory way regardless of their position around the table-top display or their physical characteristics. Hence, the table-top display has to support a wide range of viewing angles as well as minimize the effect of occlusion that is introduced in some table-top display systems and a satisfying solution to the orientation problem is also crucial for the success of table-top displays.

The designers of interactive table-top display systems should be aware of the hygienic issues that is introduced in table-top displays. The users touch the interactive surface directly with their hands and this may be a source of bacterias. Some prototypes introduce the opportunity to use a stylus pen as interaction tool instead of using the fingers directly. This might be a solution for some users, however, this introduce the problem that the users has to keep track of at least one (probably two) physical objects in order to interact with the table-top display. Just like with the use of an external keyboard it is important that the interaction tools feels natural to use in the system in order to fulfill its purpose.

Interactive table-top display systems should be suitable for complex tasks that demand a high level of collaboration such as those found in design projects. Even though the digital and technological development of for example telecommunication systems, Elfving [6, p. 74] argue that designers still require some kind of face-to-face collaboration as a part of their design work. Studies by Arias et al. [1] and Scotta et al. [25] show that table-top display systems can be a very useful tool for complex projects that require much collaboration such as urban planning and disaster management. This kind of projects often have a limited time-constraints and involve people with expert knowledge in different domains. This setting is often found in design projects today, whether it regards product development or other kinds of design tasks such as architectural projects. But these systems should be suitable for somewhat simpler design tasks such as graphical design as showed by Terrenghi [32].

In the long run table-top displays might be a starting point for the development of highly interactive environments as technology becomes more ubiquitous and interactive. Researchers at Mitsubishi Electronic Research Lab (MERL) have a developed a range of interactive products. Some of them like the UbiTable and the DiamondTouch table are interactive table-top displays. However, the researchers also work on a project called DiamondSpace which is a intelligent environment using several different products including table-top displays [26]. Streitz et al. [30] presents a future vision of cooperative and interactive buildings and just as the researchers at MERL they have table-top display technology as a central part in the environment. One reason for this may be the many advantages that Rogers and Lindley [21] identified with interactive horizontal surfaces found in table-top display systems.

6 Conclusion

This paper have examined and discussed the usability of interactive table-top displays for highly collaborative, co-located tasks such as those found in most design projects. Based on the literature study the author's opinion is that even though there are several important interaction issues such as the orientation problem, the limitation of textual input and risk of cluttering yet to be solved in this kind of systems, they offer a great potential in supporting collaboration and development of new exciting products and interaction techniques in the future. The design guidelines for interactive table-top display systems introduced by Scott et al. [24] form a basis for the success of this kind of products for increased collaboration in complex projects such as design projects.

References

[1] Ernesto Arias, Hal Eden, Gerhard Fischer, Andrew Gorman, and Eric Scharff. Transcending the individual human mind - creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction*, 7(1):84–113, 2000.

[2] Ernesto G. Arias, Hal Eden, Gerhard Fischer, Andrew Gorman, and Eric Scharff. Beyond access: informed participation and empowerment. In *CSCL '99: Proceedings of the 1999 conference on Computer support for*

collaborative learning, pages 20–32. International Society of the Learning Sciences, 1999.

[3] Mathilde M. Bekker, Judith S. Olson, and Gary M. Olson. Analysis of gestures in face-to-face design teams provides guideance for how to use groupware in design. *Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods and techniques*, pages 157–166, 1995.

[4] John M. Carroll. *HCI Models, theories and frameworks - Toward a multidisciplinary science.* Morgan Kaufman Publishers, 2003.

[5] Paul Dietz and Darren Leigh. Diamondtouch: A multi-user touch techology. *Proceedings of UIST 2001, the 14th Annual ACM Symposium on User Interface Software and Technology*, pages 219–226, 2001.

[6] Sofi Elfving. Managing collaborative product development - a model for identifying key factors in product development projects. Master's thesis, Mälardalen University, 2007.

[7] Xristine Faulkner. *Usability Engineering.* Palgrave, 2000.

[8] Gerhard Fischer. Social creativity: turning barriers into opportunities for collaborative design. In *PDC 04: Proceedings of the eighth conference on Participatory design*, pages 152–161. ACM, 2004.

[9] Gerhard Fischer, Elisa Giaccardi, Hal Eden, Masanori Sugimoto, and Yunwen Ye. Beyond binary choices: Integrating individual and social creativity. *International Journal of Human-computer Studies*, 63(4-5):482–512, 2005.

[10] Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. Direct-touch vs. mouse input for tabletop displays. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 647–656, 2007.

[11] Gerard Hollemans, Tom Bergman, Vincent Buil, Kero van Gelder, Maurice Groten, Jettie Hoonhout, Tatiana Lashina, Evert van Loenen, and Sander van de Wijdeven. Entertaible: Multi-user multi-object concurrent input. In *UIST 2006 Adjunct Proceedings: Demonstrations*, pages 55–56, 2006.

[12] Karen Holtzblatt and Hugh Beyer. Making customer-centered design work for teams. *Communications of the ACM*, 36(10):92–103, 1993.

[13] Ulla Johansson and Jill Woodilla. Designers dancing with(in) hierarchies: The importance of non-hierarchical power for design integration and implementation. *The Design Journal*, 11(2), 2008.

[14] Thomas Kvan. Collaborative design: what is it? *Automation in Construction 9*, 9(4):409–415, 2000.

[15] Andreas Larsson. Making sense of collaboration: The challenge of thinking together in global design teams. *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, pages 153–160, 2003.

[16] Gloria Mark. Extreme collaboration. *Communications of the ACM Adaptive Middleware*, 45(6):89–93, 2002.

[17] Mitsunori Matsushita, Makoto Iida, Takeshi Ohguro, Yoshinari Shirai, Yasuaki Kakehi, and Takeshi Naemura. Lumisight table: a face-to-face collaboration support system that optimizes direction of projected information to each stakeholder. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 274–283. ACM, 2004.

[18] Meredith Ringel Morris, Anne Marie Piper, Anthony Cassanego, and Terry Winograd. Supporting cooperative language learning: Issues in interface design for an interactive table. Technical report, Stanford University Technical Report, 2005.

[19] Gerhard Pahl and Wolfgang Beitz. *Engineering design - A systematic approach*. Springer-Verlag, 1999.

[20] Jenny Preece, Yvonne Rogers, and Helen Sharp. *Interaction Design - Beyond Human-Computer Interaction*. John Wiley and Sons, Inc., 2002.

[21] Yvonne Rogers and Siân Lindley. Collaborating around vertical and horizontal large interactive displays: which way is best? *Interacting with Computers*, 16(6):1130–1152, 2004.

[22] Kathy Ryall, Meredith Ringel Morris, Kathrine Everitt, Clifton Forlines, and Chia Shen. Experiences with and observations of direct-touch tabletops. *Horizontal Interactive Human-Computer Systems, International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)*, pages 89–96, 2006.

[23] Milad Saad and Mary Lou Maher. Shared understanding in computer-supported collaborative design. *Computer-Aided Design*, 28(3):183–192, 1996.

[24] Stacey D. Scott, Karen D. Grant, and Regan L. Mandryk. System guidelines for co-located, collaborative work on a tabletop display. In *Proceedings of ECSCW'03, European Conference Computer-Supported Cooperative Work 2003, Helsinki, Finland, September 14-18*, pages 159–178, 2003.

[25] Alessandra Scotta, Henk Scholten, Sisi Zlatanova, and Erik van Borkulo. Tangible user interfaces in order to improve collaborative interactions and decision making. *Proceedings of UDMS '06*, pages 9.27–9.39, 2006.

[26] Chia Shen. Multi-user interface and interactions on direct-touch horizontal surfaces: collaborative tabletop research at merl. *Horizontal Interactive Human-Computer Systems, 2006. First IEEE International Workshop on Horizontal Interactive Human-Computer systems (TableTop 2006)*, pages 53–54, 2006.

[27] Chia Shen. *From Clicks To Touches: Enabling Face-to-Face Shared Interface on Multi-touch Tabletops.* Springer Berlin / Heidelberg, 2007.

[28] Chia Shen, Katherine Everitt, and Kathleen Ryall. *UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces.* Springer Berlinn /Heidelberg, 2003.

[29] Joachim Stempfle and Petra Badke-Schaub. Thinking in design teams - an analysis of team communication. *Design Studies*, 23(5):473–496, 2002.

[30] Norbert A. Streitz, Peter Tandler, Christian Müller-Tomfelde, and Shin´ichi Konomi. *Roomware: Toward the Next Generation of Human-Computer Interaction Based on an Integrated Design of Real and Virtual Worlds*, chapter 25, pages 553–578. Addison, 2001.

[31] Hillevi Sundholm, Henrik Artman, and Robert Ramberg. Backdoor creativity - collaborative creativity in technology supported teams. *Cooperative Systems Design - Scenario-Based Design of Collaborative Systems.*, pages 99–114, 2004.

[32] Lucia Terrenghi, Torsten Fritsche, and Andreas Butz. The enlightable: Design of affordances to support collaborative creativity. *In proceedings of Smart Graphics Symposium*, pages 206–217, 2006.

The Golf terminal and its adaption to senior citizen interaction

Erik Sandin

Abstract: In the early 21th century the Swedish Golf Federation introduced the golf terminal. The purpose was to ease the amount of administrative work of the receptionist's. Through the golf terminal the golfers could book a tee time and register for competitions. Since there are a lot of active senior members at the golf clubs today this article is about giving some guidlines about how to redesign the terminal to fit the needs of the senior golfers. This showed that for example the audio feedback and the menu tree of the terminal are two examples of things that needs improvement to better suit the senior golfers.

1 Introduction

The golf terminal was introduced to the Swedish golf clubs early in the 21th century. The golf terminals act as an extra receptionist where the golfer can take care of some administrational work on their own. This can reduce the amount of work the receptionists have to do. One of the key features is tee time bookings that can be made on the specific golf club. The golfer can also confirm an already booked tee time. Another feature is that the golfer can manage registrations to competitions. There are two different kinds of terminals where both of them have the above mentioned features but the more advanced terminal can also accept payments when it comes to paying green fees and competition fees. In Sweden the average age of a golfer was 41 in 2008 [10]. In Umeå Golf Club the average member has an age of 43 years. There are 1211 active members of the golf club and 399 of those are 55 years or older. This means that a many of the golfing members of Umeå Golf Club are older seniors. The question is if the terminal is well designed to be used by the older senior members of a golf club. Most of the members use the reception to book tee times and sign themselves to competitions. If the terminal is going to be more used, the golf clubs should put more effort into marketing the terminals as a partial substitute to the reception. Another task is, as done in this article, to do a study and find out if the terminal in fact is designed to be an effective working tool for the older senior members of the golf club. The study shows that there are a couple of things that can be redesigned to better suit the senior golfers. For example can the hearing aids become a source of feedback. Another example is the menu tree, which is now visible but not clickable, and could be clickable easing the cognitive work of the senior golfer. The study discovered that some of the questioned senior golfers did not even know where they were in the menu tree when interacting with the golf terminal.

This article starts with listing the most important impairments of an older human beeing. The article continues with a discussion on how the golf terminal is built and how it works now. A user study is presented next and then there are some guidelines given that could help improve the interaction between the senior golfer and the golf terminal. The article is built up using a structure that is repeated throughout the sections. This is to give the reader a clear view of the pros and cons of each subsection.

2 Age-related impairments

As adults get older their sensory acuity declines. Below are some factors that can decline and therefore affect the elderly when using newer technology such as the golf terminal. Older adults with less declination of the sensors seem to use computers and other high technology gadgets more than those who do have some impairment [3].

2.1 Visual

The eye loses some of its capabilities when aging. The dark adaption for example loses its full spectra and can make it difficult for aging people to see clearly at night. Also, motion detections can become difficult when getting older. There are also problems with perceiving the color spectra when getting older [9]. This can lead to confusion of colors in the daily life.

Bergman and Rosenhall [1] show that there are about 1,5% of the 70 year old adults who actually have problem with their visual acuity. This percentage is increased as age increases. At age 88 only one third of the men and bearly one fifth of the women have a normal visual acuity.

Presbyopia is a common age-related visual impairment. It means that the person has difficulties focusing on an object at close range. Because of the short distance to the computer screen this phenomena makes it difficult for some to interact with a computer and notice stimuli on the computer screen [5].

2.2 Auditive

The loss of hearing is a big problem and approximately 30 percent of adults at the age of 65 and over have difficulties with hearing [12]. In this region there are many people with hearing problems and as the years pass by the hearing statistically gets even worse. Medications play only a small role when it comes to curing the hearing impairment [9]. Furthermore, older people can have difficulties understanding words from other people talking even though the older people can hear sound [12].

Bergman, Rosenhall (2001) shows in their study that nearly 25% of the adults, at the age of 70, have poor hearing acuity. At the age of 88 the percentage of poor hearing acuity is up to 80%. This shows that the hearing acuity declines at an earlier stage than the visual acuity.

2.3 Cognitive

Crystallized intelligence increases during the first 60 years and may not be the biggest problem for aging adults when it comes to decreasing the cognitive ability. There are two pieces that do decline and that is memory and cognitive speed that decline even in early adult ages [2]. Memory can be decomposed into short term and long term memory. Memory and cognitive speed are important factors when it comes to interacting with a computer.

2.4 Computer experience

Accordingly to the Swedish bureau of statistics [11], in 2006 there were 60 percent of the elderly in Sweden who had access to a computer in their home. This number referred to elderly of an age of 65 to 74 years old. If you go even further up the ladder the percentage of elderly with computers in their home decreases dramatically as age increases.

Having the computer knowledge among the older adults using a computer can help them maintain a social network by sending e-mails for example. The computer knowledge also gives the older adults a sort of independence [4].

3 The golf terminal

Back in 2002 the Swedish Golf Federation(SGF) took the decision to develop the IT-system which is still in use today. This is called GIT, which is an abbreviation of Golfens IT-system. Together with GIT SGF developed a golf terminal. The terminal is meant to ease the work of the receptionists. Since their assignments are to be the leading person within the club house the club itself can save money by rearranging tasks. The terminal is supposed to be used by everyone who wants to skip queues. The queue to the terminal is often shorter than the one leading to the reception. There is also another benefit from using the terminal and that is to be able to do some tasks by yourself. This will reduce the receptionists amount of work which in the end the golf club can benefit from. To use the golf terminal the golfer needs to have a plastic golf card issued by SGF. It takes a couple of weeks to get the card in the mailbox and during this time the golfer with no member card cannot use the golf terminal at all. As soon as a golfer has got their member card he or she can start using the golf terminal and as mentioned above there are some tasks that can be done by using the terminal. These are the following:

Tee time management Here you can book a new tee time as well as change an already existing one that has been booked by you in the past. You can also mark yourself as a player on site and then get a scorecard with your own key information on it.

Competition management This section gives the golfer a chance to sign up for everything from the smaller club competitions to the bigger more-than-one-day-tournaments. The golfer can also take back an earlier sign up for competitions.

Handicap management After a round of golf there might be the scenario that one or more golfers have to change their handicap. The golfer can change its handicap by specifying the result, played course and by giving the Golf-ID of a fellow-golfer.

See club information When golfers play on a course that he or she has never played before it could be nice to know a little bit about the golf club and the course he or she is going to play. This can be done by choosing this option in the main menu.

All of these menu options are also available on the Internet and on the webpage `www.golf.se`. Here golfers also can book a tee time or sign up for a competition in the comfort of their own home.

3.1 Visual

Visually the golf terminal screen consists of three parts, header, footer and a middle part, see fig 1. The position of the three parts do not change during the interaction with the terminal. The header consists of basic information such as the current date and time. There is also information about the logged on golfer. This header never changes its information during the interaction of the terminal. The footer includes four buttons. These buttons are rectangles with a small border around them to distinguish the edges of the buttons. To understand what the buttons do, they also have a small picture left of the text which can give a hint of the intended action. The four buttons are: Back, Info, Logout, Start. The footer also contains a strip at the very bottom which tells the golfer where he or she is in the menu tree. The menu tree is not clickable and therefore the user has to use the back button more than once to go back to a wanted choice The middle part always contains the sublevels that can be reached from the current level in the menu tree. These sublevels can be reached by clicking the rectangles with the corresponding texts. To highlight the middle part there is also a title presenting the current level in the menu tree. When a button is touched the user gets a visual feedback from the terminal. This is done by showing a button which is pressed. This is done for a short time period and it disappears when the loaded choice is shown.

The color scheme is different shades of blue except the navigation buttons which have a color of their own.

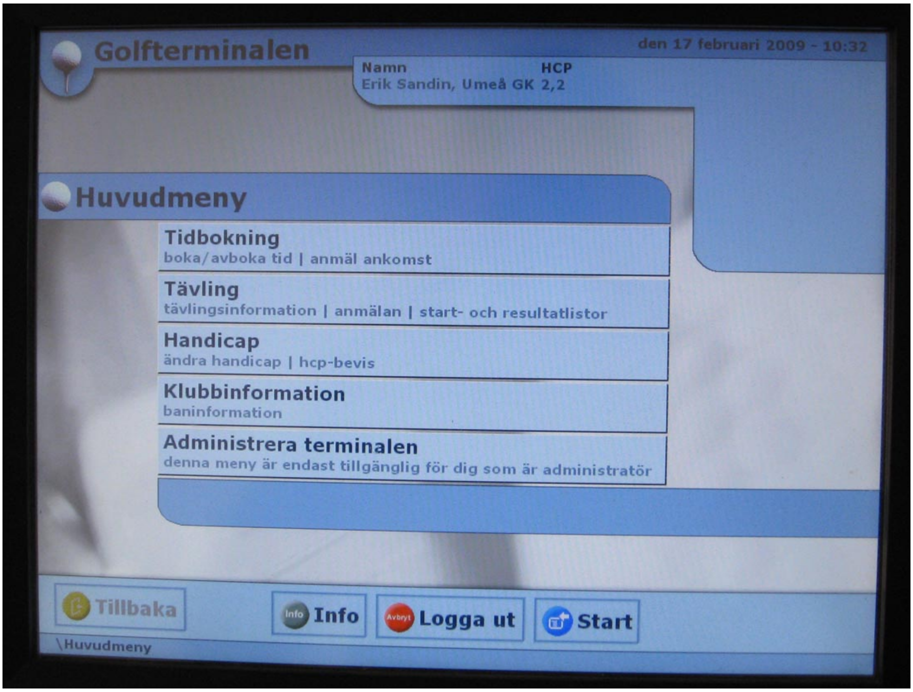

Figure 1: This is a screenshot of the menu when logged in.

3.2 Auditive

The terminal gives an audio response every time a user touches the screen. This feedback consists of a short sound. The user gets no audio feedback other than this short sound when touching a button on the screen. The same short sound is played and no special feedback tells the user if he or she has pushed the button or if they have missed the button and only touched the background.

There is one situation where there is a special feedback from the terminal to the user and that is when logging on. To log on the user has to slip his or her card in a card reader located at the right of the terminal. If this is done correctly the terminal plays one tone to inform the user that the card has been read correctly. If the user does not log on correctly there will be two tones played to the user which will inform the user that there is something wrong.

3.3 Cognitive

The terminal has a relatively small menu tree which has a maximum depth of 4. This means that it is easy for the user to remember where he or she is. The chosen path is also presented at the bottom of the screen to help the user's awareness about where in the menu tree the user is. The buttons in the middle part of the screen has a title that represents where the button leads the user. There is also an explaining text on the buttons which gives additional clues about what the user will find under each of the choices available. See figure 1.

3.4 Computer experience

The system is easy to use and requires no or at least very little computer experience from the user side. The only experience needed is how to press a button in a computer environment. The buttons give no feedback located at the actual button when pressed and that is why there needs to be some kind of knowledge behind.

4 User centered questionnaire

Since the golf terminal was introduced and up until today there is no research done in the area of HCI between senior golfers and the golf terminal. During the period of about 7 years there has to our knowledge not been any questionnaires that have dealed with the issue of senior golfers and their thoughts and ideas about the golf terminal.

To be able to find the most relevant prioritized guidelines a questionnaire has been used to fill in the gaps. During a day in May a handfull of senior golfers have been answering some questions regarding the golf terminal. The questionnaire followed the structure of subtitles in this report to give an as easily adapted explanation as possible.

4.1 The questions

As explained earlier the questions have followed the subtitles. This is to give the reader an easy way to follow the path of thought from the person answering the questionnaire. The questions are as below:

1. How old are you?
2. Is there anything wrong with your visual or hearing acuity today?
3. How much do you use a computer in your everyday life?
4. How much have you been using the golf terminal?
5. What do you think of the visual layout? What can be done differently?
6. How has the sound, that yields every time the screen is touched, helped you to interact with the golf terminal?
7. How can sound help you interact with the golf terminal?
8. What do you think of the usability of the golf terminal?
9. Is there anything else that you would like to add as a comment?

4.2 Results of questionnaire

The seniors that answered the questionnaires had an average age of 65, where the minimum age was 63 and the maximum age was 68. They all answered that the terminal was not their first choice when it came to book a tee time or register for a competition. All four of them used the web based version via www.golf.se and only used the terminal to once in a while change their handicap. They all answered that they had a computer at home and did not have any problems with the interaction of the golf terminal. Two out of the four senior golfers claimed that the knowledge of how to use a personal computer had helped them to use the golf terminal.

All four senior golfers claimed that the golf terminal was easy to use and they were all satisfied about the way the golf terminal works today. They did not say anything about how the golf terminal could be redesigned to be even more easy to use.

The visual appearance was also something that they all liked. Two of the questioned senior golfers wore glasses and had problems with the interaction if they did not look through their glasses. One of the two golfers with glasses had problems with the explaining texts under each of the titles of the choices available. There were no troubles with the titles alone though. Also the name and the handicap which has the same color were hard to read without glasses.

The sound that plays every time the screen is touched was something none of the four used as something to guide them to the wanted choice. Three out of the four had not even considered the sound as something that had helped them at all. One senior golfer mentioned that the fast switch between different choices did not make the sound that important since he got the feedback he wanted the moment after he touched a button. The sound did make all of the four a bit disturbed when it came to the menu tree. There were two different thoughts about the sound when pressing the menu tree at the bottom of the screen. The first mentioned was that if the menu tree is not clickable then why not remove the sound when it is pressed. The other proposal was to make the menu tree clickable. None of the four liked the fact that the sound was there when nothing actually could be done with the menu tree. One senior golfer mentioned that there could be a special sound when the background is clicked. That is to give the user a more direct feedback if a button is clicked or not.

As for the computer experience part all of the four senior golfers have had encounters with a touch display before. The bus station when a person buys a ticket, when doing the check-in at the airport are examples of places and situations where the questioned senior golfers have had an encounter with a touch screen.

4.3 Questionnaire issues

Since there is only four senior golfers that answered the questionnaire the results can be miss interpreted. The small amount of answers is a result of lack of time.

All four said that they used the web based version as their first choice method of interaction and this could have led to some answers that could have been otherwise if they did not use the web based version at all.

Another thing is that the senior golfers that answered all had a computer at home and therefore were unlikely to have any direct problems with the golf terminal.

To be able to give better results the questionnaire should have been done by people of different backgrounds. Due to the lack of time this was not dealt with.

The questionnaire together with a field observation could have given more material to work with.

5 Golf terminal design guidelines

There are a number of issues mentioned above both as a result of the aging process and from the golf terminal. Some design guidelines are presented below that can help the senior golfer to interact with the golf terminal. Also, after doing the user centered questionnaire some thoughts came up that gave some new input angles but also nearly discarded some thoughts about how to redesign the golf terminal to better suit the senior golfers.

The guidelines are key factors to give an as clear picture as possible of what is already adapted to senior usage and what can be improved to better suit the senior golfers and their interaction with the golf terminal. The Swedish Golf Federation can use these guidelines as a sort of advise for direction. If the guidelines are used to redesign the golf terminal the senior golfers will easily be able to use the golf terminal without major difficulties.

- **Warm colors**

  Since the user of the golf terminal is relatively old compared to the population the coloring can make a difference when it comes to operating the machine. According to [7] the older adults may have difficulties with cold colors , i e blue. The cold colors seem to disappear for some of the elderly. Therefore they claim that warmer colors are to prefer when designing towards an elderly audience, i e red. Today the golf terminal is built up using different colors of light blue, which as explained above is a cold color.

- **Contrast colors**

  The golf terminal uses, as explained earlier, different shades of light blue. This color, according to [7], can disappear for some of the older adults and can therefore become even harder to detect. This situation could be improved by using contrast colors, to emphasis the difference between a button and a normal background.

- **Big buttons**

  Big buttons in the middle part of the screen represent the choices available. These buttons are having a very clear and distinct highlighting thanks to the center spot on the screen. Thanks to the big buttons the user can easily press the wanted one without having to worry about errors.

- **Different sound when touching a button**

  By using a sound that gives the user feedback when touching a button the user can know if he or she has touched a button or not. This makes selections more obvious [7]. At the moment of writing this article there is only one single sound that is played everytime the screen is touched. It does not matter where on the screen the user touches the golf terminal will leave the same sound signal every time.

  During the user centered questionnaire the sound was discussed. The sound did not affect the users at the point of helping them in their interaction with the golf terminal. This could be because that the sound is the same every time. A special sound for the background, or no sound at all could help the user with the interaction of the golf terminal.

- **Tilted screen**

  The golf terminal is made up of a touch screen connected to a computer. The golf terminal is mounted on a wall in the building of the golf club. The screen is tilted and gives the user a nice view of the system. According to [8] the screen should be tilted somewhere between 30 and 60 degrees. This gives the best angle for the users to perform the tasks necessary to use the golf terminal to the fullest. No exact angle is mentioned as the best angle for using a touch screen but the good part is that the terminal is within the range.

- **Menu tree**

  As long as the user can read it the menu tree will be of great value. It has the possibility of cognitive reduction since the user does not have to think about where he or she is in the menu. The question of how far back the main menu is from where the user is right now will be visible to the user in a simple way. Since older people have problems with where they are in the menu [6] this feature can make a big difference. The questionnaire also gave the answer that the sound fooled some users to think that the menu tree would be clickable. Instead the sound just confused the senior golfers and made even less sence when the menu tree was not clickable.

- **The use of a touch screen**

  The touch screen makes the interaction more direct. As the questionnaire result said the senior golfers have encountered touch screens at different locations in everyday life. This experience makes interaction with the terminal an easy task.

6 Discusion

I think the terminal is a product where the designer(s) have intrdouced some big issues. Why design a terminal with cold colors? Why design it with a non-clickable menu tree? Why not use contrast coloring and shadowing to highlight buttons more than is done at present? Even if I might sound a bit frustrated I think the design team has done a good job because there are a lot of good ideas built into the terminal, and it apparently works as it is.

The youth of today have grown up using computers in many different ways. This has lead to a fearless ambition to try new solutions using computers. The elderly have not got this tool in their back pack. They do not use computer based solutions when dealing with the issues of their daily life. Instead they use the same tools as before which have proven to be sufficient when it comes to solving problems. The times that I have observed the elderly using the terminal has been during their free time. What I mean is that some of them tries the terminal but do so only in a controlled environment, e g at a time where there are fairly few people around. Since most of the golfers play during the day there is not much time left to learn a new tool inprivacy. I mean that the senior golfer of today does not have the patience and time to explore the benefits of a terminal. There is a fear of trying new things. If the old way works, why try something new?

I think that changing things to the better by using the guidelines and re-designing the terminal would make the older adults use the terminal more. The golf terminal is a fantastic technical object and it is a pity that it is not used more than it is today.

References

[1] B. Bergman and U. Rosenhall. Vision and hearing in old age. *Scand Audiol*, 30:255–263, 2001.

[2] H. Christensen. What cognitive changes can be expected with normal aging? *Australian and New Zeeland Journal of Psychiatry*, 35:768–775, 2001.

[3] C. Gonsalves and M. Pichora-Fuller. Application of cognitive theory to training and design solutions for age-related computer use. *Canadian Journal on Aging / La Revue canadienne du vieillissement*, 27(2):145–157, 2008.

[4] M. Karavidas, N. Lim, and S. Katsikas. The effects of computers on older adult users. *Computers in Human Behavior*, 21:697–711, 2005.

[5] Christopher B. Mayhorn, Aideen J. Stronge, Anne Collins McLaughlin, and Wendy A. Rogers. Older adults, computer training, and the systems approach: A formula for success. *Educational Gerontology*, 30(3):185–203, 2004.

[6] S. Mead. Application of cognitive theory to training and design solutions for age-related computer use. *International Journal of Behavioral Development*, 23(3):553–573, 1999.

[7] J. Moller. A design primer for the aging eye. *Long-term living*, 57(6):24–28, 2008.

[8] Atsuo Murata and Rina Takahashi. Optimal slope of touch panel, comarison between young and older adults, 2008.

[9] N. Nusbaum. Aging and Sensory Senecene. *Southern Medical Journal*, 92(3):267–275, 1999.

[10] SCB. Sveriges befolkning efter kön och ålder 31/12/2008, 2009. `http://www.scb.se/Pages/TableAndChart____262459.aspx`, accessed 2009-04-19.

[11] SCB. Tabell mt 9. tillgång till dator i hemmet. personer 16å – 84 år, 2009. `http://www.scb.se/Pages/TableAndChart____49660.aspx`, accessed 2009-04-03.

[12] B. Weinstein. A primer on hearing loss in the elderly. *Generations [0738-7806]*, 27(1):15–19, 2003.

How is fear evoked in video games?

David Sundelin

Abstract: Fiction literature and films has for many years relied upon emotions to deliver an experience. In order to capture the audience, emotions and emotional commitment are important aspects to take in to consideration. Fear is a recurring emotion in the world of entertainment. People watch horror movies, read thriller literature and play survival horror games. Video games are especially interesting since they allow the players of the games to interact with the fictional world, which enables an emotional experience more similar to real life than movies and books. To successfully evoke fear in video games the player has to be put in fear evoking situations. This study will investigate if there exist techniques for evoking fear in video games. Theories about human emotion and general criteria for emotion occurrence are examined. Video game emotions and some existing fear evoking techniques are investigated and discussed. Undeniably there are working techniques for eliciting fear; there are however some drawbacks and other aspects that needs to be taken into consideration.

1 Introduction

Emotion is a phenomenon people deal with daily. In interactions with each other, people use their own knowledge about emotion to predict, interpret, and try to modify moods and emotional responses [23]. Most people know what kind of conditions or circumstances that, for instance, make them feel angry or guilty [10]. In order for emotions to occur there has to exist some kind of triggering event, or factor, commonly referred to as stimulus, or emotional elicitor [16]. Emotions vary on an individual basis because of various reasons such as personality and sex [8, 19].

Fiction literature and films rely much upon emotions to be able to deliver an experience. Relative to these medias one might consider video games a rather new form of entertainment. There are reasons to argue that emotions are important in video games, for example, in terms of expanding demographics, create a "buzz", and financial profits [6]. A main difference between film spectator and video game players is that the video game players are more connected to the fictional world presented by the media. The emotions elicited by the video game are self-directed [25] instead of, as those elicited by films, directed towards a character in the film [17]. Game play emotions are emotions elicited from various actions committed in the fictional world of the video game [22]. Narratives, story, and game perspective are factors which could enhance emotional experience in video games [24, 17].

Fear is an emotion that occurs as a reaction to something threatening. For example a situation where the physical well-being of an individual is in danger [11]. Aristotle wrote two thousand years ago "fear is caused by whatever we feel has great power of destroying us in ways that tend to cause us great pain" [26]. Fear is an emotion one might encounter in fiction literature, films, and video games. The aim for this paper is to investigate if there are any specific methods and/or techniques for evoking fear in the media of video games. In particular, we want to adress the following question:

Q: Are there certain methods/techniques to evoke fear in video games?

In order to answer the question of concern we will first introduce the theory of human emotion, briefly explain what human emotion is and how emotions occur. Subsequently the emotion of fear will be described followed by an analysis of why emotions are important in video games and how they contribute to the medium. Finally existing techniques and methods for eliciting fear in video games will be investigated along with how they work. The paper will be summarized in a discussion where we will try to answer the question of concern.

2 A brief introduction to human emotion

Emotion is a phenomenon people deal with daily. The emotions of their own and other's plays a significant role in daily interactions with each other. Russell believes that emotion is of interest not only for psychologists and cognitive scientist, but also for laymen. This is because people tend to interpret moods of other's, try to foresee others' emotional response, and attempt to modify these emotional responses. In doing so, laymen must rely on their own knowledge about emotion which they have learned, organized, and summarized into a cognitive structure. [23] There are however a great interest on the subject among psychologists and cognitive scientists. There are a lot of research and literature treating the subject of human emotion.

Many studies are describing attempts to establish what human emotions are and which mechanisms that are involved in emotion processing. Many scientists seem to agree upon which ingredients results in human affect. There is one, more or less, shared notion that emotions are both nuanced and complex, that the emotions are affected by feelings and behaviour and that they involve both psychology and cognitions [19, 11]. Furthermore there is one, in literature, recurring definition of emotion which, in short, describes emotions as reactions to events in sense of goals, needs, and/or concerns [11, 12].

Various authors claim that single emotions are hard to express and describe in words. In everyday language there are words such as happy, angry, and frightened to describe current emotional states [3]. Izard argues that it is not possible to completely describe an emotion solely by describing the emotional experience, nor is it possible to describe an emotion completely by electro-physiological measures of occurrences in the brain and the nervous system. To be able to completely describe an emotion, one must take into account three aspects [10];

1. The experience or conscious feeling of emotion.
2. The processes that occur in the brain and nervous system.
3. The observable expressive patterns of emotion.

In literature dealing with emotion and emotional processing one might intermittently stumble upon the terms *basic*, *primary* or *fundamental emotion*. These three terms all refer to one and the same type of emotion which is believed to have a special status. This type of emotion is one of the most recurring notions in emotion literature [20] and is therefore described further.

The notion of basic emotions is in contrast to those who treat emotions, in most respects, as fundamentally similar. This refers to a dimensional model which will be described later on in this paper. Basic emotions infer that there are solus, discrete emotional states such as fear and anger that can be identified by differences in expression, appraisal, behavioural response, etc [4]. If there exist basic emotions is a widely discussed question, some claims that there are and some that there are not. Ortony et al. discusses the matter in the article "What's basic about basic emotions?" where he points out that the most common reason for adopting the notion of basic emotions is to explain some routine observations about emotions. The observations could include the fact that some emotions seems to exist in all cultures and in some higher animals as well, that some emotions appear to be generally associated with and recognizable by characteristic facial expressions [20].

Tomkins argues that there are basic emotions, and has developed this idea in a model called a *differential emotion theory* [26]. This model is one of the two most adopted models in emotion theory. The other is called the dimensional model and will be described later in this paper. The *differential emotion theory* is based on the assumption that there are ten fundamental emotions (fear, anger, joy, distress, etc.) which shape the human motivational system. Each of these emotions has specific properties. Furthermore does each of these emotions lead to different inner experience and behavioural effects. Emotions interact with each other which may result in strengthening or dulling another[10]. Tomkin's model argues that emotions are innate, that they are discrete from one another from a very early age, and each emotion is believed to be packaged with a specific and distinctive set of bodily and facial reactions [26].

As mentioned earlier the dimensional model is a model which stands in contrast to the notion of basic emotions and therefore also the *differential emotion theory*. This approach is the second of the two previously mentioned widely adopted models within human emotion research. The dimensional approach is based on the work of Russell who presents a dimensional, or a *circumplex model for affect*. This model is a two dimensional model consisting of a horizontal axis and a vertical axis. The horizontal axis is representing a scale of pleasure and the vertical axis is representing a scale of arousal. For example would a combination of high pleasure and high arousal result in excitement, and on the opposite side, a combination of low pleasure and low arousal would result in depression [23].

3 Stimulus and emotional reactions

In order to continue this paper, let us now take a closer look on how emotions are evoked. Most people have an idea of what circumstances makes them laugh, cry, or express some specific emotion. People might say that they are afraid of heights, while other people say they are afraid of spiders. Izard points out the fact that people know what kind of conditions or situations that interest them, disgust them or make them feel angry or guilty and that in general, people know what brings about a specific emotion [10].

It is commonly argued in emotion literature that in order for an emotion to be evoked there has to be some kind of triggering factor or event that is evoking the emotion. This factor, or event is generally referred to as stimulus. Lewis, however, calls the stimulus event the "emotional elicitor". He argues that this emotional elicitor must trigger a change in the state of the organism in order for an emotion to occur [16]. The stimulus event or emotional elicitor does not have to be physical, such as a change in the environment or interaction with another human being. The state of the organism can be a change in an idea, or it can be a change in the physiological state of the organism. The triggering event may either be an external or internal. External stimulus may be social, for example separation from a loved one, or nonsocial, a loud noise for instance. Internal stimulus may vary from changes in specific psychological states to complex cognitive activities. A major problem in defining an emotional elicitor is that not all stimuli can be characterized as emotional elicitors. For example, a blast of cold air may cause a drop in body temperature and elicit shivering, but one is reluctant to classify this occurrence as an emotional event. [16]

Norman presents a three level model of design in the book "Emotional design, why we love (or hate) everyday things" [18]. This model is developed with physical industrial design in mind, but the model is applicable on interactive media as well. The model consists of the three levels: behavioural, reflective, and visceral. The reflective level deals with how design reflects upon its' user as well as the personal satisfaction. The behavioral level measures the pleasure and effectiveness of use. The visceral level deals with the simplest and most primitive part of the brain. It cannot reason or compare situations. It operates by what humans are genetically programmed for [18]. The visceral level in Norman's model has some similarities with Pavlov's notion about conditioned and unconditioned reflexes. Many argue that emotions are reactions to events [11, 12]. Some emotions could therefore be viewed as similar to reflexes. Just as laughter could be the result from a joy evoking event, startle and/or freezing could be the result from a fear evoking event. The innate emotion implicitly states that there should be an opposite of innate emotion, that some emotion need some kind of history, that some emotional responses are learned. Pavlov presents a grouping of reflexes, where he separates certain inborn reflexes from other reflexes. The inborn, or unconditioned, reflexes have natural triggers while the others, namely, conditioned reflexes have learned triggers. Pavlov et al. exemplifies this by an experiment where a rat is given a tone, a conditioned stimulus, followed by an electric shock, unconditioned stimulus. When the rat has been exposed to the tone, and hence the shock, the rats' defensive responses will be evoked by the tone. [21]

Ortony writes in his book "The cognitive structure of emotions" that "if an emotion such as distress is a reaction to some undesirable event, the event itself must be construed as undesirable, and because construing the world is a cognitive process, the eliciting conditions of emotions embody the cognitive representation that result from such construal." [19] This leads us to individual differences and individual dissimilar construal. It is commonly known that emotions are not the same for everyone. If some certain event is evoking an emotion in one certain individual, the same event does not necessarily have to evoke the same emotion in another individual. Julie et al. explains that regional brain activity associated with emotion processing can be influenced by a range of individual differences, including differences in personality, dispositional affect, biological sex, and genotype [8]. One definition for emotion, mentioned in the previous section, is that emotions are reactions to events in sense of goals, needs and/or concerns. All three of these aspects could be individual and all three come in play in evoking emotions. Different people have different goals, needs and/or concerns. Ortony et al. exemplifies this by describing when one observes the reactions of the players to the outcome of an important game, for example world cup final in soccer. It is clear that those on the winning team are elated while those on the losing team are devastated. Both the winners and the losers are reacting to the same objective event. It is their construal of the event that is different. [19]

4 Fear

To narrow the scope down and get closer to the subject of discussion the emotion of fear should be studied further. Many researchers seem to be in agreement of what fear is, or at least the effects caused by fear. In emotion literature fear is commonly defined as an emotion that occurs as a reaction to something threatening. For example a situation where the physical well-being of an individual is in danger [11]. This reaction results in a negative state and preparation for action such as escaping or avoiding the danger [11, 2]. In emotion literature advocating basic emotions, fear is commonly listed as one of some number of basic emotions [11]. As mentioned earlier (see section 2), Tomkins argues that some emotions are innate, in other words, that some of our emotions are as if present from birth. He argues that innate activated fear is something that has not changed for many thousands of years, but on the other hand that the experience of fear varies drastically from one individual to another. For example may someone feel fear in a tightening of the throat, while someone else may feel fear in the face and stomach. [26]

There are a few different notions about fear. Some argue that fear is a level of anxiety while others claims that anxiety is a too widely adopted term. Furthermore is intense fear [5, 26] commonly referred to when speaking of, in laymen's terms, being very frightened. This implies that there are different varieties of fear. Fear could be distinguished from the word anxiety because the word anxiety has come to include a too wide range of variety of circumstances capable of eliciting any variety of negative affect. In order to separate fear from anxiety, the emotion of intense fear could be titled terror, or fear-terror. Furthermore could terror be described as designed for emergency motivation of life-and-death significance [26]. There are arguments that fear could be conceived as a network in memory which include data. The data could contain information about the situation, about the responses and about the meaning of the stimulus. This network could be viewed as a program for escaping the dangerous situation [5].

Pavlov's theories about conditioned and unconditioned reflexes is applicable on fear as well as other emotions. Le Doux adopts Pavlov's theory about conditioned and unconditioned reflexes and introduces the term conditioned fear [15]. He exemplifies this by describing an event where someone is bit by the neighbour dog, this person will be on alert every time he walks by his neighbours' house and hence the neighbours' house and yard has become emotional stimuli. Such stimuli, which ordinarily would be meaningless, are turned to signal potential danger on the basis of past experience [13].

We have already pointed out that some argue that fear as a reaction to a threatening event which makes the individual prepare for escape. There is however other reactions associable with fear. Startle and freezing for example. When a conditioned fear stimulus occurs, the subject will typically stop all movement and freeze. This is because many predators respond to movement and withholding movement is often the best thing to do when danger is near. Freezing can also be thought of as preparatory to rapid escape when the coast clears, or to defensive fighting if escape is not possible [14]. As well as for all other emotions, fear varies from one individual to another. Goals, needs and concerns as well as personality, dispositional affect and sex are colouring the emotional process of an individual [8].

5 The role of fear and emotions in video games

Fiction-literature, films and video games are examples of medias which are relying upon human emotions to deliver an experience. Films and literature has been around for many years while video games are a relatively new form of entertainment. There are a lot of arguments why emotions are important for video games. Freeman states that the greatest reason to put emotions in video games is money, in form of profits [6]. He presents nine reasons to put emotions which can lead to greater profits in video games in his book "Creating emotions in games: the craft and art of emotioneering". For instance, Freeman argues that many people never will be lured into playing video games until the games begin to offer the emotional range and depth of the entertainment these people are used to enjoy, such as in films. By merging emotional experiences with video game this could be achieved and the demographics could expand. Furthermore does he argue that by putting emotions in video games they will create a more involving game experience and hence lead to a better "word-of-mouth", or buzz. A better buzz will lead to greater profits. [6]

Video games have much in common with films, and borrow a lot from film theory in designing emotional situations. For instance does Perron state that survival horror games are designed to frighten and scare the gamer and that they do so by relying on horror mythology and conventions of horror movies [22]. Tavinor is describing video games as interactive fictions. By this he means that the premises that ground the fictional worlds of video games are dependent on the role of the player in the game [25]. The experience in film and video games does however have important differences. Tavinor argues that video games differs from film and narrative fictions in that the emotions triggered while playing video games are self-directed [25]. When designing a video game there are many factors to take in to consideration. Matthies agrees both in that video game design borrows a lot from film and film theory, and in that there are differences between movies and video games. In video games the player experiences in a different sense than the spectator of a movie, in video games the player is a part of what is happening. The player is more connected to the fictional world. In movies the spectator is not as connected, instead of connecting to the world, the spectator is identifying more with the character. [17]

Tavinor states that "emotions are involved in the affective framing of fictional worlds, making salient the goals and needs of those fictional worlds, so that our interaction in them is motivated and enhanced". The video game player may for example feel angry at their inability to overcome a monster, frustrated by the difficulty of completing a platform-jumping task, and fearful of possible loss. Emotions appear to guide participation in the video game world by boosting attention and concentration to deal with these challenges [25]. In line with Tavinors arguments about emotion in video games Perron introduces the term game play emotions which are emotions elicited by the game. That is to say fear, fright or dread that arise from the players actions in the fictional world and the consequent reactions of this world. Game play emotions come from various actions: exploring, being lost, fighting, being attacked, feeling trapped, etc. [22].

The backside of video games is that there are no, as Matthies calls it, emotional proxy. As a film spectator you can relate to the film character through his or her emotional responses. For example, if in a move the main characters wife is being killed, the spectator can relate to the main characters emotional response without having to know the wife [17]. This gives rise to the challenge of motivating the player to continue playing.

There is a great deal of different genres of video games. Some games may be story driven while other may be non-story driven. The video game character may for instance be viewed in third person or in first person perspective. Such factors have a huge impact on how a video game is conceived. Schneider conducted a study to investigate how adding narrative changes identification, presence, emotion and motivation. The results from the study made clear that the game players identified significantly more with the characters and felt more present when playing story-based video games compared to non-story-based video games. Schneider et al. draws the conclusion that story makes the players feel more as though they are the game character in the game's world. [24]

There are many factors involved in designing an emotional committing video game. Matthies mentions story, character design, and perspective as examples. He explains that by designing a game in a first person perspective, this means that the player sees the fictional world through the eyes of the character; this makes the player feel more like he is inside the game. By adding a good story, the player will be more motivated to continue playing. The player perspective has to be aligned with the characters' perspective. He argues that a fundamental advantage in video games, in relation to films, is that in a video game the player is participating, it is an experience to be in the game reality [17].

Entertainment experiences are multidimensional and highly dynamic during computer game play, with games potentially eliciting a multitude of different emotions and cognitions varying across time. In video games, there is a dynamic flow of events and action, with games potentially eliciting a large number of different emotions varying across time. [9] Cantor argues that even though people exposed to mass media drama are well understood with what they are watching could not happen in reality, they are still being frightened. She explains that if a stimulus is evoking emotional response, other similar stimuli will evoke similar, less intense emotional response. [2]

6 Fear factors, how fear is elicited in video games

"(I) if an individual conceptualizes a situation in a certain kind of way, then the potential for a particular type of emotion exists" [19].

Norman, among many, speaks of negative and positive emotions. Within the category of positive emotions one would find emotions such as joy and happiness and within the family of negative emotion one would find anger and fear. The visceral level, presented in Norman's three level model of design (see section 3), has some similarity to the innate emotions presented by Tomkins. Norman also lists a number of factors which would result in evoking positive or negative emotions. Warm and comfortable places, temperate climate, harmonious music and sounds are a few situations that would give rise to positive emotions. Similarly sudden, unexpected loud sounds or bright lights, extreme hot or cold, darkness, among others would give rise to negative emotions [18].

Cantor is categorizing three groups of stimuli and events that usually evokes fear in real-life situations and frequently occur in frightening presentations [2].

1. Dangers and injuries which contain natural disasters, animal attacks, large scale crisis etc. Danger is often present when injuries are present. Injuries have therefore come to evoke fear even in absence of danger.

2. Distortion of natural forms such as dwarves, hunchbacks and similar characters that are not distorted by injury is often encountered in thrillers.

3. The experience of endangerment and fear by others in cases where viewer respond indirectly to stimuli through, for example, the protagonist in a movie.

Figure 1: En enemy with a chain saw from the game Resident Evil 5, Capcom 2009.

Freeman claims that there are over fifteen hundred techniques for putting emotions in games [6]. He introduces the term emotioneering, which refers to the expansive body of techniques for evoking emotional breadth and depth in games as well as immersing a player in a role or in a game's world. The techniques fall into 32 categories such as the "role induction technique"; when the hero you play fights for the survival of another, and not just for his or her own survival, the heroes' likability is heightened and willingness to step into that role is increased. Another example is the "first-person deepening technique"; by putting the player where he or she must make tough choices, where these choices have real consequences (like the hero's possible death) it creates emotional depth in the player. It is similar to how, in real life, we grow emotionally by confronting difficult choices [6].

Perron explains that terror is rather identified with the more imaginative and subtle anticipatory dread. It relies more on the unease of the unseen. Terror expands on a longer duration than horror does. Crawling with monsters, survival horror games make wonderful use of surprise, attack, appearances and any other disturbing action that happens without warning. [22]

In the paper "Video Games and the Pleasure of Control" Grodal discusses emotions in video games. He argues that in order to evoke intense emotions, there has to be a stimuli presented that are central concerns of living beings, for instance threats on life or health. Such stimuli will evoke physical arousal. [7] A simple example is put forth; if you suddenly meet a lion on a savannah it would create arousal. The context will determine how the arousal is moulded into an emotion. If you are armed, you may feel aggression and shoot the lion, but if unarmed you might feel fright and look for escape. If you are safely placed in a photo safari jeep, the arousal is transformed in to delight. These emotions are phasic, that is, there is an eliciting cause of arousal, followed by an appreciation of what to do, which then leads to actions that will eventually transform the emotion by removing or transforming the causes of the emotions. [7]

To elicit phasic emotions in relation to fiction a focusing character is needed, because without such character we cannot specify any coping strategies. The emotional experience in a specific situation will be different according to whether it is cued by a film or by a video game [7]. When viewing a film the labeling of the emotion felt is determined by the viewers' passive appreciation of the film characters' coping potentials. But when the situation is part of a video game, it is the players' assessment of his own coping potentials that determines the emotional experience. The unskilled player may fell despair when confronted with a lion, but the skilled player will fuel the arousal into actions. Video games therefore simulate emotions in a form that is more similar to typical real-life experience than film. Emotions are motivators for actions and are labeled according to the players' active coping potentials [7].

Matthies states that there are many factors in evoking certain emotions, such as fear, in video games. He claims that the boo, or startle-effect is one common way to provoke a reaction through surprise [17]. Games have the core elements of the (film) threat scene's startle effect at their disposal: character presence, an implied off-screen threat, and a disturbing intrusion. This is the essential formula (character, implied threat, and intrusion). [1]

To expose the player, or spectator to sudden events is undoubtedly one of the most basic techniques used to scare someone. However, because the effect is considered easy to achieve, it is often labeled as a cheap approach and compared with another more valued one: suspense. As in the well-known example of Alfred Hitchcock, a bomb that suddenly explodes under the table where two people are having an innocent conversation will surprise the spectator for only few seconds at the very moment of the explosion. However, if this spectator is made aware that the bomb is going to explode at any minute, he will participate in the scene and feel suspense for the whole time preceding the explosion. In psychology, the concept of threat is associated with the one of "anticipatory fear" and psychological stress. [22]

Matthies suggests that by opposing traditional techniques, unpleasantness and/or horror could be evoked. Some things are frightening by default, such as the character leather-face in the movie "The texas chainsaw massacre", while something as innocent as a little girl could be more terrifying if presented as fear embodied, for example in the movie "The Ring". Furthermore he believes that colliding emotions could create discomfort and fear. He exemplifies this by describing a scene from the movie "Reservoir dogs", where a man is getting his ear cut off to the tones of the song "Stuck in the middle with you" by Stealers Wheel. This creates something that Matthies calls cognitive dissonance. The eyes are perceiving unpleasantness while the ears perceive something pleasant. This could enhance the feeling of discomfort. The spectator does not know how to handle the situation. [17]

7 Discussion

Q: Are there certain methods/techniques to evoke fear in video games?

A: Yes and no. There are stimuli that in most cases evokes fear. However, people have individual differences and their reactions to the stimulus may therefore vary. To be sure that a situation is eliciting the correct emotion, tests have to be conducted and the results from these tests may imply that a majority may feel fear as a reaction to the situation.

It is safe to say that emotions are reactions to events. Furthermore does it seem plausible to adopt the notion that some stimulus has to exist in order for an emotion to occur, and that emotions vary in sense of, for example, individual goals, needs, and concerns. As Cantor pointed out, that if a situation evokes one specific emotion a similar situation evokes similar, less intense, emotions. Since the total video game experience is getting closer to reality, because of advancements in graphics and sounds, the emotions evoked by video games are becoming more realistic as well.

The different models (*the differential emotion theory* and *the circumplex model for affect*) adopted by various scientists and authors are of little importance when discussing factors involved in evoking emotions. This statement is based on the fact that a resulting emotion, from some triggering event, will be the same regardless of which theory that has been adopted. By investigating these theories I did, however, get a better understanding about how human emotion.

Fear has been defined as an emotion that occurs as a reaction to something threatening, for example in situations where the physical well-being of an individual is in danger. Fear and the effects caused by fear vary in sense of individual differences, just as other emotions do. Personality, sex and genotype are examples of reasons why fear may vary from one individual to another. It is plausible to argue that there are many more factors, aside from the previously mentioned, that colours the individual experience of fear. In the case of video gaming, I believe factors such as video game experience, interest, dedication and skill play a huge role in how an experience is formed. For instance may a novice player experience a video game situation very differently than a skilled player might experience the same situation. An arbitrary emplacement in a video game that is perceived as fearful and exciting for a novice player may be easily overcome by the skilled player.

A number of techniques and factors are presented in the previous section (see section 6). The startle effect is a technique which I am prepared to agree in being both cheap and easy to obtain. I do, however, also believe that it can be a useful technique to some small extent. A shocking or surprising event may enhance a situation or experience, but may also make it feel boring and predictable. The startle effect should be used in small dozes and with care in order to achieve the most out of it. Suspense, however, is a more interesting method. It enables a more long lasting sensation and has a larger spectrum of situations and circumstances where it could be useful and appreciated. Cognitive dissonance, suggested by Matthies, is also an interesting technique. It enables many possible combinations which could sublimate video game experiences. By opposing traditional patterns developers may be able to keep surprising the audience with new stimuli and hence bypass the risk of being predictable and dull.

As suggested, I think it is important to distinguish fear from other negative emotions. Many negative emotions could be related to fear without being thought of as fear. Let us take stress for example, on one hand people might experience stress when running late for work, on the other hand, they may experience stress when escaping from a horde of enemies to the tunes of a stressful melody. The emotions are similar, but also different and should hence be differentiated. By not viewing fear as one emotion, but rather as emotions that come in a variety of forms and intensity it enables the possibility of designing a more nuanced and diverse experience of fear.

All the techniques and factors mentioned may evoke fear, but it is unreasonable to assume that they always will work, and that they do so for the general audience. Because of the wide spectrum of individual differences, one could never assume that a situation certainly will evoke one specific emotion, but rather assume that the majority may react in a certain way.

In the future of fear evoking video games I think a more varied game play will take form. For instance might players of the video game be able to play at opposite teams, that is, have the possibility to play both as the hero and as the villain, in both single-player and multi-player sessions. Virtual reality and alternative controlling techniques will also play a huge role in the future of video games, something that the Nintendo Wii has made clear by using controllers with motion sensing capabilities. Furthermore do I believe that the emotional stimuli will be more individualized in the future. Games will be able to adapt to the current player. If a player, for instance, is afraid of heights the game might present more game play on roof tops.

It seems reasonable to argue that the video game industry is following a similar pattern in developing as the movie industry did. I believe that in coming years video games will be a more common form of entertainment. Graphics, narratives, sound and other video game components will continue to advance and create more realistic experiences. Horror-, thriller, and other games involving fear will probably come and go similarly to as trends in the film industry. But as some genres always remain to some extent so will fear in video games.

8 Future work

This study has had its' main focus on existing techniques for evoking fear in video games. I suggest that the factors addressing the human attraction to fear needs to be investigated in order to further chart the emotional experience of fear in video games. People do obviously enjoy being scared when watching movies, reading books or playing video games. By completely make a survey of the emotional experience I believe existing techniques can be enhanced and new techniques can be invented.

References

[1] Robert Baird. The startle effect: Implications for spectator cognition and media theory. *Film Quarterly, Vol. 53*, 2000.

[2] Joanne Cantor. *Media effects: advances in theory and research.* Lawrence Erlbaum Associates, 2002.

[3] Roddy Cowie and Randolph R. Cornelius. Describing the emotional states that are expressed in speech. *Speech Communication*, 2003.

[4] Paul Ekman. An argument for basic emotions. *Cognition and Emotion, Volume 6*, 1992.

[5] Edna B. Foa and Michael J. Kozak. Emotional processing of fear: exposure to corrective information. *Psychological Bulletin*, 1986.

[6] David Freeman. Creating emotion in games: the craft and art of emotioneering. *Computers in Entertainment (CIE) Vol. 2*, 2004.

[7] Torben Grodal. *Video Games and the Pleasure of Control.* Lawrence Erlbaum Associates, 2000.

[8] Stephan Hamann and Turhan Canli. Individual differences in emotion processing. *Current Opinion in Neurobiology, Volume 14*, 2004.

[9] Jussi Holopainen and Aki Järvinen. The psychophysiology of video gaming:. *Proceedings of DiGRA 2005 Conference: Changing Views Worlds in Play*, 2005.

[10] Carroll Ellis Izard. *Human Emotions.* Springer, 1977.

[11] Julie A. Jacko and Andrew Sears. *The Human-computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications.* Lawrence Erlbaum Associates, 2003.

[12] William James. What is an emotion. *Mind, volume 9*, 1884.

[13] Jennifer M. Jenkins, Keith Oatley, and Nancy L. Stein. *Human emotions: a reader.* Wiley-Blackwell, 1998.

[14] Joseph E. LeDoux. *The emotional brain.* Wiley-Blackwell, 1998.

[15] Joseph E. LeDoux. Emotion circuits in the brain. *Annual Review of Neuroscience Vol. 23*, 2000.

[16] Michael Lewis. *The Emergence of human emotions.* Guilford Press, 2008.

[17] Jens Matthies. Private communication. Interview with Jens Matthies conducted 2009 by David Sundelin, 2009.

[18] Donald A. Norman. *Emotional Design: Why We Love (or Hate) Everyday Things.* Basic Books, 2004.

[19] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The cognitive structure of emotions.* Cambridge University Press, 1990.

[20] Andrew Ortony and T. J. Turner. What's basic about basic emotions? *Psychological review vol. 97*, 1990.

[21] Ivan Petrovich Pavlov and G. V. Anrep. *Conditioned Reflexes.* Courier Dover Publications, 2003.

[22] Bernard Perron. Sign of a threat: The effects of warning systems in survival video games. *COSIGN 2004 Proceedings*, 2004.

[23] James A. Russel. A circumplex model of affect. *Journal of Personality and Social Psychology. Vol 39*, 1980.

[24] Edward F. Schneider, Annie Lang, Mija Shin, and Samuel D. Bradley. Death with a story – how story impacts emotional, motivational, and phcysiological responses to first-person shooter video games. *Human communication research, vol. 30*, 2004.

[25] Grant Tavinor. Video games, fiction, and emotion. *ACM International Conference Proceeding Series; Vol. 123*, 2005.

[26] Silvan S. Tomkins. *Affect Imagery Consciousness: Volume III: The Negative Affects: Anger and Fear.* Springer publishing company, 1991.

Integrating User Centered Design into an agile process

Rickard Tornblad

Abstract: This article investigates how to integrate User Centered Design (UCD) into an agile process. In order to answer the question, case studies are presented, this in purpose to give a clear view of how this integration can be done. The paper will discuss some of the key conflicts between UCD and the agile approach, and give some proposals to solve these conflicts.

1 Introduction

User Centered Design (UCD) and Agile processes are two approaches for managing software development. The UCD process is used by many designers in purpose to design software with good User Experience (UX) and the software industry clearly see the benefit of having a good User Experience in the systems [13] (page 2). The term UCD or User Centered System Design (UCSD) was first announced by Norman et al. [14] and the process aims to fullfill the need of understanding the user when building a system. The two different philosophies of developing software, UCD and agile have many similarities and some conflicts. This paper will investigate some case-studies in this field and sees how research-groups have handled the different methodologies of developing software. This article will deal with the question *how can UCD be integrated into the agile process?*

The software industry realizes the importance of having an iterative process when managing software development. Something that is very common when designing software for customers is that the customers changes the requirements during the process of development. The philosophy of agile is to be able to handle these new requirements from the customer [11]. In 2001 K. Beck, et al. defined agile in the Manifesto for Agile Software Development Listing 3 [6]. Since 2001, a lot of processes has been developed that claim to be agile. The most used agile approaches are, according to Boehm et al. [7], Extreme Programming (XP) and Scrum, which are the two approaches this article will investigate further.

This paper will focus on how to follow agile as a process of developing software and also investigate how User Centered Design (UCD) can be Integrated into to an agile process. Both UCD and agile are considered as philosophies or guidelines to develop and design software. The purpose of this article is to investigate *how these two types of managing software development can support* *each other in order to get a good product.*

The software industry is drifting towards adapting agile as a philosophy of developing software for various reasons which the article will cover in more detail in Section 3. To have a good UX is becoming more important for all software applications. For many designers and researchers it is clear that UCD is a method for developing software with good UX. To get these two philosophies (agile and UCD) to work together is therefor very important.

The article will give a brief introduction to alternative approaches of developing software and then investigate further what UCD and agile is more in detail. To be able to get a better understanding of UCD and agile they will be compared with each oder, and some case studies will be presented where different teams have integrated UCD into the agile process.

2 Background

Over the years there have been many different processes for developing software. In this section some of the other major approaches will be covered in purpose to give some perspective on agile and UCD as tools for developing software.

As the industry of developing software grows mature and more experience is added, it is safe to say that many actors in the industry are adopting new proceses for developing software in purpose to work more efficient.

2.1 Waterfall

Waterfall is a sequential process for developing software, the different phases of the development goes in a sequential order (see Figure 1). The waterfall process was first announced by Royce, the article describes how the process works and in the article Royce says that the if you try to implement this process on a larger system it is "doomed to failure" [15]. Royce does not articulate the term "waterfall process" in this article but this publication is known to be the furst publication of the waterfall approach.

One of the major risks with the waterfall process is, according to Royce, that the testing phase is in the end of the cycle. If the design of the system does not work as planned, a major redesign of the system is required [15].

2.2 V-Model

The V-Model (or Vee-Model) is a process for developing software and it can be seen as an extension of the waterfall process. Instead of moving down the stream as the waterfall model it goes upwards after the coding face to form the characteristic V shape see Figure 2.

One of the big drawbacks with the V-model according to Ansorge et al. “is the lack of an organizational structure and a process allowing exchange of information between deferent project within the organization” [2].

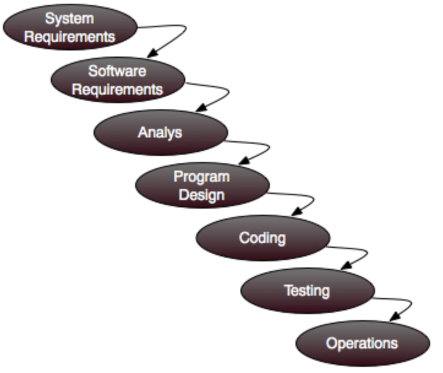

Figure 1: This figure describes each phase of the waterfall model. (adopted from Royce [15])

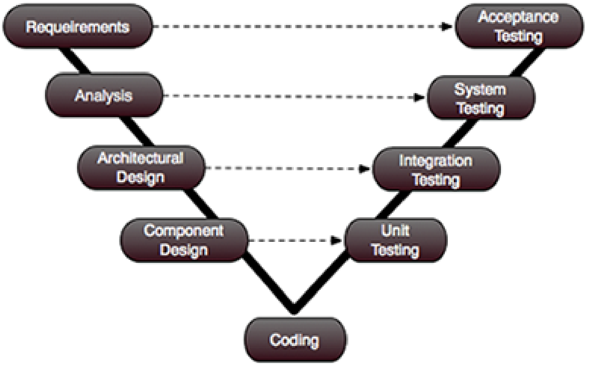

Figure 2: This picture illustrates the V-Model (adopted from Forsberg et al. [8])

2.3 Rational Unified Process (RUP)

RUP is a platform for iterative software development that was created 2003 by Rational Software [10]. RUP itself is not a "ready to go" process. It is rather a platform which an organisation can adopt and modify to fit in to the organization. It is very hard to define RUP because there are many different opinions of what it is but here is the Essential Principles of RUP according to Kroll et al. (see Figure 3).

Attack major risks early and continuously...or they will attack you.
Ensure that you deliver value to your customer.
Stay focused on executable software.
Accommodate change early in the project.
Baseline an executable architecture early on.
Build your system with components.
Work together as one team.
Make quality a way of life, not an after thought

Figure 3: The Spirit of RUP [10](page 5).

3 The Agile approach

Agile is an iterative process for developing software where solutions evolve from cross-functional teams. For a number of years there were many different processes claimed to be "agile", in 2001 the "Agile Alliance" did a publication of the agile manifesto [6] to clear the confusion about **what is agile?**

The agile manifesto

- **Individuals and interactions** over processes and tools
- **Working software** over comprehensive documentation
- **Customer collaboration** over contract negotiation
- **Responding to change** over following a plan

The agile manifesto is the essence of the "Principles behind the Agile Manifesto" wich explains what agile development is. The following list is the principles that define agile proceses [5]. The reasons behind agile development is primary to get a process that deals with the problem of develop software when conditions and requirements change dunring the time of development.

In contrast to traditional development also called "waterfall lifecycle" (see Section 2.1), agile focus on small releases with subset of features for the end release [16]. All the small releases can be seen as subgoals and when they are put together becomes the whole working system. The subgoals are created in fixed intervals that often have duration over 2-4 weeks, in agile terminology they are called "sprints" se Figure 5.

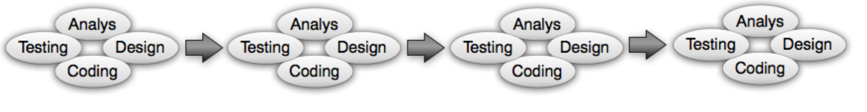

Figure 5: The figure demonstrates the agile proceses. Each cluster in the agile process demonstrates a sprint. (Adapted from Sy et.al [16])

3.1 Extreme Programing (XP)

Extreme Programing (XP) evolved as a method for solving the problem of long development cycles that emerge in the traditional developing methods [1].

One of the core values of XP is to maintain a simple design of the system in purpose to reduce the complexity. To reduce the number of classes and methods to a minimum and to avoid duplicating code, can according to Beck et al. be summarized as "Say everything once and only once". [4]. A feature of XP is to be very responsive to change and to redesign when necessary[1]. The focus in the development is to design for today not for the future.

3.2 Scrum

The term "scrum" comes from rugby and it is a way to restart the game. In software development scrum is an agile process, Hirotaka et al. [17] believes that a more holistic or rugby approach "where a team tries to go the distance as a unit, passing the ball back and forth" would fit better than the traditional way of developing a product.

Scrum fucus on how members in the scrum-team should function in order to produce a good product in a dynamic environment [1]. For each sprint the scrum team will sitt down and break down the end goal of the sprint to different tasks. The tasks that remains will be stored in the "backlogg", for the next scrum meeting the scrum-team will try to make the backlogg to an empty set.

3.3 Comparing XP and Scrum

XP and Scrum are two of the most used agile approaches and here is some major differences between them.

- **Guidance through development**

  XP is a very specific methodology, it contains practical guidance through all phases during the product development.

  Scrum does not have any defined phases.

- **Roles in the development team**

  XP has defined roles in the team.

  Scrum does not have any defined roles in the team.

- **Parallel working teams in one project**

  XP is designed for one development team per project.

  Scrum handels many different teams that works in parallel with the same project.

In short words one might say according to Abrahamsson et al. that XP focus on practices while Scrum is more focused on managing software projects [1].

4 User Centered Design (UCD)

Norman et al. introduced the term UCD in 1986 [14] where they pointed out the importans of having a good understanding of the user and focusing on usability throughout the entire development process. To clear out the confusion on how to adopt UCD (or UCSD) Gulliksen et al. [9] has defined 12 key principles of what UCSD is (see Figure 6).

The list can be seen as guidelines when designing a user-friendly product, rather than a list that should be implemented.

5 Comparison between Agile and UCD

This section will investigate what differentiates agile and USD and what they have in common.

There is a number of things that agile and UCD have in common as processes for managing software development. Iterative development is one of the cornerstones in both agile and UCD. This in purpose to have an ongoing relationship with the customers during the development. One of the many advantages with having a iterative development is to be able to handle new requirements from the costumers during the development.

There is a number of things that separates agile and UCD, *Understand the big picture* is one of the main things were UCD argue that it is important to understand the big picture before starting to implement. The agile approach is to start implementing small chunks (sprints) of working software and not focus on the big picture.

In aspect of *Prototyping* it is importent for UCD to start evaluating using "low-fi designs", where agile argue for implementing working software that is part of the end product.

The *Development focus* is different between agile and UCD were the main focus for UCD is to produce useable software, and agile argue for working software is the primary measure of progress.

In aspect of *designing* the UCD approach is that the development process should contain dedicated design activities, were agile argue that the best architectures, requirements, and designs emerge from self-organizing teams.

This is some of the main things of how agile and UCD differentiate as processes for developing software.

One of the core values of agile software development is to deliver working software Listing 3. This is very important from a usability perspective as well, but with too much focus on the implementation, it can lead to the usability issues being put aside [9].

6 Analyse case-studies

This section presents three different empirical studies on how different teams of developers and UX designers have analysed how UCD and agile can support each other in order to get a good process that are likely to generate good products.

6.1 Agile UCD -Design Studio

This is an empirical study on how to make agile and UCD support each other to become more effective. The experience and findings from this article [18] are largely based on how a UX team work with six scrum-teams over a year.

The chosen way of integrating UCD and Agile is to implement a design studio. The concept of a design studio is often used in the areas of fine art and graphic design. Listed below is there way of implementing a design studio that fits to the purpose of making designer and developer work in an iterative process.

The idea of a design studio is ment to provide a rapid process that includes stakeholders developers and designers. Design studio is a pragmatic example of how it is possible to combine UCD and Scrum (see 3.2). The Design Studio have four main components [18]:

- **Research**

  User-research in order to get the necessary information to base decisions on

- **Design**

  Also known as "pre-work", the goal here is to generate many ideas and explore them, not to refine a single idea.

- **Studio**

  A workshop for a whole day to discuss design alternatives, make decisions and agree on one design

- **Participants**

  A team of designers and non-designers who are willing to learn and grow with the designprocess.

Rapid development of designs are very important for the design studio instead of the "big design up front" approach, which does not fit the agile way of designing software. It is very important for designers and developers to work very tightly together, the designers must be able to produce "lightweight" designs early in each sprint to the developers. To get a good understanding of the end users the design studio argues that it is important that the UX team performs some user research. In order to get the design right, some user research has to be done in purpose to understand the users at a deeper level, before starting to produce working code for the end product [18]. This concept of practice UCD in an agile environment has according to Ungar et al. been very beneficial and the author encourage others to attempt the design studio concept.

6.2 UCD in Agile Projects: Dream Team or Odd Couple?

This is an empirical study where three different designers with extensive experience in UCD investigates how UCD could fit in to an agile process [12]. All the designers that participate in the cases-studie work in different organisations and this is there experience regarding "how does UCD fit in an agile process?". These will refer to the positive and negative effects that has be seen when working in an agile process with a UCD approach. The article have analyzed the cases-studie under four headings, (1) Making the case for UCD, (2) Understanding users, (3) UI design, (4) Evaluating design usability.

Making the case for UCD: The agile approach does not identify a specific role for UCD. This however did not seam to be a big issue for non of the designers in this cases-studie. Instead of having one specific role that is responsible for the design, everybody feels shared responsibility for the design aspects.

Understanding users: This is one of the cornerstones in UCD, it is usually done with a user research before the designing starts. The agile approach is to have close contact with the costumers/users during the development and ask questions when they occur. McInerney el al. recommends to do some upfront analysis of the users before project start.

UI design: Lack of user interface (UI) design ownership in agile means that "everyone wants to be involved in design, which can lead to design by committee". This can lead to arguing in the group about who is responsible for the design.

Evaluating design usability: Testing is very important for both UCD and agile, testing is done in the end of each iteration (sprint). This seams to be a good example of when UCD and agile really support each other.

Overall the UX designers that participated in the empirical study were very positive about using Agile as a process for developing software. One positive effect with agile was that the UX designers "felt actively engaged in a common goal".

6.3 Are agile methods good for design?

This article is written by John Armitage, Armitage has served as the designer for a software development team where they were using XP (see Section 3.1) [3]. The task for the team that John Armitage participated in was to produce a system with high requirements on producing software with good usability.

Based on this experience Armitage has formed some guidelines for designers working in an agile environment Listing 7.

Armitage is skeptical of using XP as a tool for developing software. One of the drawbacks using XP from a design perspective is that agile methods have very rapid iterations. This have the consequence that designers needs to design out the symptoms instead of tackling the core.

Armitage clames that agile methods is not suited to fit all cases of development, “agile methods are best used in cases of exotic technology, volatile requirements, lack of high-level architecture expertise, or lack of high UX standards”.

7 Discussion

As this article presents and from my own experiences many of actors in the software industry are adapting agile processes. One of the main reasons why so many are using agile is that it allows the costumers to change the requirements during the development witch reduces the risk of ending up with a product that the customer is not satisfied with [3].

UCD is the approach that many of the UX designers are used to work with and the main advantage with this approach is that the user is always in center. The primary focus is to develop software with high usability that is well suited for the target-group.

In order to get a good product UCD claims that it is importent to have a good understanding of the users before starting to develop the end product (see Section 4), this is something the agile process does not take as much account of. According to the cases-studies this seems to be a big problem with the agile approach form a UX point of view. There is a risk that this can lead to misunderstandings and wrong priorities during the development [12]. To avoid this I therefore recommend to analyse the users before project start. However another way to integrate UCD into agile in order to understand the user better is the use of personas.

When making importent design decisions, discussions about the different design options in an agile team can arise. The reason why this situation can occur is because of the lack of UI design ownership, and because everyone can have an opinion about the design [12]. This is not necessarily a bad thing when many people within the team feels like they can express there opinion, but this can lead to some frustration for the designer because of the need to convince everyone. This is something a designer should be aware of when working in a agile environment. One of Armitages guidelines for designer working in an agile environment (see Section 6.3) is that “design solutions and work products that can easily be changed” [3], this is one way for a designer to work in an agile environment.

To be in close contact with the user and to preform testing continuously in the end of each iteration is something that UCD and agile have in common. The main difference here would be that the agile approach would focus on functionality, while UCD would focus more on the usability. The iterations however in an agile approach is 2-4 weeks, the UCD approach has not specified a maximum duration for each iteration. According to Armitages designers in a agile team could “become frustrated with fixing symtoms of lager problems that could not be solved within one iteration” [3].

One of the main differences between UCD and the agile approach is there view of understanding the big picture early in the development process. UCD argues about the importance to see the big picture in order to understand how different parts of the system should work. While agile argues for building up the system without focusing on the big picture. As a designer it can be very hard to do important design choices without an overview of the system. One way of facing this problem is to let the designers work ahead of the development according to Ungar et al. [18]. One can argue that the agile view is to have a very unspecified vision of what the system should look like in the end, while the UCD argues for a more explicit vision of what the end result should be. I believe that this comes from the different viewpoints of each approach for managing software development. Where UCDs main concern is to produce an usable system while agiles main concern is to produce a working system.

In this article I have done an investigation of how UCD could be integrated into an agile process. In order to get a clear view, UCD and agile have been presented in detail, some case-studies of how this integration can be solved has been described. It is very clear that the two different approaches for managing software development have very different perspectives. UCD is formed of designers and have the main priority to produce useable software, while agile is formed from a perspective of developers and have the main priority to produce functional software.

As the industry adopt agile approaches more and more it is very important to investigate these approches from a design perspective. The question *how to integrate UCD into an agile process?* is not by any means fully answered by this article and further work needs to be done in this field. This article gives a perpective on some of the key problems of integrating UCD into an agile process. In further work it would be interesting to see more empirical studies on how UCD can be integrate into an agile process.

References

[1] P. Abrahamsson, O. Salo, J. Ronkainen, and J. Warsta. Agile software development methods. *Review and Analysis VTT Publication*, 478, 2002.

[2] D. Ansorge, K. Bergner, B. Deifel, N. Hawlitzky, C. Maier, B. Paech, A. Rausch, M. Sibling, V. Thurner, and S. Vogel. Managing Componentware Development-Software Reuse and the V-Modell Process. *Lecture Notes in Computer Science*, pages 134–148, 1999.

[3] J. Armitage. Are agile methods good for design? *Interactions*, 11(1):14–23, 2004.

[4] K. Beck. Embracing change with extreme programming. *Computer*, 32(10):70–77, 1999.

[5] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunnimgham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, et al. Principles behind the Agile Manifesto. *(2009-05-05) http://agilemanifesto.org/principles/*, 2001.

[6] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, et al. The agile manifesto. *(2009-05-05) http://agilemanifesto.org/*, 2001.

[7] Barry Boehm. Get ready for agile methods, with care. *Computer*, 35(1):64–69.

[8] K. Forsberg, H. Principals, TN Chattanooga, and N. Oslo. The relationship of system engineering to the project cycle. *Center for Systems Management*, 9:11, 1994.

[9] J. Gulliksen, B. Goransson, I. Boivie, S. Blomkvist, J. Persson, and A. Cajander. Key principles for user-centred systems design. *Behaviour & information technology*, 22(6):397–409, 2003.

[10] P. Kroll and P. Kruchten. *The rational unified process made easy: a practitioner's guide to the RUP*. Addison-Wesley Professional, 2003.

[11] M. Lindvall, V. Basili, B. Boehm, P. Costa, K. Dangle, F. Shull, R. Tesoriero, L. Williams, and M. Zelkowitz. Empirical findings in agile methods. *Lecture Notes in Computer Science*, pages 197–207, 2002.

[12] P. McInerney and F. Maurer. UCD in agile projects: dream team or odd couple? *interactions*, 12(6):19–23, 2005.

[13] J. Nielsen. *Usability engineering*. Morgan Kaufmann Publishers, NY, 1994.

[14] D.A. Norman. Cognitive engineering. *User centered system design*, pages 31–61, 1986.

[15] WW Royce. Managing the development of large software systems: concepts and techniques. In *Proceedings of the 9th international conference on Software Engineering*, pages 328–338. IEEE Computer Society Press Los Alamitos, CA, USA, 1987.

[16] D. Sy. Adapting usability investigations for agile user-centered design. *Journal of Usability Studies*, 2(3):112–132, 2007.

[17] H. Takeuchi and I. Nonaka. The new product development game. *Harvard Business Review*, 64(1):137–146, 1986.

[18] J. Ungar and J. White. Agile user centered design: enter the design studio-a case study. 2008.

1. Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. Business people and developers must work together daily throughout the project.
5. Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
6. Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
7. The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
8. Working software is the primary measure of progress.
9. Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
10. Continuous attention to technical excellence and good design enhances agility.
11. Simplicity–the art of maximizing the amount of work not done–is essential.
12. The best architectures, requirements, and designs emerge from self-organizing teams.
13. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

Figure 4: Principles behind the agile manifesto [5]

1. **User focus**-the goals of the activity, the work domain or context of use, the users goals, tasks and needs should control the development.
2. **Active user involvement**-representative users should actively participate, early and continuously throughout the entire development process and throughout the system lifecycle.
3. **Evolutionary systems development**-the systems development should be both iterative and incremental
4. **Simple design representations**-the design must be represented in such ways that it can be easily understood by users and all other stakeholders.
5. **Prototyping**-early and continuously, prototypes should be used to visualize and evaluate ideas and design solutions in cooperation with the users.
6. **Evaluate use in context**-baselined usability goals and design criteria should control the development.
7. **Explicit and conscious design activities**-the development process should contain dedicated design activities.
8. **A professional attitude**-the development process should be conducted by effective multidisciplinary teams.
9. **Usability champion**-usability experts should be involved from the start of project to the very end.
10. **Holistic design**-all aspects that influence the future use situation should be developed in parallel.
11. **Process customization**-the UCSD process must be specified, adapted and implemented locally in each organization. Usability cannot be achieved without a user-centered process. There is, however, no one size-fits-all process.
12. **A user-centered attitude must be established**-UCSD requires a user-centered attitude throughout the project team, the development organization and the client organization

Figure 6: Key principles of UCSD [9]

- Embrace a larger context and judge success by the success of the team or project (which is perhaps the essence of being interdisciplinary).
- Appreciate that providing partial solutions earlier can be more valuable than providing full solutions later on.
- Design solutions and work products that can easily be changed.
- Learn to design the simplest possible version of your idea, and add to it later.
- Think hard about what to design first and what to leave for later.
- Be willing to throw out what is done if it is not working or if it was the wrong thing to do in the first place.
- Learn to quickly jump from low-level to high-level design tasks.
- Branch out from the build iterations to sketch and model an overall vision, yet still respect the learnings from early technical trials.

Figure 7: Guidelines for designers working in an agile environment [3]

Tangible User Interfaces for Teaching Children Language

Erika Vestman

Abstract: Tangible User Interfaces are a new kind of interfaces where the whole environment serves as the interface. This article will explore the possibilities of using Tangible User Interfaces when teaching children language. As children are active and spend most of the day playing, the traditional user interfaces, such as mouse and keyboard, are limiting to the child. When using a mouse or a keyboard the child is required to sit still at a desk. Tangible User Interfaces could combine playing and education which could enhance the child's motivation. To answer the question if Tangible User Interfaces can be used for teaching children language, an investigation of how children learn language has been done and what the designer needs to consider when developing Tangible User Interfaces for children are also discussed.

1 Introduction

Learning language is an important part of life. The child starts speaking in the early years of his or her life and the learning process will continue throughout the whole life. Almost everyday new words will be introduced to him or her and no one will ever be fully learned. Computers have been used in schools for educational purpose for a long time and today almost every child in Sweden has a computer in their home environment. When people think of computer interfaces they often think about traditional interfaces, such as mice, keyboards and displays. Tangible User interfaces (TUIs) are a newer kind of interfaces that use the surroundings of the user as the interface instead of just the traditionally used mouse and keyboard [9]. The user can interact with physical objects to manipulate and interact with the digital world. When using the environment as the interface, the computers will disappear into the periphery of the user. By hiding the computers from the user and bringing the focus to the task rather than to the technique [8], we will get a more ubiquitous society. The TUIs were developed from the Graphical User Interfaces (GUIs), and help to accomplish a more ubiquitous society [8].

Using graspable objects in the physical world to interact with the virtual world has big potential, for instance in the area of children's education [7]. It is an interesting area and TUIs can be further developed to fit the purpose of teaching children language. Some research has already been done on how the TUIs can be used for children's education and many researchers have come to the conclusion that by using TUIs for the purpose of teaching children, we can combine playing and learning and thereby increase the will to learn [5].

Using technologies, such as computers, for educational purpose puts a great responsibility on the designer. When developing an interface, the designer needs to take the user in consideration. The area of human-computer interaction (HCI) argues that the usability of the system is what determines if the user is able to interact with the computer in a satisfying way. It is important to design the computer system to act as a tool, such that the main focus can lie on the specific task and not on manipulating the interface of the tool [17].

The question to be answered in this study is; can Tangible User Interfaces be used for teaching children language? First we will look into the area of human-computer interaction and the importance of usability and affordance. As mentioned earlier the usability is of great importance when it comes to designing interfaces and by using affordance it will be easier for the user to interact with the interface. Affordance invites the user to certain actions. Next we will give some background on the two areas which established the foundation for TUIs, Graphical User Interfaces and Ubiquitous computing. A short study of TUIs and TUIs for children will be presented, and before the final discussion and conclusion we will also look into the fieled of children learning.

2 Human-Computer Interaction

Human-computer interaction (HCI) became commonly known to the public in the early 80s. In the beginning HCI was referred to as man-machine interaction due to the fact that the interaction could take place between a human and any kind of machine. When the focus shifted to the interaction between humans and computers, the expression human-computer interaction became more frequently used [1]. Today the area of HCI is widely known and is something the designer of computer systems needs to be aware of and practice in everyday work. When the computer became more frequently used, the interaction between the human and the computer became even more significant. To be able to develop systems that are easy to use both usability and affordance are of great importance. These subjects will be discussed in the following sections.

2.1 Usability

To make the interaction between the human and the computer successful and fairly fuss-free the designer needs to take the usability of the computer into consideration. In the area of HCI the term usability is often used. The importance of usability is clear but what does it exactly implicate?

In the beginning of 1980 the term "user friendly" was used for describing what we today call usability [3]. The term usability is used in many different fields and has been defined a number of times, but do not have one established definition. One thing that many of the researchers in the HCI field do agree on is that "user friendly" is not a suitable term [3]. It is not about the computer being "friendly" to the user, and what one user believes is "friendly" might not considered to be friendly to another [11].

Usability implicates that the computer does what the user wants without standing in the way while doing so. As mentioned previous, there are many different views on usability and how it should be measured. There is however one definition that is well-defined and descriptive and it is developed by The ESPRIT MUSiC project. They defined usability as [3];

"the ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment; where 'ease of use' affects user performance and satisfaction, and 'acceptability' affects whether or not the product is used". [3]

In other words usability is when the specific system lives up to the expectations of the specific user and gives the user the support that he or she needs at that specific moment, without affecting the user in a negative way.

To achieve usability in a system the designer needs to do extended user tests. What is naturally the best solution at the time, according to the designer, does not have to be the best solution for the user. It is nearly impossible for the designer to know all needs and expectations that the user has, due to the fact that not even the user knows them.

2.2 Affordance

Affordance is often associated with Donald A. Norman, but was first invented by a psychologist called J.J Gibson [12]. Affordance refers to what actions an object invites to. For example, vertical doorhandles afford pulling and flat horizontal doorhandles afford pushing [6]. When Norman is talking about affordance, he likes to call it "perceived affordance". He makes an distinction between "real affordance" and "perceived affordance", and he argues that as designers "we care much more about what the user perceives than what is actually true" [12]. By this he means that designers must develop systems that are designed based on what the user perceives.

3 The foundations for Tangible User Interfaces

As mentioned in the introduction, the fields of Graphical User Interfaces and Ubiquitous computing served as foundations for the Tangible User Interfaces. TUIs help to establish a more ubiquitous society by hiding or masking the interaction tools from the user [8]. Both Ubiquitous computing and Graphical User Interfaces will be discussed next.

3.1 Graphical User Interfaces

The Graphical user interfaces (GUIs) served as a foundation for the TUIs. GUIs were first developed in the early 80s, but were not brought to the public's attention until 1984 by Apple Macintosh [9]. In GUIs there are clear differences between input and output devices. Traditional interface devices, such as mouse and keyboard, presents input and the output is presented by the displays and monitors [8].

An early attempt at the GUI was the "desktop metaphor". This metaphor is used for describing the desktop on our personal computers and is still widely utilized [10]. Even though the "desktop" on our computer has little to do with a real desktop, the metaphor makes it easier for the user to understand how to interact with the computer. The user can save files on the desktop in a similar way as he or she can put papers on a "real life desktop" and when the user throws away a file he or she puts the file in the trash can as he or she would have done in the physical world. Studies have shown that the desktop metaphor is significant superior the conventional user interface that uses menu selection [15].

3.2 Ubiquitous computing

One of the former leading researchers in the field of Ubiquitous computing, Mark Weiser, coined the term when he was working at the Computer Science Lab at Xerox PARC [9]. The term became known when he published his first article "The computer for the 21st Century" in 1991 [18]. Weiser argued that the computer often is in the focus of our attention, even though the focus should be on the task instead. The computer should mainly work as a tool [17]. In Ubiquitous computing the technology disappears into the periphery of the user and Weiser wrote that:

"The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it" [18].

Computers in the ubiquitous paradigm are embedded in the physical world of the user. The computers will be so embedded that the user will not pay any attention to them. In other words, the computer will become invisible to the user. In ubiquitous computing there will be not only one computer embedded in the environment of the user but maybe hundreds or more. Instead of one user interacting with one computer there are many computers that interact with each user. Weiser meant that this could for example result in rooms that greet people by their names, computers that are aware of who is using them and people knowing where other people are located[18].

4 Tangible User Interfaces

The term Tangible User Interfaces was coined in the middle of the 90s by the Tangible Media Group at the MIT Media Laboratory [9]. By using TUIs the area of ubiquitous computing is further developed and a society without visible computers is more likely to be achieved [8].

"To make computing truly ubiquitous and invisible, we seek to establish a new type of HCI that we call "Tangible User Interfaces" (TUIs). TUIs will augment the real physical world by coupling digital information to everyday physical objects and environments" [9]

By making digital information, in forms of bits, graspable, the physical world will be the interface instead of the traditional interface of the computer (see Fig.1)[9]. This digital information in form of bits, are called tangible bits. According to two of the leading researchers in the field, Hiroshi Ishii and Brygg Ullmer, when coupling the digital information with real life objects, such as for example boxes or cards, the world will be the interface. In the article "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms" they are discussing the use of not only real life objects, but also of the architectural environment, such as walls and windows as input and output devices. To be able to use the periphery of the user's perception as a part of the interface, Ishii and Ullmer use ambient media, such as sound and light to provide feedback to the user [9].

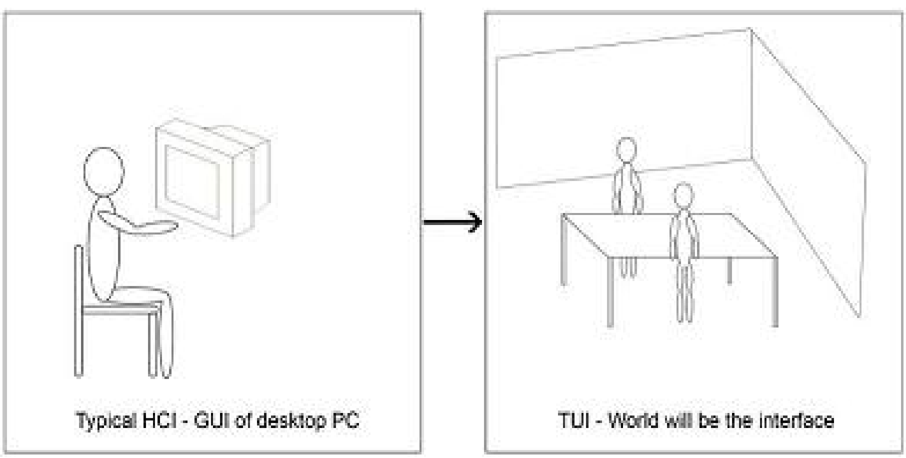

Figure 1: Instead of the typical GUI the world will be the interface. Adapted from [9]

The goal of the tangible bits was defined by the same researchers. According to them it is "to bridge the gaps between both cyberspace and the physical environment, as well as the foreground and the background of human activities." [9]

4.1 Tangible User Interfaces for children

Playing games and having fun are some of the most important aspects of being young. Children are running around and exploring the environment, which is just a part of being a child. By using Tangible User Interfaces we can exploit the natural habits of children and for instance make a learning experience much more intriguing and fun for the child. The TUI gives children the opportunity for action based learning which would combine learning and playing. This can enhance motivation, and according to some theories [13], when combining the learning experience with playing and exploring, the child shows more tendencies to learn. How children learn, and how TUIs can be used for the purpose of teaching children language will be discussed later in this study.

Another advantage of using TUIs as the method for interaction is the fact that very young children do not have the ability to speak and write. These abilities are developed in different later stages of each child's life. Previous studies have shown that the young child also has problems to use the normal interaction tools, such as mouse and keyboard [7]. According to the article "Tangible user interfaces for children" [7], these problems are based on the child's lack of ability to use their fine motor skills. By using physical and virtual objects to communicate knowledge to the child, the child is given additional and different channels to learn. Revelle believes in the potential of TUIs for children and wrote that "Tangible user interfaces, which provide interactivity using real physical objects, hold enormous promise for children." [7]

When designing TUIs for children it is important to consider some basic aspects. For example the designer must take the children's active nature in consideration and develop the interface according to this. Other aspects are, as Antle argues in her article "Tangibles: Five Properties to Consider for Children", that the child needs to understand the mapping between the objects in the physical and the digital world. The child needs to understand why the objects behave in a certain way, how the objects are connected and how they affect each other. Also what the objects represent both in the physical and the virtual world is necessary for the child to understand. Antle brings up five properties that needs to be considered when developing TUIs for children and they are [2]:

1. Space for action: TUIs provide the opportunity for interaction through the whole environment. Larger actions are possible and the user are not required to be close to the computer, as a more traditional user interface would require.

2. Perceptual mappings: This is one of the mappings between physical and virtual objects which occur when using TUIs. The perceptual mapping is

the mapping between the physical and virtual object which can be seen by the user.

3. Behavioral mappings: The child needs to understand why a certain behavior results in an certain effect. Which behavior leads to which effect?

4. Semantic mappings: For the child to understand the meaning of different objects, regardless the representation form.

5. Space for friends: Multiple users makes it possible for children to interact with each other and collaboration and imitation are important parts of the child's learning process. The designer needs to understand how children collaborate and imitate.

Taking all these aspects in consideration can be a challenge for the designer, but is necessary when developing a system with high usability that is suited for children. As mentioned in an previous chapter, the children are suppose to use the computer as a tool for education and the focus needs to lie on the specific task of learning [17].

5 Children learning

To be able to successfully teach children, no matter what the topic is, we need to consider the true characteristics of children and the way they learn [5]. As mentioned earlier, it is in the child's nature to be active and curious and there are no differences when it comes to the learning situation.

Piaget is one of many who argues that there are differences between how an adult and a child think [14], and this will induce different ways of learning. Piaget argued that by constant interaction with the environment the child will develop and learn new abilities and that the intellect of the child develops by change [5]. Another researcher named Tricia David means that for a child to be able to learn he or she needs more than just a demonstration or that someone tells them what to do. To improve the learning process it is important for the child to experiment and execute the task themselves [5]. So the best way for a child to learn is by combining the learning experience with some physical activity or game. By letting the child play and interact with other children the child learns to adapt new skills and behaviors [5]. David also argues that motivation is an important aspect of the learning process of the child. By including the interests of the child and combining them with the education, the child tends to be more motivated to learn and the learning process will get more effective [5].

5.1 Children learning language

The question of how children best learn language has not one definite answer, due to the fact that no child is similar to another [19]. Naturally there are children that have similar ways to learn but there are no guarantees that every part of the learning process is the same.

Known is the fact that children learn language by imitating their parents and other people in their surroundings, but how much the imitation really contributes to the learning process is widely discussed [19]. Wells argues that the imitation is not the primary explanation for how children learn language. He means that the main explanation is the difficulty of the task. How hard the task is to execute has a great influence on the process of learning and he also means that the characteristics of the child have great affect on how well the child learn [19].

According to Wells, the reason for why children learn language is to be able to communicate with the people in the surrounding and that most children rather reciprocate then imitate. This means that the child is more likely to answer questions than giving them, despite the fact that the child probably has heard more question than answers [19]. This is one argument that consolidates why imitating might not be the main explanation for how children learn language.

The Associationism theory argues that the child learn language by connecting a word to a certain object. By showing an object to the child while repeating the name of the object, the child will understand what the name stands for. For example if the parent shows a book to the child, and repeatedly saying the word book, eventually the child will learn[4]. The Theory of Mind is another theory on how children learn language and it is saying that instead of connecting the word to an object, like in Associationism, the child will know what object the parent is referring to. The child is observing everything in his or her surrounding and follows finger pointing and the glance from the parent[4].

6 Discussion

So the question to be answered is if TUIs can be used for teaching children language. We will discuss the advantages and disadvantages when using TUIs for children's education. As mentioned earlier in the study, many researchers do believe that TUIs can be used to enhance the child's learning experience. Children are active and curios and spend big parts of the day playing. If we use this knowledge to combine playing and education, this will lead to more motivated children. The motivation is one great aspect when it comes to learning and needs to be considered when developing systems for teaching. When the child feels motivated and as a part of the process he or she will be more engaged in the education. TUIs also enables collaboration because the interfaces makes it possible for multiple users and this can also lead to enhanced motivation. Small children do not have the ability to speak or write and TUIs makes it possible for very young children to interact with the computer.

By studying how children learn language we can find ways for developing techniques for TUI systems. There are many different ideas on how children learn and how they learn language as mentioned earlier. This topic has been studied for a long time and there are still a lot of mixed opinions. One thing that we believe is that parts of the learning process of the child are developed from imitating. Although the imitating is not the only aspect, it is important enough to take in consideration in this study. One commonly used technique by adults when learning a new language is listening and imitating. This technique could be developed and used in systems with TUIs to teach children language.

Another explanation on how children learn is by mapping words to objects, as we mentioned in previous section. This can easily be implemented in a computer system where TUIs serves as interaction tools. For example when the computer program asks the child for a certain object, the child can get the specific object in the environment. If the computer asks for a book the child will get the book, and if the child is already holding an object, the computer can pronounce and spell the name of it.

However, there are some disadvantages that need to be considered when using TUIs for children education and we will discuss what we believe are the three main disadvantages. One is the fact that it is a challenge for the designer to develop systems with TUIs for children. The designer has a lot of aspects to consider, for example the usability of the system. It is very important that the child understands the mapping between the objects in the physical and the virtual world. It is also important that the child is able to interact with the interfaces and still keep the focus on the specific task. If the interface is too complicated or distracting, the focus will shift to the interaction with the system instead and the actual purpose will be neglected. This is where affordance can be helpful. By using affordance and be aware of the affordance that both physical and virtual things have, the designer will have better chance of succeeding.

Another disadvantage is that systems with TUIs can be expensive, not only to develop but it can be expansive to buy the equipment and the computer programs that are necessary. This is of course dependent on what kind of TUI system that is used and how many interface components that are needed for the computer interaction. It is also dependent on if the equipment is used in the school environment or at home. Naturally the schools have more assets that are earmarked for the purchase of new equipment for educational purpose, which might not be the case at home.

The last disadvantage that will be discussed in this study is that the TUIs can have loose components that are not physical attached to the system. As children likes to play and run around, equipment can easily be lost, which can make the interface useless. These components can often be replaced but this might take time and will cost money.

7 Conclusion

Although the use of TUIs for teaching children is in theory promising and the TUI systems has been even more developed during the last years, the TUIs for this purpose are not very often used in practice. We believe that this is due to the fact that at the moment the disadvantages weight more than the advantages. However, we do believe that TUIs will be further developed and will be applied in the area of education in the future. TUIs do have a lot of potential when it comes to children education and should be something that we exploit.

References

[1] Gregory D. Abowd Alan Dix, Janet Finlay and Russell Beale. *Human-computer interaction.* Pearson Education, 2004.

[2] Alissa N. Antle. Tabgibles:five properties to consider for children. 2009.

[3] Nigel Bevan, Jurek Kirakowski, and Jonathan Maissel. What is usability? In *Proceedings of the 4th International Conference on HCI*, 1991.

[4] Paul Bloom. *How Children Learn the Meaning of Words.* MIT Press, 2000.

[5] Tricia David. *Young children learning.* SAGE, 1999.

[6] William W. Gaver. Technology affordance. *Rank Xerox Cambridge EuroPARC*, 1991.

[7] Allison Druin Glenda Revelle, Oren Zuckerman and Mark Bolas. Tangible user interfaces for children. In *Proceedings of CHI '05 extended abstracts on Human factors in computing systems, Portland*, pages 2051–2052, 2005.

[8] Hiroshi Ishii. The tangible user interface and its evolution. *Communications of the ACM*, 51(6):–79, 1953.

[9] Hiroshi Ishii and Brygg Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of CHI'97 conference on human factors in computing systems, Atlanta, Georgia*, pages 22–27, 1997.

[10] Jeff Johnson, Teresa L. Roberts, William Verplank, David C. Smitha, Charles H. Irby, Marian Beard, and Kevin Mackey. The xerox star:a retrospective. 1989.

[11] Jakob Nielsen. *Usability Engineering.* Morgan Kaufmann, Boston, Massachusetts, 1994.

[12] Donald A. Norman. Affordances and design. 1991.

[13] Jean Piaget. How children form mathematical concepts. *Scientific American*, 189(5):74–79, 1953.

[14] Jean Piaget. Biology and knowledge. *Chicago: University of Chicago Press*, 1971.(Originally published, 1967).

[15] Matthias Rauterberg. An empirical comparison of menu-selection (((cui) and desktop (gui) computer programs carried out by beginners and experts. *Behaviour & Information Technology*, 11(4):227–236, 1992.

[16] Brygg Ullmer and Hiroshi Ishii. *Emerging Framworks for Tangible User Interfaces*. Addison-Wesley, Boston, Massachusetts, 2001.

[17] Mark Weiser. Ubiquitous computing. *IEEE Computer "Hot Topics"*, October 1993.

[18] Mark Weiser. The computer of the 21st century. *Scientific American*, 265(3):94–104, Sept.1991.

[19] Gordon Wells. *The meaning Makers: Children Learning Language and Using Language to Learn*. Hodder and Stoughton, 1987.

Location-aware mobile fitness coaches – Current applications and future possibilities

John Vikström

Abstract: This paper explore positioning-dependent functionality in mobile fitness coaches and current limitations and possibilities. Lately, GPS technology has increasingly more been used in mobile phones and similar devices due to lower manufacturing costs and fewer technological drawbacks – however, usefulness has been limited and there have been interaction challenges of various severity. What this article explores, is the how and why mobile fitness coaches can be viewed as cross-media systems by simply implementing location-awareness, in order to fully take advantage of, and further develop, positioning functionality in such systems.

1 Introduction

This paper will highlight some important characteristics in GPS positioning in general, and existing positioning functions and possibilities in fitness equipment specifically.

With increasing use of mobile, handheld devices, several new areas of interest have emerged concerning the aspect of successful interaction design. One of these areas is mobile, or rather wearable, devices aimed at aiding the user when running or working out through passively providing information such as heart rate, energy consumption or speed. A call to store keepers reveals that sales of so called mobile fitness coaches has grown notably during the last years.

In recent years, GPS technology has been used more and more in the final product to give the end-user increased possibilities (to maximize the results), ahead of, during and after a training session. The devices which uses this technology, and functions or services that depend on it, differ from brand to brand. Another important question under the same topic, is that of technical limitations in GPS systems – however, new algorithms and approaches to effective positioning seems to further strengthen the tendency to bring location awareness to more products. Thereby manufacturers will be able to offer new functionality in already existing devices.

The paper starts by presenting a background to the subject. After that, a general description of what makes a location-aware mobile fitness coach is given. Some common, existing devices and applications are then studied in order to detect similar traits among brands. After this, some important notions on future development possibilities are presented.

1.1 Questions in focus

- What common interaction challenges might occur when introducing location-awareness into a mobile fitness coach?
- What design guidelines could be of use when developing such devices?
- What new possibilities are presented to the user of a location-aware mobile fitness coach?
- What could future applications and solutions in this area look like?

2 Background

The evidence that physical performance strongly affects health and at the same time can prevent health problems are continously growing. Studies have shown that there seems to be a linear dependency between physical performance and personal health [22].

As a result of this, heart rate monitors seems reach new user groups. Several companies now design and manufacture pulse watches with the spoken intent of delivering a simple solution to the users' basic heart rate monitoring needs. But while users get accustomed to these basic functions, manufacturers are trying to look ahead and find new functionality to incorporate into their products.

This is where GPS (Global Positioning System) comes into the picture. It gives, in short, the user new possibilities to plan and analyze his or her training. Several solutions are already on the market, of which two of the main ways of solving the technical aspects are *built-in* or *external* GPS devices which connects to a wristop computer. However, due to technical limitations, the usefulness of location awareness in this type of equipment is often questioned. In order to increase popularity, new technical solutions and functions depending on these, could be developed.

2.1 Location is not context

Mark Weiser wrote an article in 1991, foreseeing a future where the computers work in our background, being invisibly embedded into so called "smart spaces". Humans would not have to directly interact with them – instead they would operate automatically based on current *context* [13].

For the reader of this paper, it is important to recognize the difference between *location* and *context*. As Schmidth et. al. states, *location* can be seen as a *part* of context. A system is *context-aware* if it can use the current context to provide information that is relevant to the user [4]. For instance, a mobile phone that enters silent mode when it senses that a meeting is under way. In this aspect, the phone must register and understand the notions of a meeting – that is, what characterizes a meeting?

Location, though, is a very good starting point for interpreting context. The number of application areas for personal and commercial use is growing, and the future for devices making use of geographic position is bright [5]. Bringing location awareness to mobile devices could therefore create new possibilities that we have not seen today for the user.

2.2 Concerning GPS technology

As mentioned earlier, GPS technology for civil use has in recent years gained popularity and is not longer limited to only being used in one-purpose devices such as handheld devices.

Prices for manufacturing GPS circuits for handheld devices have dropped steadily. This is one reason to why we now find them, among other places, in cameras, mobile phones, built into new cars, in maritime equipment and, as a more unusual example, in games such as [1]. Therefore, wrist watches and mobile or wearable fitness equipment is not an unexpected market that can, and has, made use of this technology.

2.2.1 Limitations in traditional GPS technology

Users' expectations have always been high, but the usefulness of location awareness in mobile devices has several times been proven too low. This might be the reason to why positioning has not reached the popularity manufacturers would have hoped for. Limitations of the "traditional" GPS technology include [14]:

- *Long signal acquiring time* (up to 60 seconds in common devices [6, 8])
- *GPS "signal shadow" due to environmental factors* (e.g. urban landscapes, forests, indoor use etc.)
- *Use decreases battery life of device*
- *Manufacturing costs*

2.3 GPS development

The development of GPS technology has not stalled, and there are new techniques that builds upon previous solutions. Two major breakthroughs has been made during the last decade, Assisted GPS (A-GPS) and eGPS.

2.3.1 A-GPS

By providing the current GPS satellite constellation as a part of the signal, the receiver – for instance, a handheld device for civil use – can get a more accurate position calculation. This is done by using a location server and a mobile station which is able to receive a strong satellite signal. This information is sent to the receiver as "assistance information" [11, 10]. Compared to standard GPS, benefits include [11]:

- Works satisfactory even when used in bad communication areas such as urban landscapes
- Reduced signal acquiring time (30-40 seconds)

2.3.2 eGPS

Even though positioning systems using GNSS (Global Navigation Satellite System) has been developed for several decades, they have been of little or no use in signal-degrading environments. In order to strenghten signal or provide a good-enough position report to the user, the solution needs assistance from other systems such as cellular networks. The eGPS, or Enhanced GPS, uses such cellular signals and calculates position based on timing from satellites. The eGPS builds upon A-GPS but improves performance to a near ubiquitous state. What this means, is that it works indoors and other places where signal strength is zero [17].

According to Rowe et. al., the eGPS solution will provide a position to the user up to three times as fast as A-GPS. Other benefits include low power consumption and manufacturing costs.

3 Devices and applications

Bearing in mind the increasing use of location awareness and technological improvements, what does some common solutions in the field of mobile fitness coaches look like? In this section, some existing devices and applications are reviewed in order to detect characteristic features, and see how knowing ones location might prove useful.

3.1 Location-aware mobile fitness coaches

The notion of a mobile fitness coach is a device that is mobile – or rather wearable – in essence, and that has the necessary sensors to measure factors that are related to physical performance. Primarily, this means measuring the user's *heart rate* and the *duration* of a training session, but there is no need to set limitations here. Other, more advanced, factors may include:

- *VO2-max* (maximum oxygen uptake)
- *Distance cleared* (primarily for cardio-vascular workout) and/or
- *Min/max/average pace* (primarily for cardio-vascular workout)

These devices, for the most part, bear the similar characteristic traits that are listed below.

- Wearable systems consisting of two or three devices (a watch, a heart rate belt and, alternatively, an external GPS sensor)

- Interface navigation using physical buttons
- Small, monochromatic screen with low resolution
- Low computational power

Furthermore, the location-awareness adds position dependent variables such as longitude, latitude and height above sea level (altitude). As a consequence of this, the user is able to record their training sessions by geographic waypoints and retrace their route on a personal computer. Positioning-related functions vary depending on brand, some examples are presented in the next section.

3.2 Devices and applications on the market

Some examples of popular devices that are available on the market today, and of interest for this paper, are the *Suunto T3 with GPS Pod*, *Garmin Forerunner* and *Polar RS800CX with G3 GPS device*. These will in this paper be studied in order to recognize common factors when it comes to features based on location-awareness.

3.2.1 Characteristics of Suunto T3 with GPS sensor

The Suunto T3, according to the manual [19], uses an external GPS sensor in order to provide location awareness to the system. As a first step, the user must "pair" the watch to this device. This must be done to let the two initially separate devices become aware of eachother and be able to send and receive data within the system.

The start of a training session is marked by simply pushing a button. The heartrate belt then starts transmitting data, and the system starts calculating personal training effect based on previously recorded sessions. Other factors that has an impact on this result are weight, length, gender and other personal attributes, according to the user guide.

During workout, the system enables new features since the GPS sensor is present and active. More precisely, the user can now switch between viewing current pulse or current speed/pace.

The Suunto T3 with GPS Pod (as well as the *Garmin Forerunner 305*) has the *AutoPause* feature – this means that the speed and distance calculating system can detect when the user is standing still or moving very slowly, something that is useful if running or cycling through a city with traffic lights. In the end, all time spent standing still is automatically subtracted from the total time of the training session, according to the manual.

The 15 latest training sessions are stored on the watch. To get more detailed information and analyze information more intensely after a workout, the user must purchase the external PC Pod to transfer data to a computer. When using a GPS sensor together with the Suunto T3, total distance, average and maximum speed, distance per lap and average speed per lap is logged in extent to workout data (such as heart rate variations over time).

Figure 1: Suunto T3C with GPS sensor

Figure 2: Garmin Forerunner 305

Figure 3: Polar RS800CX with G3 GPS sensor

3.2.2 Characteristics of Garmin Forerunner 205/305

The Garmin Forerunner 205/305 are examples of models which makes use of a built-in GPS sensor. This eliminates the need of pairing devices to eachother.

As with the other devices eplored in this paper, the satellite signals must be acquired before the user is able to use GPS functions. According to the user manual, this may take up to 30–60 seconds or more [8].

When workout has begun, the Forerunner can provide time and distance alerts, based on user input – an alarm sounds when the preset distance or time is reached. There is, however, a note in the manual that states *"NOTE: Alerts* *do not function during quick, interval or advanced workouts".*

Other alerts can be based on current pace or heart rate, and sounds when the user diverts from a preset pace or heart rate level. The Forerunner manual states that the 205/305 models also makes use of the AutoPause feature, as the Suunto T3 with GPS Pod described earlier.

The Forerunner's *Virtual Partner* function enables the user to compete against a "ghost" of him- or herself. Timing is based on geographic location, and a lap can be marked as a location that the user passes during a, for instance, a run. The information is then presented on the display using colors – if the background is black the user has fallen behind, and if it is white the user is ahead.

Intervals and advanced workouts can also be set up to make use of the GPS, according to the user guide. The user enters how long, specified as distance (or time), he or she wishes an interval should be. The device then asks for the number of repetitions, and sounds an alarm when the goal is reached.

It is also possible to create and edit *courses* with the Forerunner, using the *Garmin Training Center* PC software [9]. When working out, the course map and direction to next waypoint is shown on the display, along with waypoints inserted using the software. The user can compete against him- or herself using the "ghost" function described earlier. This software also enables the user to view recorded training and position data in greater detail.

The GPS functionality is, as with the other devices explored in this paper, also available when not exercising. Viewing the map in order to navigate to waypoints is possible at any time, provided that the GPS signal is strong enough. However, according to [8], accuracy may drop if current speed drops below three kilometers per hour.

3.2.3 Characteristics of Polar RS800CX with GPS sensor

Polar RS800CX offers positioning by using an external GPS device, called G3. As is the case with Suunto T3 and it's GPS Pod, the G3 sensor needs to be paired with the RS800CX in order to work [6].

During workout, the user can switch between several training related views, among which those based on positioning technology are the following:

- the *speed/pace* view
- the *distance* view
- the *altitude* view – total ascent/descent and current altitude in meters above sea level.

After a workout is completed, the user can analyze the exercise results on a computer, using Polars PC software.

4 Discussion

There are several questions raised when introducing location awareness into the area of personal fitness. In this section some personal viewpoints related to the described systems are presented, along with possible future application areas.

Considering the fact that GPS technology is continuously developed, and with the predictions of eGPS about to reach the market, positioning and positioning-related features and devices will probably grow strongly in the coming years. Positioning technology has since long been implemented in some, more expensive, mobile fitness coaches – still, the overall usefulness seems not to have proven high enough to overcome the different drawbacks stated earlier.

Drawbacks of these systems comes down to technical aspects. Specifically, the presentation and accuracy of GPS data is sensitive to changes in the surrounding environment – that is, while working out, the user may pass through a forest or run between high buildings. This, in turn, often has impact on accuracy as the GPS signal weakens. Since solutions to this are underway – through eGPS, for example – this might not be a problem for much longer.

As stated earlier, the time it takes to *acquire satellite signals*, which according to manuals each time may take up to a minute to accomplish [8, 19], may lead to the user to feel impatient and simply not activate the positioning functions. This is something that eGPS might fix, and in turn keep the owner using the device.

Another factor that may reduce usefulness is *power consumption*. When comparing, for instance, the Polar RS800CX with its G3 GPS sensor, the difference between having GPS enabled or disabled is a approximate decrease in battery life with 65 %, from 11h 50m down to 4h 10m (when recording rate is set to one second) [6]. A higher record rate allows for greater accuracy when logging, for instance, a run. A low record rate increases battery life but the "resolution" of the logged track will in turn become lower, and the margin of error in accuracy increase.

Enabling GPS may, in some multipurpose devices such as a mobile phone, *decrease overall performance*. This, however, is not the case in the examples discussed in this paper, since they were designed with location-awareness technology in mind.

To summarize, recent GPS development – like eGPS – seems to fix a lot of current problems with positioning-enabled training coaches. This is likely to lead to more devices having positioning-related features. However, as will be discussed below, this could lead to some interaction issues that are important to be aware of when designing such systems.

4.1 Common features

The devices mentioned earlier bears some rather obvious similarities in functionality. These can be separated in what activity takes place on what platform.

An overview is presented in Table 1 below.

Using LAMFC [1]	**Using computer software**
– View workout history	– View and print workout history
– View waypoints	– View, print, edit waypoints/routes
– Setup advanced workouts	– Analyze terrain impact on performance
– View heartrate	– Possibility to share GPS-data
– View speed/pace	– Create routes based on own preferences
– Use GPS to navigate	

Table 1: Common features of location-aware mobile fitness coach systems

The limited interaction techniques that follows with using a mobile fitness coach are often complemented by use of computer software, where advanced route planning and exercise analyzing can take place.

4.2 Interaction challenges

According to Cooper et.al., the use of devices with small screens can be a challenge in interface design [2]. Several of the factors that were identified as common for a location-aware mobile fitness coach earlier (see 3.1 *Location-aware mobile fitness coaches*) are also recognized by Cooper et.al. – in particular, he feels that the designer must understand that these devices often are not standalone systems. The same conclusion can be drawn for wearable fitness systems, since the most efficient use of a location aware mobile fitness coach is not by interacting with the device itself, but rather lies in effective planning on an a computer with the appropriate software.

What this means, according to us, is that these current devices are not designed to let the owner of a location-aware mobile fitness coach make use of GPS functionality to it's full extent. Rather they seem to try to streamline the training process (plan workout, perform workout, analyze workout) via a computer. The more advanced functions, such as viewing a map, editing waypoints etc., are more easily performed on a computer screen, and the pulse watch should be seen as a "satellite device", as Cooper states [2].

To describe the usage of several technologies as a whole system, Boumans detected five characteristic traits to the term *cross-media* [12] – all of which coincides with the description of a location-aware mobile fitness coach (see Table 1):

1. More than one media is involved in supporting a message/story/goal

2. The aim is on integrated production of support functionality
3. Content is deliverd on multiple devices
4. More than one medium is needed to support the *whole* message/story/goal
5. The common message/story/goal is spread on different platforms

The current, most used designs of location-aware mobile fitness coaches fits this profile according to us, which means care needs to be taken when guiding the user between platforms. To furthermore support this, Stone et.al. states that

> "While the same terminology can be used between handheld applications, you will need to think carefully when adapting an application from a desktop to a handheld device – it is not necessarily the case that terminology that works for a desktop will work for a smaller screened handheld device." [3]

Interaction challenges, techniques and guidelines when using mobile training equipment has been explored before. Based on Nielsens ten heuristic guidelines for succesful interaction design, Esquivias detects important factors for an interaction designer developing systems like these to consider [16, 21].

4.3 What design guidelines could be of use when developing location-aware mobile fitness coaches?

The location-awareness brings much new information to the screens. The possibilities for what the device can do with this data are many. However, the usefulness of all this data limits the amount to the basic information (such as current speed or position) that we see in systems such as those explored in section 3.2.

Nielsen and Molich stated ten rules of thumb in 1990 for evaulation of user interfaces [16]. These are listed here.

1. *Visibility of system status*
2. *Match between system and the real world*
3. *User control and freedom*
4. *Consistency and standards*
5. *Error prevention*
6. *Recognition rather than recall*
7. *Flexibility and efficiency of use*
8. *Aesthetic and minimalist design*

9. *Help users recognize, diagnose and recover from errors*

10. *Help and documentation*

These guidelines can be hard to apply to devices such as a location aware mobile fitness coach. However, when viewed as a traditional small-screened device, then, there are some guidelines that may seem more important and applicable than the others.

Above all, *Consistency and standards* is important. Since the systems discussed most often can be viewed as cross-medial systems as stated earlier, the need for a consistent feel and positioning-related features that depend on the current platform is high, according to us. The user is much more limited in interaction capabilities on a mobile fitness coach than when using a computer. That is why we encourage more focused development from manufacturers when it comes to interface design in crossmedial use, if we fully want to exploit the possibilities of location-awareness in mobile fitness coaches.

In other words, much care must be taken when guiding the user in-between different platforms in order to let him or her make full use of the GPS data.

In the next section, we identify functionality that is unique to location-aware mobile fitness coaches compared to traditional, non-GPS, devices of the same kind.

4.4 What new possibilities are presented to the user of a location aware mobile fitness coach?

Users need motivation for training. This can be everything from just wanting to get som fresh air every now and then, to lose weight or to achieve some personal goal such as a marathon. This, in turn, usually ends up with the user making a schedule or a training plan with some fixed-distance routes.

The social aspects of working out are at the same playing an increasingly important part of an active life. As stated earlier, sales of fitness equipment such as heart rate monitors seems to have grown more rapidly during the last years, according to local store managers. At the same time, we see more and larger communities where people meet and share workout experiences and tips.

One of these is *Funbeat*, a Swedish community with more than ten thousand active users [7]. On the website, they have the possibility to share geographic tracks with each other. This is done by using the Google Maps API and letting the user draw the track in the form of lines on the map. These can be set as public, which lets others users see them. However, this does not make use of an existing GPS recording.

This is what software such as *Trailrunner* for Mac OS X can do [20]. It is a route planning and training software for devices such as the Garmin Forerunner 205/305, which was discussed earlier. The user here has the possibility of importing his or her workout data from a .gpx-file into the computer, which in turn presents a number of available options. The GPX format is an XML file used to describe and transmit GPS information between applications [15].

The user collects geographic data via some GPS-enabled device (according to the website, Trailrunner supports, among other, the Garmin Forerunner 205/305 and the Polar RS800CX discussed in section 3.2) and can later import this information to the computer. As routes are collected and rated, the program can make suggestions on new routes based on user preferences [20]. As the program allows for routes to be recombined in new constellations based on user input, this provides a new way of planning aerobic exercises for greater efficiency. For example: if the user wants steeper tracks to run (or pure hill training), in order to raise heart rate quickly, Trailrunner can give approppriate route suggestions. Furthermore, the user can view his or her maps and even export them to devices that are not location-aware but has the ability to show the map on a display (for example, the Apple iPod Nano).

All topographic data collected during a workout session can be displayed along with recorded heart rate, in order to analyze how an ascent or descent affects physical performance.

The possibilities stated above can act as examples to give an overview of what extra functionality becomes available when implementing location-awareness into a mobile fitness coach and in the computer software.

Next, we will explore some future possibilities for devices like these.

4.5 What could future applications in this area look like?

With better accuracy and response times in GPS systems (which breakthroughs such as A-GPS and eGPS has provided) usefulness of location-awareness in mobile fitness coaches will most likely continue to increase. If so, GPS in fitness equipment can provide further motivation for runners, mountainbikers and similar in a number of ways, mainly because of new features but also, for example, because of reduced time for GPS signal acquiring. Below, we present some examples, bearing in mind the technological development and current applications described in sections 2.2 and 3.2.

- Increasing possibilities for social interaction in all phases of workout
  - *Before workout* – increasing use of virtual communities, both via computers and handheld devices, can provide the option of planning, creating and sharing routes online. This will allow for people with similar training interest and fitness level to more easily meet, exercise and talk.
  - *During workout* – as Esquivias states, the use of music when exercising is strongly increasing [21]. Navigation using sound (music variations) when training is a field of interest, especially as there are several interaction challenges with mobile devices when moving [21]. An example is the gpsTunes system: it guides the user via GPS data through variations in music [18]. Further development in this

field could be possible to enhance the workout experience with music without interfering. Means of locating and navigating to ones friends could also be developed.

  - *After workout* – analyzing recorded fitness and geographic data is the foundation for creating a succesful training plan based on personal goals. We see a great oppurtunity in developing software for location-aware mobile fitness coaches with the possibility of uploading workout data to various virtual communities. The growing number of members on such websites that are devoted to fitness seems to welcome the possibility to share GPS data, even though it may still be limited in usefulness.

- Better chances of more focused workouts for different levels of athletes

  - Beginners can collect routes and see how different terrain affects physical performance. This phase provides an oppurtunity to simply explore ones surroundings, and can provide enough motivation for the user to continue exercising.

    At the same time, more advanced athletes can make use of collected tracks and create new ones by linking routes together. This can aid the user through providing an environment which continuously offers new elements that has an impact on the training session, such as steep hills, long flats or terrain suitable for interval workout. The advanced runner can therefore push his or her limits further, and create tailor-made workouts that targets his or her personal weaknesses (whether that is high intensity training, long duration training, recovering from more demanding training sessions or similar).

5 Conclusion

Technological advancement in the area of personal positioning presents more and more oppurtunities for increasing usefulness of location awareness in mobile fitness coaches. From a social viewpoint, chances of more human-to-human interaction can grow because of more information to discuss either in real life or in virtual communities. Routes can be shared, rated and commented. Regarding fitness, more detailed workout plans can be made to give better chances of greater end-results.

However, as with the systems explored in this paper, the need for a computer is increasing if one wants to take full advantage of all functionality that comes with location-awareness. For example, if the user wants an overview of topographic influence on physical performance, this can only be done if the collected physical and geographical data is imported to a computer. This crossmedial use provides a great interaction design challenge that probably will evolve as the technology it depends on does.

References

[1] Brown M. Chalmers M., Bell B. Gaming on the edge: Using seams in ubicomp games. In *to appear in Proc. ACM Advances in Computer Entertainment (ACE'05)*, pages 306–309, 2005.

[2] Cronin D. Cooper A., Reimann R. *About Face 3: The essentials of interaction design*. Wiley Publishing Inc., Indianapolis, Indiana, 2007.

[3] Mark Woodroie Debbie Stone, Caroline Jarrett and Shailey Minocha. *User interface design and evaluation*. Elsevier Inc., San Fransisco, 2005.

[4] AK Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5:20–24, 2001.

[5] Kaasinen E. User needs for location-aware mobile services. *Personal and Ubiquitous Computing*, 7:70–79, 2003.

[6] Polar Electro. Polar rs800cx manual, 2009. `http://support.polar.fi/gip/PKBStoGIP.nsf/4eb122f6011156bec22573e0003779ed/c225742500419a8ac225747500228955/$FILE/RS800CX%20manual_ENG.pdf`, accessed 2009-05-01.

[7] Funbeat. Funbeat website, 2009. `http://www.funbeat.se`, accessed 2009-04-28.

[8] Garmin. Garmin forerunner 205/305 manual, 2009. `http://www8.garmin.com/manuals/984_OwnersManual.pdf`, accessed 2009-05-01.

[9] Garmin. Garmin training center software homepage, 2009. `http://www8.garmin.com/products/trainingcenter/`, accessed 2009-05-01.

[10] R.E. Richton G.M. Djuknic. Geo-location and assisted gps. *IEEE Computer*, 34(2):123–125, 2001.

[11] Jie Chen et.al. Guolin Sun. Signal processing techniques in network-aided positioning - a survey of state-of-the-art positioning designs. *IEEE Signal Processing Magazine*, 22(4):12–23, 2005.

[12] Boumans J. Cross media: E-content report 8, published in a series of e-content reports by acten, 2009. `http://www.acten.net/`, accessed 2009-04-15.

[13] Weiser M. The computer for 21st century. *Scientific American*, 265(3):94–104, 1991.

[14] Angus P. Andrews Mohinder S. Grewal, Lawrence R. Weill. *Global Positioning Systems, Inertial Navigation, and Integration*. John Wiley and Sons Inc., 2001.

[15] Barnard K. Morris S., Morris A. Digital trail libraries. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 63–71, 2004.

[16] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. *Proc. ACM CHI90 Conf.*, 33(3):249–256, 1990.

[17] Jarvis M.R. Rowe R.W., Duffett-Smith P.J. and Graube N.G. Enhanced gps: The tight integration of received cellular timing signals and gnss receivers for ubiquitous positioning. In *Position, Location and Navigation Symposium, 2008 IEEE/ION*, pages 838–845, 2008.

[18] Murray-Smith R. Strachan S., Eslambolchilar P. gpstunes - controlling navigation via audio feedback. Technical Report Mobile HCI 2005, Dept. of Computing Science, University of Glasgow, 2005.

[19] Suunto. Suunto t3c manual, 2009. `http://media.suunto.com/media/suunto/manuals/en_US/t3c_User_guide_en_v3_m56577569830755179.pdf`, accessed 2009-05-01.

[20] TrailrunnerX. Trailrunner for mac os x website, 2009. `http://trailrunnerx.com/`, accessed 2009-04-29.

[21] Esquivias V. Mobile music fitness coach: Interaction techniques and guidelines. In *Proceedings of Umeå's 12th Student Conference in Computing Science, USCCS'08*, pages 111–122, 2008.

[22] Bredin S. Warburton D., Nicol C. Health benefits of physical activity: the evidence. *CMAJ*, 174(6):801–809, 2006.