OULU 2010
C 355

ACTA UNIVERSITATIS OULUENSIS

Sami Ronkainen

DESIGNING FOR ULTRA-MOBILE INTERACTION

EXPERIENCES AND A METHOD

*FACULTY OF TECHNOLOGY,
DEPARTMENT OF ELECTRICAL AND INFORMATION ENGINEERING,
UNIVERSITY OF OULU*

ACTA UNIVERSITATIS OULUENSIS
C Technica 355

SAMI RONKAINEN

DESIGNING FOR ULTRA-MOBILE INTERACTION
Experiences and a method

Academic dissertation to be presented with the assent of the Faculty of Technology of the University of Oulu for public defence in Auditorium TS101, Linnanmaa, on 28 May 2010, at 12 noon

UNIVERSITY OF OULU, OULU 2010

Acta Univ. Oul. C 355, 2010

Supervised by
Professor Olli Silvén

Reviewed by
Professor Stephen Brewster
Docent Antti Oulasvirta

ISBN 978-951-42-6178-7 (Paperback)
ISBN 978-951-42-6179-4 (PDF)
http://herkules.oulu.fi/isbn9789514261794/
ISSN 0355-3213 (Printed)
ISSN 1796-2226 (Online)
http://herkules.oulu.fi/issn03553213/

Cover design
Raimo Ahonen

JUVENES PRINT
TAMPERE 2010

Ronkainen, Sami, Designing for ultra-mobile interaction. Experiences and a method
Faculty of Technology, Department of Electrical and Information Engineering, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland
Acta Univ. Oul. C 355, 2010
Oulu, Finland

Abstract

Usability methodology has matured into a well-defined, industrially relevant field with its own findings, theories, and tools, with roots in applying information technology to user interfaces ranging from control rooms to computers, and more recently to mobile communications devices. The purpose is regularly to find out the users' goals and to test whether a design fulfils the usability criteria. Properly applied, usability methods provide reliable and repeatable results, and are excellent tools in fine-tuning existing solutions.

The challenges of usability methodologies are in finding new concepts and predicting their characteristics before testing, especially when it comes to the relatively young field of mobile user interfaces. Current usability methods concentrate on utilising available user-interface technologies. They do not provide means to clearly identify, e.g., the potential of auditory or haptic output, or gestural input. Consequently, these new interaction techniques are rarely used, and the long-envisioned useful multimodal user interfaces are yet to appear, despite their assumed and existing potential in mobile devices.

Even the most advocated and well-known multimodal interaction concepts, such as combined manual pointing and natural language processing, have not materialised in applications. An apparent problem is the lack of a way to utilise a usage environment analysis in finding out user requirements that could be met with multimodal user interfaces. To harness the full potential of multimodality, tools to identify both little or unused and overloaded modalities in current usage contexts are needed. Such tools would also help in putting possibly existing novel interaction paradigms in context and pointing out possible deficiencies in them.

In this thesis, a novel framework for analysing the usage environment from a user-centric perspective is presented. Based on the findings, a designer can decide between offering a set of multiple devices utilising various user-interface modalities, or a single device that offers relevant modalities, perhaps by adapting to the usage environment. Furthermore, new techniques for creating mobile user interfaces utilising various modalities are proposed.

The framework has evolved from the experiences gathered from the designs of experimental and actual product-level uni- and multimodal user interface solutions for mobile devices. It has generated novel multimodal interaction and interface techniques that can be used as building blocks in system designs.

Keywords: body user interface, mobile device, multimodal interaction, sound design, usability, user interface, user-centered design

To my family

Preface

The research for this thesis has been conducted in various business units of Nokia during the years 2003–2010.

I would like to express my deepest gratitude to Professor Olli Silvén for supervising this thesis. One could not hope for a better supervisor. Comments to the manuscript were always provided extremely quickly and they were full of insight. I am also much obliged to the pre-examiners of my thesis, Professor Stephen Brewster from the University of Glasgow and Docent Antti Oulasvirta from the Helsinki Institute of Information Technology. Their costructive critique profoundly improved the final version of the thesis.

Next thanks go to my co-authors, Ms. Emilia Koskinen, Ms. Ying Liu, Mr. Panu Korhonen, Dr. Jonna Häkkilä, Dr. René Hexel, Mr. Juha Marila, Ms. Saana Kaleva, Mr. Ashley Colley, Dr. Jukka Linjama and Ms. Leena Pasanen. With Panu I have had the privilege to work both in a research setup and in product creation, both of which are areas where he excels. Panu's group at Nokia Research Center has been the boiling pot for an enormous amount of HCI research. Jonna has been a driving force behind several of the publications that this thesis is based on. Without her enthusiasm this thesis very likely would not exist. Jonna also introduced me to René whose help in creating and presenting Publication II is greatly valued. Another enthusiastic character I have had the joy of working with is Jukka. His vast experience from both science and life has been a source of many interesting discussions. With Juha I have published several studies, only one of which ended up in this thesis. I am especially grateful to him for introducing me to the world of psychological studies and statistical analysis. However, what comes to statistical analysis, I would like to give a special mention to Leena whose mathematical skills were of utmost importance in the creation of Publication V of this thesis. Emilia, Ying, Saana, Ashley – with you it has been a joy to work both as professional colleagues and as co-authors.

I also want to acknowledge my supervisors and group leaders at Nokia who have trusted me to do valuable work for Nokia's business even when one result has been a scientific publication. Mr. Juha Vaihoja, Mr. Timo Tokkonen, Ms. Satu Kalliokulju, Ms. Maini Williams, Ms. Helena Tokkonen, Mr. Urpo Tuomela, Mr. Jukka Parkkinen, Mr Sami Paihonen, Dr. Harri Kiljander, Mr. Hannu Konttinen, and Ms. Sanna Belitz. I also want to honor the memory of the late Mr Hannu Pirskanen who was a great colleague and one of the founding characters of Nokia's user interfaces.

I would like to express my respect to the great scientists who have been both forefigures and friends during my scientific work. Dr. Pertti Huuskonen and Dr. Pekka Isomursu – in addition to the already mentioned Mr. Vaihoja, Mr. Tuomela and Mr. Tokkonen – were the men behind the Research and Technology Access organization at Nokia Mobile Phones Oulu, which was the starting point of all scientific work I have conducted. Dr. Jyri Huopaniemi and people in his teams have been both a bunch of great people and a great source of knowledge. Professor Kaisa Väänänen-Vainio-Mattila I value both as a great person and as one of the top experts in the world what comes to user research. And speaking of top quality user research, I would like to acknowledge Professor Roderick Murray-Smith of the University of Glasgow. His vast scientific knowledge has both widened my views in several areas and introduced me to completely new ones. In this context, I want to thank also Professor Stephen Brewster even though I know I already did.

I also want to thank Mr. Panu Kanervo for his expertise in correcting the grammar of the final edition of the thesis. Thanks are in order also to all of my colleagues in the groups I have worked in. I have been extremely lucky to have enjoyed the company of people who are bright, creative and fun to be with.

Thanks are also in order to all the anonymous reviewers of the publications in this thesis. Further thanks belong to all the participants in our user tests.

My warmest thanks belong to the family that raised me. I am extremely grateful to my mother and father for their unconditional love, support and encouragement. My mother Inkeri always motivated me to take those extra courses at school that I never thought I would really need. That has taught me perseverance and will to go that extra mile needed for a thing like a doctoral thesis. My father Niilo showed me the curiosity to find out why things are the way they are, which is the foundation for all research. My brother Jussi and my sister Anu-Liisa are my siblings, which really says it all. Both the games and the fights were important in learning to deal with social contacts.

Finally, I want to thank my belowed wife and companion, Johanna. I fully realize how largely this thesis is founded on the work of her taking care of the everyday routines and childcare. Children, of course, should not be regarded as obstacles for creating a thesis: I do want to thank also our two handsome sons, Joonas and Topias, for the happiness, joy and the sense of miracle they bring every day to our lives.

"Many strokes overthrow the tallest oaks." - John Lyly c.1554–1606, British writer

Abbreviations

ACM	Association for Computing Machinery
CPS	Characters per Second
DISL	Dialog and Interface Specification Language
EMMA	Extensible Multimodal Annotation
EPIC	Executive Process – Interactive Control
GOMS	Goals, Operators, Methods and Selection rules
HCI	Human-Computer Interaction
HTML	Hypertext Markup Language
I/O	Input/Output
MHP	Model Human Processor
MMI	Multimodal Interaction
MML	Multimodal Markup Language
PDA	Personal Digital Assistant
PPWS	Percentage of Preferred Walking Speed
UI	User Interface
UIML	User Interface Markup Language
UMPC	Ultra-Mobile Personal Computer
SIID	Situationally Induced Impairments and Disabilities
SSKC	Successive Same Key Character
TLX	Task Load Index
W3C	World Wide Web Consortium
XML	Extensible Markup Language

List of original publications

I Ronkainen S, Koskinen E, Liu Y & Korhonen P (2010) Environment analysis as a basis for designing multimodal and multi-device user interfaces. Human-Computer Interaction (in press).

II Ronkainen S, Häkkilä J & Hexel R (2004) Speech-assisted mobile text entry. Proc 10th International Conference on Auditory Display ICAD 2004. Sydney, Australia. URI: http://www.icad.org/Proceedings/2004/RonkainenHakkila2004.pdf.

III Marila J & Ronkainen S (2004) Time-out in user interface: the case of mobile text input. Personal and Ubiquitous Computing 8(2): 110–116.

IV Ronkainen S, Häkkilä J, Kaleva S, Colley A & Linjama J (2007) Tap input as an embedded interaction method for mobile devices. Proc 1st International Conference on Tangible and Embedded Interaction TEI 2007. Baton Rouge, Louisiana, USA: 263–270.

V Ronkainen S, Häkkilä J & Pasanen L (2005) Effect of aesthetics on audio-enhanced graphical buttons. Proc 11th International Conference on Auditory Display ICAD 2005. Limerick, Ireland: 121–126.

VI Häkkilä J & Ronkainen S (2003) Dynamic auditory cues for event importance level. Proc 9th International Conference on Auditory Display ICAD 2003. Boston, MA, USA: 233–237.

VII Ronkainen S (2008) User interface sound design for handhelds – case: internet tablet. In: COST Action IC0601 – Proceedings of the CHI SID Workshop, Florence, Italy: 71–76.

Table of contents

1 Introduction

The mobile devices of today are better described as portable than truly mobile. Typically using them requires the user to stop his current activity, dig the device up from its storage location, and start using his hands and eyes to use it. Some functionalities, such as alerts of incoming calls or messages, can take place when the device is still in its storage. The alert happens through the auditory or tactile modalities. However, even reacting to the alert typically requires manual and visual activity.

The amount of functionalities embedded in mobile devices has increased rapidly, and is likely to keep increasing. The available computing power has grown to enable adding new functionalities such as mobile gaming or navigation. Perhaps even more importantly, the number of services available through mobile devices is growing. The networking and data-processing capabilities in mobile devices are approaching the level that traditionally has only been possible to computers. It has lately become easy, e.g., to follow up the events on one's favourite social networking Web sites, or to receive up-to-date local weather forecasts. Many of these tasks have effects on the user's plans and behaviour, so it will become more and more important to be able to use them while on the move. Navigation is a clear example of such a task; the mobile application guides the user in moving towards a location. But other tasks have similar effects. The calendar application may tell a busy worker where he should be. For a youngster, the social networking applications or messaging applications can be important for finding out what everyone else is doing and where to should one head next.

As a result, the number of times when a user would like to use a function in a truly mobile usage context is likely to increase. At the same time, the wide variation of other activities that take place in such contexts can lead to problems with interacting with the device. It would be desirable if one could perform some activities on the device, without having to stop what one is currently doing, and to switch completely to device use.

User interface design has not kept up very well with the increasing amount of interaction while on the move. Posting a comment to one's Twitter account while walking could be desirable, but is currently quite difficult. In fact, presently even talking to the remote participant of a phone call currently requires the user to hold one of his hands up against the ear. This seems ridiculous. We do not do this when speaking to a person next to us, so it would seem natural that this would not be required of telephone communication either. The headset has been introduced as a solution, but also it has disadvantages. It is uncomfortable to wear one all the time. If the headset is in a pocket when a call comes in, it requires manual interaction to put it onto the ear and answer the call. The headset introduces new aspects to the call handling task, which need to be taken into account when evaluating its usefulness.

Holding a device and operating it is not necessarily even a trivial task. E.g. walking through traffic and carrying a bag poses several restrictions to what one can do with a handheld device. One's eyes and much of the concentration is required for navigating through traffic. The carried items may prevent using the hands for operating a mobile device. Moreover, the weather conditions may require the user, e.g., to wear gloves, which makes it difficult to press small buttons or to operate a touch screen.

In Figure 1, an example of a user in a typical mobile usage context is presented. The environment has a certain illumination level, certain background noise exists (possibly some of which consists of people speaking, affecting the user's cognitive abilities and the performance of a possible speech recognition system), the mobile device is carried in a bag or pocket and so on. The weather may affect what the user is wearing, and the social situation affects what kind of mobile device usage the user considers to be appropriate.

Fig. 1. User in an ultra-mobile usage environment (image from Microsoft Word 2003 clip art © Microsoft).

If, for instance, the mobile phone of the user pictured in Figure 1 suddenly rang, it is clear that reacting to it would require rather drastic operations with traditional user interfaces. The user would have to find the device and pick it up in order to press a button to silence the ringtone. The bags she is carrying would have to be laid down, which might be undesirable if the user fears that they may become dirty as a result. Nevertheless, the pictured usage situation is far from being uncommon today.

The challenges in designing user interfaces for varying mobile usage contexts with limited input/output (I/O) capabilities have already been seen as an important trend in future research (Dunlop & Brewster 2002, Wobbrock 2006, Lumsden & Brewster 2003). Schmidt *et al.* (1999b) define a term, *ultra-mobile computing,* as follows*:* "*Ultra-mobile devices are a new class of small mobile computers, defined as computing devices that are operational and operated while on the move.*" The definition is originally aimed at context-aware computing, but it is useful also for general purpose: an ultra-mobile device is one that is designed to interfere as little as possible with the user's activities in the mobile usage context. The term "ultra-mobile" has been largely overtaken by the ultra-mobile personal computer (UMPC) concept promoted by Microsoft and Intel (Microsoft 2006). However, it has also been used to describe small, mobile devices in general (Wagner & Schmalstieg 2003, Tarasewich *et al.* 2004, Adipat & Chang 2005) and is suitable for describing the problem space that is the focus of this thesis. Interestingly, the UMPC in fact is just a miniaturized personal computer, not at all following the definition of Schmidt *et al.*

As a solution for designing user interfaces for such ultra-mobile devices, multimodal user interfaces have been suggested. For instance, Oviatt *et al.* (2000) claim the following: “They have the potential to expand computing to more challenging applications, be used by a broader spectrum of everyday people, and accommodate more adverse usage conditions than in the past.”

A multimodal system can offer the user a way to utilise a combination of modalities simultaneously, or to switch between modalities depending on the requirements of the task or the environment (Oviatt *et al.* 2000). The utilisation of other interaction techniques than the visual sense and the manual input method is still quite limited in mobile devices. Auditory and tactile modalities are typically used only in situations where the device is alerting the user of, e.g., an incoming phone call. Utilising them, e.g., for augmenting the small graphical display to improve its usability is still quite limited. The biggest advances seem to be made in the field of accessibility, where applications are available for reading the screen contents, allowing eyes-free usage. However, these originate from the needs of people with impairments – not from the demands of the mobile context of use.

Why has multimodality not been utilised more widely in commercial products? The benefits seem intuitively obvious. If one's eyes are occupied, e.g., due to walking in a crowd, then it would seem intuitive that other senses such as hearing could still be used for presenting information. But perhaps it is less clear which modalities should be utilised in which contexts, and how the presentation should be made.

Many of the current multimodal user interface concepts concentrate on enhancing the naturalness of the interaction. A classic example is the “put-that-there” concept, as described by Bolt (1980). Oviatt *et al.* (2000) describe five systems that combine simultaneous spatial pointing and spoken input. Martin *et al.* (1998) have analysed ten multimodal user interfaces, six of which combined spoken input with pointing. Further two of the remaining four combined keyboard-based language input with pointing. It has been shown that in spatial domain, people do prefer to interact multimodally when given the free choice to use either speech or pen input (Oviatt 1997). However, as pointed out by Oviatt

(1999), this kind of interaction is inherited from desktop-like systems, where pointing and object selection are common.

In ultra-mobile computing with today's technology, combined speech and pointing is not necessarily a very good user-interface candidate. Pointing today typically means pointing an object on a touch screen, which requires the usage of the eyes and often both of the user's hands (one to hold the device, the other for pointing). The used speech recognition technology also needs to be quite robust in order to work in the often noisy mobile environments. Obviously pointing can also mean using the handheld device to point to a physical world object (Fitzmaurice 1993), in which case the situation resembles the one described by Bolt (1980) – but this kind of technology is not in everyday use yet.

Wexelblat (1997) points out that utilising hand gestures alone, without speech, has not emerged as an interaction technique either. This is largely true even today, more than ten years after Wexelblat's paper. Wexelblat argues that utilising discrete (as opposed to continuous) gestures is pointless since they are unnatural and their functionality is no more powerful than using a simple button. He then continues to point out further open issues preventing the wide utilisation of gestures in user interfaces.

Technical advances enable designers to create many kinds of interaction concepts using various modalities. Creating and integrating them to commercial products requires money and effort, so there need to be ways to show their benefits. The sheer novelty of a certain concept may support it for a while, but in order to become widely accepted, it has to solve existing usability problems. One must ask: “what is this good for?” in order to avoid a concept from becoming a mere niche. Wexelblat (1997) points out this issue as one problem preventing the wider use of gesture input. It is interesting that in Wexelblat's paper, the lack of a method for analysing the utility of gestures is not pointed out as a problem. Instead, the utility of a discrete gesture is simply labeled as low; Wexelblat argues that it would be all the same to replace the gesture input with a button press. However, a method for analysing interaction problems might reveal that the button press actually requires finger precision that may not be available in the situation. If the user is wearing thick gloves, the ease of grabbing a device and performing a gesture may be a benefit over using a button. Furthermore, with certain technologies, gestures can be recognised without the user having to pick the device up to use it. An accelerometer can detect a gesture being performed on a device sitting in a pocket. A camera can be mounted on a desktop (or a mobile device sitting on the desk) so that it can detect hand or head gestures. It is just a question of finding sensible uses for the gestures.

Sometimes the benefits of the new interaction methods can be easily found. For instance, making a phone call while driving is such a clear concept that speech-recognition based systems have become mainstream in mobile devices. However, in order to come up with new useful concepts, or to test the value of newly invented ones, a method is needed for revealing the current problems in the interaction and the potential solutions for them. Furthermore, there is a need to evaluate whether the new concepts bring new problems into the interaction while solving old ones.

A comprehensive method for evaluating mobile contexts of use by focusing on the available interaction modalities is currently missing. The ISO13407 standard (ISO/IEC 1999) mentions that relevant characteristics of the physical and social environment should be described when specifying the context of use. The question of relevance remains open in the standard. It is not specified how the relevance should be measured or studied.

The majority of HCI literature discusses the problems of user groups and user tasks – fitting the tool to the task. The context of use is often analysed only in reflection to the current task. As an example, the Contextual Design method (Beyer & Holzblatt 1997) describes a work model called *physical model,* which consists of a set of observations of how the physical environment reflects the user's task. It also describes a *cultural model*, which again describes how the task is affected by the attitudes and other cultural aspects. A problem with this kind of an approach is that it is difficult to describe the ever-changing mobile usage contexts, and how exactly they, and the changes in them, affect the use.

Hackos and Redish (1998) describe a process for *user and task analysis,* where the user environment is analysed in more detail. They list variables, such as the existence of noise and dirt, or the typical lighting conditions and temperature in the study context. Those seem to be constructed by everyday experiences – Hackos and Redish mention that the list is not exhaustive but is aimed to be food for thought. The list obviously stems from analysing a workspace environment and places that people have arranged to their needs – just like in the contextual design process.

In ultra-mobile computing, the usage environment can change constantly, and the user rarely has had a chance to arrange the environment according to his needs. Therefore, it is not sufficient to analyse the usage environment only as a reflection of the user's task. There is a need for a deep and detailed analysis of how the different aspects of the environment affect the user or the device he is using.

Tamminen *et al.* (2003) have pointed out the same problem. In their study, the mobile usage context has been analysed by observing people using public transportation. Some of the findings are likely to be common for other mobile contexts as well, but in general the study is concentrated on urban navigation only. It is difficult to tell whether one would in other mobile contexts get similar results that could perhaps be generalized into more common rules.

Pascoe *et al.* (2000) present a way to use tables for describing characteristics of interaction modes in a mobile usage context of field workers studying the behavior of giraffes. In their work, two important principles were found. The first one, of a *minimal attention interface* means that in a real-life context, all the cognitive abilities of the user may not be available for using a mobile device. The second principle, *context awareness*, means that the mobile device is aware of its surroundings. The research, however, concentrates completely on studying field workers. It is probable that many of the findings could be extended to general use of mobile devices, but currently it is unclear how widely that extension could be done. Furthermore, there are aspects in general mobile device use that obviously are not revealed by the study of Pascoe *et al.* For example, the social context was not analysed as it is mostly irrelevant when studying wild animals in the African savanna.

Sears *et al.* (2003) present a framework for describing a context of use from the viewpoints of the user, the environment, and the application that is to be used in the environment. They present the notion of situationally-induced impairments and disabilities (SIID), meaning that due to some restrictions in the context, the user is not able to use all of his resources to use an application. In a similar manner, the application may be designed so that not all the available user resources can be used since, e.g., only unimodal input and output has been made possible. The framework presented by Sears *et al.* appears rather complete in the way that all of the different parts related to the context of use have been listed and their meaning has been analysed. However, the practical steps for showing what is relevant for the user and how, are not described.

In a similar manner to Sears *et al.*, Obrenović *et al.* (2007) point out that the usage environment can induce temporary restrictions to device use. They describe human-computer interaction (HCI) as a set of channels based on different interaction modalities, and the features of the device, the environment, the user and the social situation as filters constraining the usage of these channels. They present some example modalities, for example textual presentation, and which capabilities those require – e.g. visual sensory processing or grouping of lines by proximity. They also provide examples of how the environment affects these modalities. For example, rainy weather causes the user to become more focused and stressed. Obrenović *et al.* present a generic model for interaction constraints. However, what is still lacking is a way to analyse in detail how those restrictions emerge from the usage environment. For instance, Obrenović's model does not reveal what exactly restricts the usage of a textual presentation in the context of, e.g., walking on a street.

Lemmelä *et al.* (2008) present a method for designing multimodal interaction concepts for mobile environments. The method is based on the design team observing target users and analysing the restrictions that the everyday usage contexts place on the HCI. In the presented method, restrictions are classified into five classes where the environment is considered to affect the interaction. The classes are “aural”, “visual”, “physical”, “cognitive” and “social”. The physical class includes all issues affecting on one hand the physical – typically manual – interaction with the device, as well as the possible tactile output the device produces. The method describes how taxing the analysed environments are, with respect to different classes, on a four-point scale. However, the method does not describe in detail from where the restrictions originate in each environment, as the findings are based on the researchers' individual experiences in the studied environments. There is no systematic way of making records of the environment's characteristics and how they cause restrictions to the user's resources.

On the other hand, also usability testing should reveal problems related to how the mobile usage context affects HCI. It has been pointed out that the usability testing of mobile devices should be performed in the real environment (Goodman *et al.* 2004, Nielsen *et al.* 2006, Duh *et al.* 2006) or at least in simulated conditions that are close enough to the intended usage environment (Kjeldskov *et al.* 2004). In a literature review by Kjeldskov and Graham (2003), most of the usability testing of mobile devices and applications presented in the papers they analysed was still conducted in laboratory conditions. The user interfaces of currently available commercial products are quite limited in the way that they utilise all the available resources of the mobile user. Creating and testing a user interface mostly in laboratory conditions is not likely to reveal the cost of interaction in the real world. However, as will be shown later in Chapter 2 of this thesis, mobile-specific usability measures do exist. Perhaps they are not used widely enough when creating commercial products and the lack of rich usage of various modalities results from that. On the other hand, testing reveals problems in interaction but does not necessarily suggest solutions. When usability problems are found, ideas for solving them would benefit from as rich as possible description of the whole context of use.

Thirty years after Bolt (1980) presented the put-that-there concept, multimodal user-interface concepts aimed at ultra-mobile usage contexts are still only emerging. Such concepts could allow the user to switch between modalities at will, and the offered modalities would be selected so that they would be efficient with respect to the characteristics of the ultra-mobile usage contexts. An example of such a system is the Shoogle, presented by Williamson *et al.* (2007). In the system, the user can check, e.g., the state of the device battery by shaking it and feeling or listening to the output, or simply by looking at the device display.

Alternatively, modalities could be used for augmenting each other so that when one modality would be more loaded in a certain usage context, the same information would be provided for the user through other, less burdened modalities. This would also allow switching some modalities off according, e.g., to the social context of the situation. Brewster presents such a system, where both visual and auditory output is used for presenting the state of a graphical button (Brewster *et al.* 1995). In practice, this kind of multimodality exists in today's mobile devices. On a mobile phone, alerting of an incoming call can happen simultaneously through the visual, auditory and tactile modalities. A different set of visual, auditory and tactile alert patterns can be used when a user-defined calendar alert goes off. When the user switches the phone to silent mode, the auditory pattern is omitted from the alert, but the visual and tactile patterns can stay the same.

When designing such systems, a question of how to present the same piece of information on different modalities in a user-understandable manner rises. This kind of cross-modal design has been discussed, e.g., by Hoggan and Brewster (2007) on designing icons that present the same information in the auditory and tactile modalities. The definition of 'crossmodal' in the work of Hoggan and Brewster is related to the presentation of the same information through different modalities. In neuroscience, the same word is used to describe the ways in which stimuli through different senses affect the overall perception of a single event (see, e.g., Driver and Spence (1998) for a discussion of crossmodal attention). This thesis follows the practice in which Hoggan and Brewster use the word.

Crease (2001) has described the design of multimodal widgets, concentrating on presenting the same information using auditory and visual modalities. In the field of auditory interfaces, two design concepts prevail. Especially the concept of *earcons* (Blattner *et al.* 1989) has emerged from the idea of presenting similar information using sound as traditionally was presented by using graphics. Blattner *et al.* state: “Earcons, then, are the aural counterparts of icons.” The other prevailing design paradigm is that of *auditory icons* (Gaver 1986). It, too, originates from the idea of using an auditory presentation for things traditionally presented using graphics. However, Gaver discusses ways of using the qualities of the perceived source of a sound, for mapping the sound to its meaning. Blattner *et al.* concentrate on sounds created using an abstract structure, even though they also mention that representational earcons could be created in a similar manner as in Gaver did.

For commercial products, it is not enough to present information over different modalities in a user-understandable manner. It is equally important to take into account the various aspects that users relate to the products they own and use. Han *et al.* (2001) have presented a set of criteria for product usability evaluation – half of which emerge from the image/impression people want to convey. In the case of mobile user interfaces, especially sound has the quality of conveying information – wanted or not – to people around the user. This has been seen as an obstacle for the auditory user interfaces becoming more popular, and the need for design quality in user interface sound design has been pointed out (Leplâtre & McGregor 2004). Special attention must therefore be paid to the aesthetics of sounds when utilising them in commercial mobile user interfaces. Furthermore, the identity of sounds is extremely important for the brand of the product and the company that built it. These issues limit the ways for creating user-interface sounds. The resulting sounds must carry usability value while simultaneously fulfilling the requirements on product aesthetics and the brand identity.

1.1 Novel contributions of the thesis

This thesis handles the dilemma of finding out obstacles that the ultra-mobile usage context places on the HCI, and finding commercially applicable solutions for overcoming those. The author believes that the lack of wide usage of multimodal mobile user interfaces is that their benefits are not clear. This results at least partially from the lack of a comprehensive and practical tool for analysing the requirements for interaction in various mobile contexts. In this thesis, it is shown that no such tool currently exists, and an analysis framework is presented as a candidate for such a tool. New concepts are presented as possible solutions to certain selected interaction problems. These problems were revealed when performing the study on which the analysis framework is based. Details related to UI sound design and utilising a time-out as an interaction component are presented as tools for creating such concepts. A case study of creating UI sounds for a commercial device is presented. The novel contributions of this thesis are the following:

- A framework for systematically analysing and modeling the mobile usage environment, to reveal which aspects of the environment restrict mobile interaction and how. Based on the analysis, clear design requirements can be created, pointing out how different modalities can be utilised to overcome the restrictions (Publication I)
- New interaction concepts for ultra-mobile usage contexts. A technique utilising graphics and speech as output modalities for entering text in visually demanding situations (Publication II). Comparison of techniques utilising accelerometer-based gesture recognition in different social contexts, as well as a more detailed study of the accelerometer-based double tap gesture. (Publication IV)
- Studies of interaction techniques to be utilised in interaction concepts. Special attention is paid to the commercial utility of techniques. A study on short and long term learning of system time-out lengths and different feedback schemes for the time-out occurrences. (Publication III). A comparison of feedback sounds to touch-screen interactions based on different aesthetic principles. (Publication V). New methods of altering sounds to present different levels of event importance (Publication VI)
- A case study analysing how user interface sounds could be better utilised in commercial handheld products (Publication VII).

Finally, the research done in the context of this thesis has contributed to several experimental and product-level designs.

1.2 Contributions of the author

This thesis is based on seven original publications. The contributions of the publications and the author of this thesis are as follows.

Publication I presents a novel framework for analysing different environments and illustrating how they affect the user's capability to operate a device in each environment. A novelty in the framework is that no previous research shows a systematic and formal way to analyse the usage environment starting from the viewpoint of the user's senses and capabilities. Another novelty in the framework is that it offers a way to model the restrictions caused by the environment. This way it is possible to compare the restrictions in several environments. If similarities are found, a single interaction concept can prove out to be useful in several environments. This is an important aspect when designing commercial devices. The framework also provides a way to create a mapping between characteristics of the environment and the restrictions they cause to the user's capabilities. This information can be used to illustrate how a detected feature of a context is relevant for HCI. This has been a major open question also in context-aware computing.

The contribution of the current author in Publication I is that the author participated in the designing and performing the environment analysis on which the framework is based. The author participated in developing the idea of analysing environments from the perspective of each modality separately. The author also created the idea of the resource space in Publication I. The author was the main responsible for creating the framework and writing the publication.

The second author (Koskinen) of Publication I was responsible for developing and testing of one of the interaction concepts presented in the publication, namely that of the touch screen augmented with haptic feedback. The third author (Liu) of Publication I participated to the environment analysis, especially what comes to analysing the context of bicycling in Beijing. The fourth author (Korhonen) of Publication I was the project manager responsible for the environment analysis project with the analysis framework and results. He prepared the first version of the framework and the publication together with the current author.

Publication II describes a system aimed at enabling almost or completely eyes-free text entry on a mobile device, by utilising both spoken output and non-speech sounds as user feedback for keypad-based text entry. A comparison of laboratory-based and mobile testing of the system was conducted. No previous concepts exist that would have combined manual input with audiovisual output in the same manner in mobile text entry.

The contribution of the current author in Publication II was generating the original concept to be tested, as well as the technical implementation of the test system. The author also conducted the user testing and analysis of the results in co-operation with the first co-author (Häkkilä) of the publication. The current author was the main responsible for writing the Publication II. The co-authors

(Häkkilä and Hexel) of Publication II participated in writing the publication and interpreting the results of the user testing.

Publication III presents a study on using a time-out as an input method, with the specific example of studying its effect on multi-tap mobile text entry on an ITU-T keypad. The study consisted of testing various lengths for the time until a cursor reappears after entering a single character, with presenting the cursor using different modalities.

The contribution of the current author in Publication III was developing the original idea for the study, preparing the demonstrators used in the test and conducting the tests in co-operation with the first author (Marila) of the paper. The current author also participated in preparing the publication.

The first author (Marila) of Publication III was mostly responsible for arranging the user tests and analysing the results, and had the main responsibility of preparing the publication.

Publication IV is a study on an eyes-free input technology, namely that of accelerometer-based gestures performed on a mobile device. The study consisted of finding out the acceptance of gesture input in public, and the usability of the best accepted gestures. Gesture input, while not being a completely novel input paradigm, is still very rarely used in commercial products and the question of how well different gestures will be accepted among users has been open. The study utilised a Web-based questionnaire containing video representations of different gestures in different mobile environments. So far the public acceptance of gesture-based input had not been studied before. The results showed that a simple and discreet gesture of the user tapping the mobile device with his fingers was the best accepted one. This gesture was studied further by arranging usability tests where it was utilised to control various features. The results have also been used in productization of the gesture in Nokia mobile devices.

The contribution of the current author in Publication IV was preparing the Web-based test and analysing the results, as well as participating to the design of the studied gestures and the usability testing of the tap input feature. The current author was also the main responsible for preparing Publication IV.

The co-authors (Häkkilä, Kaleva, Colley and Linjama) of Publication IV created the demonstrators used in the usability testing, and were mostly responsible for arranging the usability tests of the tap input feature. The co-authors also participated in preparing the publication.

Publication V presents a study where a concept of adding auditory feedback to graphical buttons was added, to aid in recovering from input errors, was tested with varying designs for the auditory output. The concept is a replication of an earlier study (Brewster *et al.* 1995) that proved the benefits of the auditory output, but the sounds used in the test were unsuitable for a commercial product. The aim of the study presented in Publication V was to create an aesthetically more pleasing, potentially less audible set of sounds, and verify that the users can gain the same usability benefit from them, and to study whether the new design would be perceived as more pleasing among the users. The novel findings in the study were that the usability benefit remained even with much smaller sounds as in previous studies, and that users almost unanimously preferred the most discreet sound design over a more obtrusive one, and also over complete silence.

The contribution of the current author in Publication V was developing the original idea for the test, as well as designing the new set of sounds and preparing the simulation to be used in the test. The author also conducted the user tests in co-operation with the second author. The current author was also the main responsible for creating Publication V.

The second author (Häkkilä) was the main responsible for arranging the user tests, and also participated in preparing Publication V. The third author (Pasanen) was responsible for the mathematical analysis of the results.

Publication VI presents a study where new parameters affecting the perceived importance of sounds were studied. In the existing literature certain well-known and rigorously studied parameters exist for varying the perceived urgency people relate to a sound. However, sounds generated with those parameters are typically unsuitable for commercial products, as they are often aesthetically unpleasing in nature. The parameters also set strict restrictions on a sound designer, providing only very constrained ways for designing interaction sounds. The aim of the study presented in Publication VI was to find new parameters that would be more easily applicable to product sound design. This approach has not been presented in existing research. The findings revealed some new parameters whose usefulness for this purpose had not been known before.

The contribution of the current author in Publication VI was developing the overall idea of finding new parameters affecting the perceived importance of a sound, and preparing the sounds used in the testing. Planning and conducting the user tests as well as analysing the results were conducted in co-operation with the second author.

The first author (Häkkilä) was the main responsible for preparing Publication VI. Planning and conducting the user tests were done in co-operation with the current author.

Publication VII presents a synthesis of product sound design for commercial handheld devices. The publication presents a new formulation of information from existing knowledge on things affecting sound design and usage of sound, and suggests solutions to the current problems. The publication also points out some new relevant research problems related to the field. As an example, the publication describes the sound design process for a commercial handheld product.

2 Usability research of ultra-mobile devices

The history of ergonomics is probably as old as mankind. When conducting everyday work, humans strive for ease of use, leading to development of better tools. Moreover, work processes can also be improved according to the movement, dimensions and other human characteristics. The word “ergonomics” was defined by Jastrzebowski (1857).

The usability profession emerged in the 1980s with the history of product ergonomics and human factors engineering (Dumas 2007). With the explosion of computer applications, a specific field of human-computer interaction was identified. The first meeting of the ACM HCI special interest group (SIGCHI) was held in Gaithersburg, USA, on March 15–17, 1982. The fact that the work done on computers mostly strains the mind of the user – not the body – led to the definition of “cognitive ergonomics”, emphasizing the difference to traditional ergonomics.

In the 1980s, HCI typically meant interacting with desktop computers or computer terminals. Laptop computers were yet to be invented. Commercial mobile devices were yet to become widely used, even though the first commercial cellular telecommunications systems had been launched in Japan in 1979 and in the Nordic countries in 1981. Information work was done in offices. This background can still be seen in today's usability methods.

The definition of *usability* as stated by the ISO 9241-11 standard (ISO/IEC 1998) is:

> “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.”

The standard further defines *effectiveness* as “accuracy and completeness with which users achieve specified goals”, *efficiency* as “resources expended in relation to the accuracy and completeness with which users achieve goals” and *satisfaction* as “freedom from discomfort, and positive attitudes to the use of the product”.

As can be seen, the definitions do not yet contain methods or metrics related to measuring the usability. Preece *et al.* (1994) present various methods in the Interaction Design: Evaluation chapter of their book “Human Computer Interaction”. These include the thinking-aloud method, which was originally described by Lewis and Rieman (1994). The method requires that the test participants speak constantly aloud what they are thinking when using the tested UI. In this way, the method should reveal if something is misunderstood, not noticed or causes usability problems for some other reason. Preece *et al.* list also quantitative methods such as logging the user's actions in the tested software or performing traditional, empirical experiments with well-defined dependent and independent variables. As other methods, Preece *et al.* describe predictive evaluations such as inspections against specified standards or heuristics or performing expert walkthroughs.

Ecological validity has been pointed out as a problem with usability testing also in traditional HCI (Preece *et al.* 1994). The problem is that the test setup and the environment in which the test is performed affect the results. This problem is emphasized in the field of ultra-mobile computing, since the traditional laboratory setup may be considerably different from walking on a crowded street or riding a bicycle. Usability problems originating from the mobile context of use are not revealed if both the design and testing of the user interfaces are conducted in typical laboratory conditions where the user is sitting down. Intuitively it would seem obvious that mobile-specific usability measures would be needed.

2.1 Comparing mobile and laboratory testing

There has been some debate over whether it pays off to perform usability tests in the field or whether it would be enough to conduct the tests in laboratory settings. A question has been raised whether testing in the field reveals more usability problems than testing in the laboratory. The argument for not testing in the field has been that well designed simulations of conditions resembling the real mobile context may provide relevant results with the benefit of the test conditions being controlled and repeatable.

Kjeldskov and Graham (2003) analysed 102 research papers on mobile HCI, published in the top-level conferences of the field. The study revealed that 71% of testing of mobile devices and applications presented in the analysed papers took place in a laboratory environment. This number includes also the cases where a mobile usage context was simulated in an indoors environment. The authors speculate that this may be because laboratory experiments are easier to set up and conduct than field studies. Another reason seen by the authors is that mobile HCI has roots in computer science and HCI. As a result, the research has concentrated on product functionality instead of revealing issues related to the mobile usage context. It is as if the researchers in the field assumed that problems specific to mobile user interfaces are already well known.

Consequently, much of the potential of available user interface technologies are missed today. The majority of mobile user interfaces is built under the assumption that the user's visual, manual and cognitive resources are available for product use – just as they typically are in a laboratory setting. Kjeldskov and Graham point out the same by saying that it appears that a mobile computer is widely seen as a generally applicable system for a mobile context of use.

Kjeldskov *et al.* (2004) have also compared a PDA-based system for nurses both in a real hospital setting and a laboratory setup resembling the hospital. In their study they found few benefits in testing in the real hospital, except for the fact that automatic actions taken by the device were less understandable in the hospital. The conclusion of testing in the field having no benefits over laboratory testing has been criticised (Nielsen *et al.* 2006, Roto & Oulasvirta 2005) about the laboratory environment being too similar to the hospital environment. Nielsen *et al.* (2006) performed a study where the same tasks were performed on the same equipment both when sitting down in a traditional usability laboratory and by working in an environment similar to a construction site. In their study, they did find significant differences in the number of usability problems found. Moreover, problems related to the user's cognitive load and the interaction style were only found in the field setting.

In a study by Kaikkonen *et al.* (2005) testing a mobile application in a laboratory setting was compared to testing in a mobile usage context. In the study, no significant differences were found between the methods in the number or nature of usability problems found, or task times, or the user's performance. The application used in the study was one running on a Nokia mobile phone with the Series 60 user interface. The test method used in the study was the think-aloud protocol, which reveals high-level usability problems but does not measure low-level performance of the users. Nevertheless, some issues related to the mobile usage context were revealed – e.g. the users sometimes stopped walking and went to a quieter spot at certain times when the task was cognitively demanding.

The possible problems of the think-aloud protocol when used in a field test have been pointed out by Kjeldskov and Stage (2004). In their study, 116 research papers were analysed for usability testing methods of mobile devices. Kjeldskov and Stage present six different methods for laboratory testing and compare the results to those found from testing in a real pedestrian street. The number of usability problems found was not significantly greater in the simulated mobile contexts or the real street context, compared to a situation, where the user was sitting down when using the system. In fact, most usability problems were found in the sitting down condition. The authors expect this to be caused by the think-aloud method that was used to detect usability problems. In a relaxed condition the user can put all his effort to the device use and thinking aloud while in more demanding contexts it was more difficult to think aloud while using the device. In the study, tasks involving motion and navigation led to more errors found related to the user interface layout, and the placement and size of the UI elements, compared to when the user was stationary or walking at a constant speed. Errors in this category include hitting a wrong button or missing a button altogether.

Duh *et al.* (2006) performed a similar comparison on performing usability tests on typical tasks performed on a mobile phone, both in the laboratory and in a train. They used the think-aloud method for collecting usability problems. In their study, testing in the mobile context revealed significantly more usability problems than what could be found in the laboratory. Prominent sources for usability problems in the train setting were the overall noise level, the fact that the train was moving, and the lack of privacy in the train.

Baillie and Schatz (2005) describe the design of a mobile messaging system that utilises both visual and spoken I/O. After testing the system in a laboratory setting, users reported that they did not see much value in the multimodal output, but once having tested the system in the field, the opinions were reversed. In the study, even the performance of the users was better in the field than in the laboratory. It must be noted, though, that apparently the system was not really running on a speech recognition system but the test facilitators listened the spoken commands. Recognition errors therefore did not exist in the performance data. Another interesting finding in the study was that in the laboratory, users tended to stay within one modality of using a device while in the field they mixed and switched modalities.

It seems that traditional usability measures such as the think aloud method have problems when utilised in the ultra-mobile usage context. Kjeldskov and Stage (2004) make a reasonable claim that that the mobile context may be so demanding that simultaneously performing a task on a mobile device and thinking aloud while also, e.g., navigating through traffic is simply too much for the test user. When using traditional usability measures in a mobile setting, there may be interesting findings on the side. Examples of these are the tendency to switch between modalities, as shown by Baillie and Schatz (2005) or the tendency to stop walking when the task gets too demanding, as shown by Kaikkonen *et al.*

(2005). However, finding the causes for the problems may still be speculative. More specific measures are needed for detecting what causes the errors and how severe the effects are.

2.2 Mobile-specific usability measures

The usability of mobile devices and applications is a two-level issue. On one hand, one cannot argue that the application logic and the provided features need to match the expectations and the mental model of the user. This part of the picture can be revealed in traditional, laboratory-based usability testing. But on the other hand, the restrictions caused by the mobile environment may prevent the usage of the application or the device completely. Or they can make the interaction so difficult that the user chooses not to use the device at that time. Or, the environment may cause interrupts to the task flow. Because of those, the user may forget which parts of the task have already been conducted, and it becomes an issue to create the UI so that the current task step is easily recognised. Interrupts may also disturb the usage of certain turn-based interaction methods such as speech recognition. These effects caused by the mobile environment are not revealed in the traditional usability test setup where the user is sitting down, talking aloud.

In a mobile context, thinking aloud may not reveal where the interaction problems originate from. Interacting with the mobile device may disturb the other tasks the user is performing in the context, such as riding a bicycle. This is, of course, a problem that should be revealed by usability testing. However, a user may be able to perform a task equally efficiently both in the laboratory and in the mobile context, but at the cost of an increased workload. These issues will be reflected in the effectiveness, efficiency and satisfaction scores of tested user interfaces, but revealing the underlying reasons for the observed problems require usability measures tailored for the ultra-mobile usage context.

Goodman *et al.* (2004) point out the lack of testing in the field and propose a systematical method for conducting field experiments on mobile applications, in order to remove some of the obstacles researchers typically see for field testing. They point out that in some cases, such as when studying location-based applications, testing in the real environment is necessary to obtain reliable results.

Goodman *et al.* also present a set of usability measures relevant for evaluating location-sensitive applications. Most of these are relevant for other mobile applications, too. Some of the measures are specific to the mobile context of use – for instance, the “percentage of the preferred walking speed” (PPWS).

MacKay *et al.* (2005) compared three different methods for a very basic user interaction: that of navigating through a larger-than-screen document using a touch-screen device. These were compared in three different levels of user mobility: sitting down, walking but stopping to use the device, and using the device while walking. The results showed significantly better performance when sitting than when walking. The performance measure used was the time needed for target selection. There were also significant differences in the user preference between the interaction methods while walking.

In a similar manner, Barnard *et al.* (2007) studied the effects of different lighting conditions and levels of user mobility (sitting and walking) on the user's performance when conducting simple tasks (reading comprehension and word search). The results showed that both variables (lighting level, user mobility) had significant effects on the performance in certain areas, such as the task time or how taxing the users evaluated the tasks to be. The measure used for the perceived task load was the NASA Task Load Index (TLX) (Hart & Staveland 1988).

Price *et al.* (2006) studied the effect of different levels of user mobility (sitting vs. walking) on the performance of a speech-recognition based system. Price *et al.* found that the walking condition increased the error rates. The increase was not due to increased noise level when walking, since the walking condition of the test was conducted on a treadmill (and the treadmill noise was compensated for in the sitting condition). In the walking condition the users were breathing more heavily, which the authors believe to be the cause of more insertion-type recognition errors occurring in the walking condition than in the sitting condition.

Lin *et al.* (2007) studied the usage of stylus-based user interface in different mobility settings (sitting, walking on a treadmill, walking through an obstacle course). Target selection times did not differ between different conditions, but overall task times, error rates and several measures for workload did. This indicates that people are able to perform even in difficult conditions, but at the cost of increased cognitive load. The increased overall task time without increases in the individual target selections suggests that people are able to multitask effectively by switching between the mobile task and the task related to the mobile environment.

This assumption is backed up by the results of a study by Oulasvirta *et al.* (2005). The user's attention and cognitive resources have been shown to switch rapidly between the used device and the environment in a mobile context. This was measured by constantly monitoring the user's gaze and his surroundings using a multi-camera system. In some cases, there was a tenfold increase in the attention dwelling in the user's environment in a mobile context compared to a laboratory condition. During a download of a Web page, users spent on average 5% of their time attending the environment instead of the mobile device, compared to the figure of 51% for a busy street context.

Crossan *et al.* (2008) point out the problems related to mobile usability testing and propose an objective method of utilising an accelerometer built into the device to detect the user's context. Timestamped context information is then compared to the user's performance during tasks of text entry and hitting toucscreen objects. The results show that new information of the details of interaction can be obtained with the method. For instance, the method revealed that tapping touch-screen objects is correlated with the phase of the user's gait. Preferred gait phase angles were found, and the users' accuracy when tapping at those angles was shown to be higher than in other phases. The walking speed was also analysed in the study. However, walking speed was found to be relatively constant for all users. The authors assume this to follow from the rather quiet path that the participants traversed in the test.

The usefulness of the percentage of the preferred walking speed (PPWS) measure is still under debate. Pirhonen *et al.* (2002) found consistent results between the PPWS and other usability measures of tested mobile music players, but Mizobuchi *et al.* (2005) only found a text entry task to only slow down the PPWS in general – no differences were found in the PPWS between text entry tasks of varying difficulty (proven with other usability methods). Furthermore, in the study by Mizobuchi *et al.* walking speed did not affect the text entry performance of each individual user. The result of Mizobuchi *et al.* has also been criticised (Lin *et al.* 2007) for the walking task being too easy.

A widely studied mobile environment is that of driving a car. In addition to studying the driving task itself, the performance of the driver while simultaneously operating a mobile device has been examined. The direction of the user's gaze and the duration of fixations have been used as one measure (Sodhi *et al.* 2002, Nunes & Recarte 2002). Variations in the lateral position of the car on the lane have been used as a measure for the driving performance (Salvucci 2001).

Measuring the workload of the user during a task has also been used as a usability measure. In her book “Human Performance, Workload and Situational Awareness Measures Handbook”, Gawron (2008) presents over 100 different measurement methods. Not many of these have been applied to analysing mobile user interfaces, but many of them certainly could. As an example, the Nasa Task Load Index (Hart & Staveland 1988) is described in Gawron's book, and has also been widely used when testing mobile user interfaces.

To summarize, the different measures that have been seen relevant for mobile usability testing in the described literature (Goodman *et al.* 2004, Kjeldskov & Stage 2004, MacKay *et al.* 2005, Oulasvirta *et al.* 2005, Sodhi *et al.* 2002, Nunes & Recarte 2002, Salvucci 2001) are presented in Table 1.

Table 1. Mobile usability measures (adapted from Goodman *et al.* (2004), published by permission of authors. By request, parts added by the current author are emphasized with *italics*)

Measure	What does it tell us?	How is it measured?
Timings	Performance. *In some conditions, difficulty of interaction (e.g. Fitts's Law for pointing).*	By the experimenter via a watch or stopwatch. Automatically by the device.
Errors	Performance. Particular sticking points in a task or design.	By success in completing the tasks or answering questions correctly, through experimenter observation, examining the route walked and the buttons pressed.
Performance (CPS, tapping accuracyt) linked to walking movement	*Details of how the user's performance is affected by his or her movement.*	*Using an accelerometer to record movement, and analysis algorithms for showing the details of the user's gait, or classifying movement into contexts such as “sitting”. Analysing timestamped information of performance and movement.*
Perceived workload	Effort invested. User satisfaction.	Through NASA TLX scales and other questionnaires and interviews.
Distance travelled and route taken	Depending on the application, these can be used to pinpoint errors and to indicate performance.	Using a pedometer, GPS or other location-sensing system. By experimenter observation.

Measure	What does it tell us?	How is it measured?
Percentage preferred walking speed (PPWS)	Performance – *but limitations may exist on reliability of this measure.*	By dividing distance travelled by time to obtain walking speed, which is then compared with normal walking speed.
Comfort	User satisfaction. Device acceptability.	Using the Comfort Rating Scale and other questionnaires and interviews.
User comments and preferences	User satisfaction and preferences. Particular sticking points in a task or design.	Through questionnaires, interviews and think-alouds. Using open-ended questions and forced choice.
Experimenter observations	Many different aspects, depending on the experimenter and on the observations.	Through observation and note-taking by the experimenter.
Glance direction, duration, variability	*Where the user's attention is focused. How much the user's eyes can be occupied with device use.*	*Using a gaze-tracking system or manually by recording a video stream of the user's face.*
Driving performance	*Difficulty of in-car applications.*	*Measuring average lateral deviation and average lateral speed.*

As can be seen, usability methods relevant for the mobile context of use exist, even though they may not be as widely used as they should (Kjeldskov & Graham 2003).

2.3 Cognitive models for testing mobile user interfaces

Another way to study the usability of a user interface is to use cognitive models. The philosophy behind cognitive models is to create a detailed enough model of the human perception and information processing systems, and to analyse how various aspects of the task affect those. In a multitasking concept, the task must be considered to be a combination of the overlapping tasks.

Preece *et al.* present the Keystroke Level Model as an example of a cognitive model. It can be used when the user interface is based on manual interaction, as it depends on giving cost values to atomic interactions related to manual use, and calculating the total cost of specified tasks. The Keystroke Level Model is a special case of the generic Goals, Operators, Methods and Selection rules (GOMS) model, described by Card *et al.* (1983). The GOMS is one of the most widely known theoretical concepts in HCI (John & Kieras 1996). Especially its variant CPM-GOMS (where CPM can stand for both the Cognitive-Perceptual-Motor to describe its level of analysis, or Critical-Path Method to describe how it is used to predict a total task time) has been created to describe the parallel activities taking place when a user performs a task. The parallelism of activities is highly relevant to the ultra-mobile usage contexts where many of the usability problems root from the user having to conduct several tasks simultaneously.

GOMS is based on the human-processor model (MHP, Model Human Processor) described by Card *et al.* (1983). It models how human information processing works based on a set of perceptual and motor processors and a central cognitive processor. A newer architecture called EPIC (Executive Process – Interactive Control) has been proposed (Kieras & Meyer 1997) and its usefulness in modelling multimodal interaction has been evaluated (Kieras *et al.* 1997). Multimodality is an aspect that is very relevant for ultra-mobile usage contexts, as certain modalities such as vision is likely to become easily overloaded. Kieras *et al.* created a number of different models and compared their ability to predict the performance of a human telephone operator. At best, the prediction power of the models was very good. The smallest average absolute error of prediction was only 6.6% in the study.

The work of Baber and Mellor (2001) also concentrates on modelling multimodal interaction. Baber and Mellor use critical path analysis to determine the overall user performance when using different types of multimodal user interfaces. They point out how bottlenecks in the interaction can have many different roots. Certain steps in a task may have to be performed in a serial manner. Certain tasks may require the usage of the same modality and therefore cannot be performed simultaneously. There can also be technical limitations to performing the tasks.

Brumby *et al.* (2007a) have utilised cognitive modelling especially in the field of mobile user interfaces. They have described a cognitive constraint model to predict user performance in performing a secondary task while driving. Variations in the user behaviour can be modelled and predictions for the upper and lower boundaries for the duration of the dual task can be calculated. The cognitive constraint model has been used for modelling a task of dialling a phone number while driving (Brumby *et al.* 2007a) and a task of using the Apple iPod rotating wheel UI for making selections in lists while driving (Brumby *et al.* 2007b).

Salvucci (2001) presents a study where cognitive architecture models were prepared for a driving task and different dialling tasks. These were used to predict the user's performance in the driving task and the predictions were compared to the results from actual user testing. In other words, the user's performance in the driving task reflected the difficulty of each dialling task.

Cognitive models have been shown to be able to predict the performance of a user with a well-known and well-described task. They can be exploited as surrogate users, enabling quick iterations of changing the user interface and testing the effect of the change (Ritter *et al.* 2000). However, the problem with cognitive models is that they root from task analysis. There must be a well-defined task conducted using a well-defined user interface in order to conduct the analysis. Hence, their use is restricted to analysing and comparing existing user interfaces or user interface concepts.

The same problem also exists when applying workload analysis or usability testing. There first needs to be something to test. What is more, the analysis or test methods should be able to point out not only problems in the interaction but their causes. In order to point out solutions to the problems, the methods should also be able to point out less strained resources of the user.

Currently, mobile-specific usability methods provide measures for comparing different interaction methods in a mobile context of use. As such they do not provide tools for creating new concepts otherwise than pointing out problems with current ones. A cognitive model may reveal that at some stage in the interaction one of the user's senses becomes overloaded. However, it will not directly offer choices for solving the problem. A more generic method for describing the user's overall situation can help the designer to see where there is room for potential solutions and which ideas are more likely to just lead into further problems. In other words, the user's context must be analysed and presented in enough detail to help new ideas to emerge.

3 Analysing ultra-mobile usage contexts

With good user-interface design that would take into account the challenges related to the ultra-mobile usage contexts, many of the interaction problems could perhaps be solved. However, if most of the usability testing of mobile user interfaces is done in the laboratory, all of the true requirements of the ultra-mobile usage context do not get revealed. As a result, they do not affect the user-interface design. Besides, as pointed out by Cockton (2008), usability testing is often done at a late phase of product design and most of the major selections for the used technology and for the product concept have already been made. Therefore it would not be enough even if usability testing were mostly done in the field. It is much more important to include the requirements related to the mobile usage context in the definition of the product concept as early as possible.

The ISO13407 standard (ISO/IEC 1999) proposes that the human-centred design process should start at the earliest stage of the project. The process includes four human-centred design activities. These activities are to understand and specify the context of use, to specify the user and organizational requirements, to produce design solutions, and to evaluate designs against requirements. The first two activities are about specifying the requirements for the design solution to be created. The standard states that relevant characteristics of the physical and social environment should be described. The workplace and its furniture are given as examples of the physical environment. Temperature and humidity are listed as examples as the ambient environment. Work practices, organizational structure and attitudes are listed as examples of the social and cultural environment. Tools for analysing these are discussed next.

3.1 Gathering user requirements

Methods for analysing the users and their needs in general, are abundant. Often they are qualitative in nature, collecting information that is interpreted by the researchers. A common way of collecting user requirements is to go and observe users in their natural environment. Patton (2002) lists examples of this kind of field-based observation methods such as *participant observation, fieldwork, qualitative observation, direct observation* and *field research*. Millen (2000) presents a method called Rapid Ethnography, concentrating on the issue of conducting the ethnographic study in a timely yet effective manner. Beyer and Holzblatt (1997) present a formal ethnographic method called the *Contextual*

Inquiry. Ethnographic methods are so widely used in finding out user requirements that in fact, Hughes *et al.* (1992) claim that there is no one method of ethnographic analysis.

Another qualitative method for collecting user data is to collect a representative set of potential users and conduct interviews with them. Patton (2002) lists interview-based methods such as the *informal conversational interview, the general interview guide approach* and *the standardized open-ended interview*. Interviews can be conducted on a one-to-one basis, especially when they are used as a part of field studies. Perhaps more commonly, interviews are arranged as group interviews. Structured group interviews are often called *focus group interviews* (Morgan 1997).

The main goal in using these tools is to clarify the needs of the intended user of a product, in order to clarify the problems to be solved by the product designers. Typically the resulting product is an application or a device that enables the users to carry out certain tasks – for instance, an application for following the results of the user's favorite sport (Holzblatt 2005) or a mobile device specially aimed at teenage users (Berg *et al.* 2003).

However, as the research methods stem from the background of workspace analysis, the tools are inadequate for analysing the ever-changing ultra-mobile usage contexts. More specifically, the analysis of the usage environment has been defined in a superficial manner. There is no systematic way to illustrate precisely how the environment affects, e.g., the availability of the user's senses or cognitive resources. The handbook of qualitative design written by Patton (2002) mentions that the physical environment in which a field study takes place, should be described. But the book does not go on to describe exactly why and how the description should be done. It is left to the researcher to notice if and how the environment affects, e.g., device use – or how it would affect it, if a person were to use a device.

As a result, the design solutions typically are inherited from the world of computers. For example, the mobile sports application design, described by Holzblatt (2005) utilising the Contextual Inquiry method, does not include at all a design phase, where other modalities than vision and manual use would have been considered. As a result, the paper states: “How could we display all that information on such a small screen and make it readily available and readable at a glance” – with no consideration that all the information would not necessarily have to be presented on the small screen, but perhaps as sound or a tactile pattern.

3.2 Research of ultra-mobile usage contexts

Currently, context-aware computing is an active field of research studying mobile usage contexts. Several attempts to provide frameworks for analysing and modeling the mobile context have been proposed. Within the field of HCI, analysing the role of the ever-changing ultra-mobile context of use is still just taking shape. This is probably due to the history of usability research being in HCI which traditionally took place in a controlled working environment. Lately the fields of context-aware computing and HCI have been approaching each other. On one hand, more applications are being available through mobile devices, and on the other hand, researchers of context-aware computing becoming more aware of that measuring and modeling the context is rather irrelevant unless one can show how the user can benefit from it.

Mobile usage contexts are often so demanding that solutions are not obvious when problems are spotted. For instance, it is not straightforward to replace manual input with spoken input, even if the accuracy and the vocabulary of the speech recognition system were adequate. The mobile device is typically in the user's pocket or purse, so some manual interaction would be required to access its microphone. Social restrictions may prevent the usage of a speech recognition system in mobile contexts. Environmental noise can disturb speech recognition. In an office context, the microphone can be more easily arranged to be constantly near the user. The level and quality of noise is often rather constant, so the speech recognition system can be tuned to compensate for it or noise level can be reduced with room structures. In offices, adequate sound insulation can be arranged, and the social situation in the workplace is more tolerant to work-related interaction. In other words, in an office context, problems related to manual input may more easily point towards utilising speech recognition. In a mobile usage context, problems related to speech recognition may point towards considering also other methods, such as modifying the manual input to require less accuracy by utilising gesture recognition. Another reason for a different selection of modalities is the typically available user interface technologies. Paternò (2005) points out that the differences in the available screen sizes have an effect on selecting modalities for presenting information on an office PC or on a mobile device. Currently the Tablet PC form factor combines the large screen with at least some level of mobility. However, it is still questionable how much one would use a Tablet PC while walking. A good method for analysing the ultra-mobile usage context would reveal the possible problems in digging up and using the device in an ultra-mobile usage context. Those findings would then enable the designer to evaluate whether usage of such a device is realistic or whether, e.g., a wearable accessory providing control for the PC would be more useful. Such an analysis method would fulfil at least the following requirements:

- It would have to be simple enough to be used in practical product design. The results should be easy to interpret and their implications on product design should be clear. In other words, the effects that the environment has on the interaction should be well illustrated by the method.
- It should be as complete as possible. It should take into account all the user resources that can be practically considered to be utilised in the user interface.
- It should also be complete in the sense of pointing out all the relevant environmental factors related to user interaction. For instance, there can be several factors disturbing the usage of visual displays. One aspect is that the user's vision may have to be occupied elsewhere. Another aspect may be that the user is unable to hold the device so that he can see it. Yet another aspect may be that the environment is so bright that it is difficult to present information adequately using current display technologies.
- In order to fulfil the previous requirement, the method should be able to point out what kind of relevance environmental factors have on the user interaction. This would be a benefit also for context-aware computing systems, since the environmental factors are issues that can be sensed by the system, and by knowing their effects on the user interaction, the sensed data could be utilised to modify the interaction.

The requirement for practicality roots from the fact that product design is often hectic and there is little time for interpreting results and turning them into product requirements. The requirement for completeness aims at pointing out opportunities related to all of the user's resources that in practice can be considered to be utilised in the user interface. It is the available UI technologies and interaction techniques that define what is practically useful. The completeness also means that no issues causing restrictions to a resource are missed in the analysis. It is obvious that when considering the utilisation of a certain resource, all things affecting it need to be taken into account.

The relevance of the environmental aspects gets defined as a side-effect when considering all things affecting each resource. When analysing how restricted the usage of a certain resource is, one needs to define where the restrictions come from and how. This information can be used for enriching the description of the environment, making it more clear how the environment affects use and how the restrictions could perhaps be overcome. If an analysis method does not make a point of describing how an aspect of the environment affects interaction, there is a risk that analysis is done in a superficial manner. For instance, if one lists only “bright sunlight” as an aspect affecting the sense of vision, it is left for the designer to estimate how that affects device use. Often it is the designer's own, previous experience that then defines how the illumination is taken into account in the device design.

Current proposals for analysing the context of use either only touch the relations between the environment and the user, or when going more deeply into the analysis, still fail to describe those relationships adequately. In context-aware computing, the aim typically is to recognise a context based on sensor readings and the internal state of a device. A good overview of existing research on describing contexts is presented in the PhD thesis of Korpipää (2005). It can be seen that the focus of the research is recognising the user's context, but the next step of describing how the device should react to the context is typically left open.

In the field of user interface research, few studies exist that would concentrate on describing how the environment affects the human-device interaction. None of the existing studies takes a systematic approach in describing the effects of the environment. Context-aware computing provides tools for recognising contexts and different characteristics of them. User studies should reveal what the effects of these characteristics are and how user interface solutions could overcome them.

3.2.1 Existing user research methods and ultra-mobile context of use

Let us now analyse some of the current user research tools with respect to how they propose the usage environment to be studied, and how well that fits to the ultra-mobile usage contexts and the situationally-induced impairments in them. Table 2 presents a list of existing research on studying usage contexts and the variables proposed to be analysed in the presented methods.

The Contextual Design method (Beyer & Holzblatt 1997) - part of which is the Contextual Inquiry - aims at conducting field studies and building models of the intended users' usage contexts. The work models are the following:

- The flow model
- The sequence model

- The artifact model
- The cultural model
- The physical model.

Of these, the physical model aims at describing the physical environment in which the intended product is to be used; a description of the space where the work takes place. Beyer and Holzblatt state: “The physical environment is the world people live in: the rooms, cars, buildings, and highways they move about and work in; how each of these spaces is laid out so that it supports work; and how they use these spaces in the process of working.”

The key phrase in the description is “is laid out” – in other words, the model assumes that the environment has been constructed to aid the work. While this principle makes great sense when building office applications, it is far less useful when designing mobile devices. Mobile devices are often used in situations where the user has had no chance of affecting the environment – for instance, when walking on a street.

Furthermore, the physical environment in an office building has typically been arranged so that there are no great limitations to using the human senses. It is therefore no surprise that the physical model in the Contextual Design method makes no explicit point of describing how the environment affects the user's senses and abilities to operate devices.

The Contextual Design method has been designed as a tool for aiding everyday product design. It is very practical and aims at generating clear requirements and ideas for product design. However, it does not aim at utilising all possible resources of the user. It is incomplete both in analysing all practically possible user resources or things affecting them. Hence, it is far less useful for analysing the challenges of the ever-changing mobile environment. Problems caused by the environment can be pointed out in the different work models of the method, but there is no model explicitly describing the environmental effects, their causes or effects.

Hackos and Redish (1998) present a method for user and task analysis. Its aim is similar to that of the Contextual Design in that it starts from doing field research among the intended user, and utilising data gathered from there in designing systems and applications. Hackos and Redish describe the phase in which the user's physical environment is to be studied in more detail than Beyer and Holzblatt.

These factors (and others that have been observed) are to be used when creating Environment profiles that describe the place where the intended product is to be used. Hackos and Redish mention that the list is not exhaustive. However, expanding the list would most likely happen by studying how the environment currently disturbs the tasks the user is performing, using the current devices. The method therefore would be unlikely to reveal possible problems with future user interfaces. This would, however, be important as when designing a solution for a current problem, one can easily create an equally problematic solution. For example, a current problem of difficult initiation of phone calls when cycling (due to the user's hands and vision being mostly needed for the cycling task) could be solved by a speech recognition system. In practice, though, the wind noise picked by the device microphone could easily render this solution useless. Therefore it would be important to make a note that there is wind noise in the environment, even though it does not necessarily pose a great problem for current usage.

Like the Contextual Inquiry, also the method of Hackos and Redish is suitable for practical product design. In a similar manner, it also lacks a systematic way of analysing how the environment affects the user's resources. It does not provide a method for going through all the available resources in a context to see which ones of them would best suit the situation. Hence, it is mostly useful for analysing contexts with relatively few or obvious restrictions.

A more complex framework for creating mobile scenarios has been presented by de Sá and Carriço (2008). Their aim is to analyse mobile usage contexts, and utilise the data in creating scenario descriptions that aid in designing mobile devices and applications. The framework consists of three parts: Contextual scenarios, Scenario transitions and Scenario variables. The Scenario variables are used as building blocks when building definitions of the Contextual scenarios. The framework described by de Sá and Carriço has a practical approach, as it defines easily understandable variables such as whether the user is sitting or standing. However, it is incomplete in many respects. For instance, the framework lists “movement and posture” as one scenario variable. As another variable, affecting the user's workload, the framework mentions “activities”. However, if the activity of the user is, e.g., bicycling, and the posture is related to the bicycling task, the framework provides little means for describing exactly how those affect the usage of a mobile device. The framework does not describe a comprehensive way to analyse how all of the factors of the environment affect each resource of the user. It is left for the researcher to spot or miss these effects.

Kim *et al.* (2002) present a framework for analysing mobile usage contexts and how they affect the usage of mobile Internet. The context variables presented by Kim *et al.* are rather limited and it is questionable whether they are enough to describe a usage context. For instance, the cognitive load implied on the user by the different tasks in the context has not been listed among the context variables at all. Furthermore, the study by Kim *et al.* relied on users reporting their context while using current mobile Internet services with the currently available user interfaces. So, possibilities for future user interfaces were not logged at all – for instance, cases where a system is currently *not* used due to some restrictions in the environment. Nevertheless, Kim *et al.* do present some realistic parameters of a context – e.g. by studying whether the user is walking or stopped when using a mobile device. The method of Kim *et al.* has a practical starting point – analysing a concrete task of using the Internet with a mobile phone. Its underlying principles certainly could be used for analysing also other kind of mobile device usage. However, it too is far from being complete in describing how the environment affects device use. For instance, the method mentions "visual and auditory disturbances in the environment" but leaves it to the researcher to define what the effects are. As a result, the method does not generate very practical requirements for the user interface.

In the field of context aware computing, Sears *et al.* (2003) have pointed out the lack of analysis methods for mobile usage contexts. In a previous article Sears & Young (2003) have defined the term "situationally induced impairments and disabilities" (SIID), pointing out that depending on the usage situation, any user can have similar disabilities to use a mobile device as those with permanent physical conditions such as a vision or hearing impairment. This definition certainly is useful in any analysis of mobile device use in an ultra-mobile usage context. The philosophy is, in fact, the basis for any analysis technique that concentrates on studying how the environment restricts the usage of certain user resources. The context model defined by Sears *et al.* (2003) mentions that the available I/O channels should be taken into account when analysing an environment. In practice that means the available UI technologies, so from this respect the framework is very practical. However, the framework apparently does not even aim to provide a complete method for analysing mobile environments but rather points out what all can have effects, and presents some examples.

Obrenović *et al.* (2007) take a similar approach, pointing out that there are several layers of constraints between the device and the user. A message to be conveyed is constrained by the characteristics of the device, the environment, the social context and the user. Some of the user's characteristics may be static; for instance, one may be blind. The device constraints can be affected by designing new devices. The rest are dynamic and depend on the context of use. Obrenović *et al.* provide examples of what the constraints may be and where they originate from, but these are provided in an ad-hoc manner. No formal way of collecting the constraints and analysing their effects is provided. However, the model does aim at pointing out the restrictions of the environment (or the user's characteristics), so from this respect its underlying philosophy is very pragmatic. What is still missing is the systematic approach of listing which resources of the user could be utilised in the user interface, and in which possible manners the environment affects them. The study of Obrenović *et al.* only provides examples of whence the restrictions generally could stem.

Lemmelä *et al.* (2008) present a method for designing mobile user interfaces by observing users in mobile environments and making records of how the environment affects human-device interaction. One viewpoint in their method is to list restrictions according to what they affect. Lemmelä *et al.* classify the findings into groups of aural, visual, physical, cognitive and social restrictions. This information is then used for building storyboards describing the contexts. The findings are used as food for thought when creating the storyboards. This is a very practical approach and the findings should be useful in product creation. However, it is still easy to miss opportunities with the method. It does not include an explicit step where each resource of the user would be analysed according to how the environment affects it. The effects are also listed on an implicit level only. The method is not complete in the sense of analysing all available resources or on the other hand in the sense of analysing everything in the environment that has effects on certain resource.

Sears *et al.* (2003) present an analysis of proposed descriptions of the notion of 'context' in context-aware computing, and point out that many of them have good high-level descriptions, but the relation of the characteristics of a context and their relevance to the user is often left vague. Sears *et al.* present a combination of how the notion of 'context' has been defined by Dey *et al.*(2001) and Schmidt *et al.* (1999a), adapting their selection of context variables from those publications.

In another paper Schmidt *et al.* (1999b) have provided a model for context that is a bit more thorough than in the one (Schmidt *et al.* 1999a) used by Sears *et al.* It, as well, is intended to be used in context aware computing, but like the paper by Sears & Young (2003), it includes considerations that are relevant for any analysis of the mobile usage context.

Table 2. Previously proposed environment analysis variables (I, reprinted by permission of Taylor & Francis).

Originator	Human	Environment	Application/ activity
Hackos& Redish (1998)	- User (background knowledge, mental models, vocabulary) - Individual differences (personal characteristics, physical differences, cultural differences, motivational differences)	- The working environment (home, office, laboratory) - The working space - Noise - Dirt - Lighting - Temperature - Speed - Power sources - Proximity to information sources - Danger	-Task flow (workflow analysis, job analysis) - Frequency - Criticality - Time to complete - Difficulty - Division of responsibility
de Sá & Carriço (2008)	- Sitting - Standing - Walking - Movement impaired *(sic!)* - Visually impaired - Heterogeneity	- Lighting - Noise - Weather - Obstacles - Social environment - Cognitive distractions - Physical distractions	- Criticality - Activities - Single-handed - Dual-handed - Stylus/ Finger/ Keyboard/ Numeric pad
Kim *et al.* (2002)	- Internal context; goal and emotion during system usage	- Physical context; visual and auditory disturbances in the environment - Social context; number of people around the user and the amount the user interacts with them	- External context; usage of hands and legs while using a system
Sears *et al.* (2003)	- User; identity, preference, biophysiological conditions, emotional state	- Social environment; existence of other people, social interactions - Location - Physical conditions; e.g. lighting, noise, temperature, speed - Infrastructure; embedded	- Activities; walking, filling out a form, looking for a phone number, following a map - Functions; e.g. backlight control, font size adjustment, zoom - I/O channels; e.g. touch

Originator	Human	Environment	Application/ activity
		devices, sensors, communication protocols	screen, voice command, gesture recognition, audio output, vibration
Schmidt *et al.* (1999b)	- User	- Social Environment - Light - Pressure - Acceleration - Audio - Temperature - Infrastructure - Location	- Task
Obrenović *et al.* (2007)	- Disability (an extensive list, e.g. blindness, poor acuity, clouded vision) - User's current state (emotional context)	(Lists issues related to a driving context as an example) - Traffic situation - Noise level - Visual conditions - Weather condition - Social context	- Driving (listed as an example)
Lemmelä *et al.* (2008)	- User's long and short-term abilities	- Aural restrictions (e.g. traffic, conversation, footsteps) - Visual restrictions (e.g. amount of natural light, artificial lighting) - Physical restrictions (e.g. clothing) - Cognitive restrictions (social situation, amount and type of traffic, walking surface) - Social restrictions	- Aural restrictions (e.g. conversation) - Visual restrictions (e.g. following the environment) - Physical restrictions (e.g. person's plan and schedule, pace of walk)
Pascoe *et al.* (2000)	- User (with possible impairments)	- Visual / Auditory / Tactile restrictions based on environment	- Visual / auditory / tactile restrictions based on current task - Available I/O techniques (display, microphone, loudspeaker, buttons)

All of the methods presented above aim at practical findings that can be used to guide product development. However, none of them aim at completeness in addressing either all the potentially available user's resources or all factors that affect them. When this approach is lacking, there is a risk that the solutions which designers create are not fully utilising the available resources of the user, as there is no systematic way to analyse each potential resource at a time. Furthermore, when creating new solutions, there is a risk that the solutions themselves create problems in the target environment. Speech recognition may enable eyes-free interaction, but for example wind noise because the user is moving while bicycling may prevent its use in practice. As another example, a designer might compare the benefits of auditory and tactile output for eyes-free presentation of information. If the environment analysis has not been conducted thoroughly, it is easy to miss things such as the environment causing vibrations that mask the tactile output of the device.

One important factor lacking in all of the presented methods is that of modelling how the environment affects interaction. Instead, they present examples, and in the case of Lemmelä *et al.* (2008) the findings are used for creating storyboards. For designing commercial products, it is important to evaluate how widely useful an interaction concept would be. In order to do that, similarities between different environments must be analysed. It is noteworthy that the similarities should exist in the restrictions the environments cause for the interaction. It is not necessary for the environments to be physically similar; a restriction may in one environment originate from the physical characteristics of the environment, and in another from a typical task the user is performing. Therefore it is important to model the restrictions that the different environments cause, not necessarily the physical characteristics of the environments. The importance of recording the physical – and other – characteristics is in keeping track of what causes the restrictions.

3.2.2 Using models in analysing ultra-mobile usage contexts

If cognitive models can be used to analyse the usability of user interfaces, a question rises whether they can also be used for analysing the effects of the environment.

A benefit in cognitive models over other environment analysis methods is the capability to model cognitive constraints. At some point, performing several tasks such as walking and operating a mobile device becomes simply too difficult, regardless of which UI technologies are used. To a great extent, the restrictions in current usage of mobile devices are related to the environment causing restrictions to the same resources that the mobile user interfaces utilise. These can be pointed out by using the environment analysis methods described in the previous chapter. However, showing when the cognitive effort becomes too great requires a way to model the human information-processing system. Sigman and Dehaene (2005) present a simple but effective model for characterising the “mind’s bottleneck” and analysing which components of a mental operation can be run in parallel and which must be run serially.

This is the field where cognitive models become important. With those, it can be shown how taxing a certain interaction concept is likely to be, and problematic steps in the interaction can be pointed out. Baber and Mellor (2001) use critical path analysis to show dependencies between different resources of the user. They present a number of simulations of concurrent tasks utilising various resources and show that the user's performance predicted by those simulations is quite accurate when compared to real performance of users.

However, as a generic analysis tool, cognitive models are too constrained. They are based on well-defined tasks and user interfaces that can be analysed down to the last detail. When concepting new interfaces for mobile usage contexts, the user interfaces have not yet been created.

Sutcliffe and Faraday (1994) present a model-based approach for designing multimodal multimedia user interfaces. They start by building a task model representing a task of handling emergency situations on a shipboard. They then proceed into selecting presentation modalities based on the type and time of information to be presented. However, selecting modalities based on which user resources are available, is missing from the study. This surely results from the fact that the shipboard environment is relatively static and the available resources remain constant for the most of the time. Choosing the best presentation modality based on the information type is definitely important when designing multimodal user interfaces. However, if the user's context prevents or disturbs the usage of the needed modality, it should be considered whether the presentation would be changed. The amount of presented information can perhaps be changed, or perhaps the second best modality for the presentation should be chosen. The model-based design method presented by Sutcliffe and Faraday supports finding the best modalities for presenting information and receiving user input. It starts from abstract models representing the user's task and on the other hand the information types and media offered by the computer. What is still lacking is a way to include the aspects of the user's environment that affect interaction.

3.3 Analysis framework for usage environment

Publication I describes a novel framework for analysing and modelling the usage environment, taking the user's senses and capabilities as a starting point. User's senses are typically related to the output modalities provided by the device, and the capabilities to the input modalities. A capability can, for example, be the ability to manually operate the device. Each capability can be affected by many different aspects of the environment. Therefore the framework separates features of the environment (naming them “environment variables”) and capabilities of the user (naming them “resource variables”). There can be a many to many mapping between the two variable sets. One resource can be affected by many aspects of the environment, and one aspect of the environment can affect several resources.

The framework in Publication I defines a space called the “resource space” with the user's senses and capabilities (jointly called as “resources” in the publication) as axes. The capabilities and senses are analysed not only from the point of view of the environment itself, but also through typical activities in the environment. For instance, one could define an axis of how well the user can use his vision, and place the zero value to, e.g., a situation where it is completely dark, or when the user's vision is completely required for some other task than mobile device use.

The resource space affects both the input and output modalities that could be used for operating a device in the studied context. The number of possible output modalities for the device is rather fixed, as it is related to the five senses and the cognitive capabilities of the user. Currently, mobile user interfaces utilise the senses of sight, hearing and touch, but even the senses of smell (Brewster *et al.* 2006) and taste (Maynes-Aminzade 2005) have been suggested as output modalities. The number of resources related to input modalities of the mobile device is even more limited today. In Publication I, only manual input and speech recognition have been considered. However, this simply reveals a lack of currently existing input technologies. While manual input is a rather large area, covering things like gesture recognition in addition to pressing buttons, hands are still not the only body part that could be used for device operation. For instance, the user could use head movements or blink his eyes to control a device. Since these technologies were not practically available when the study presented in

Publication I was conducted, they were omitted from the studied capabilities of the user. The framework itself does not restrict the number of studied modalities in any way.

Another space that is mentioned in Publication I is called the “environment space”. It describes the physical (and more abstract such as social or cognitive) aspects of an analysed environment. It is simply built from the different measured (or observed) pieces of data in the different environments. A measurement could be, e.g., the average background noise level, or the average number of people around the user, or the amount of interaction the user has with the other people around. In many cases, these values would not be direct measurements but classifications. For instance, one variable that has been proposed in Publication I for describing a certain aspect of the user's environment is that of how demanding the situation is socially. If the user is alone, the situation is not socially demanding at all while in a meeting where, e.g., disturbing others would be not be looked kindly upon, the situation can be considered to be socially very demanding. In practice, it is somewhat subjective how socially demanding a person considers a situation to be. This aspect of the environment is therefore best described by just classifying the situation roughly, e.g., on a three-step scale varying from “not demanding at all” to “very demanding”. In a similar manner, the cognitive demands were collected mostly by self-evaluation and simple reaction time tests. As has been discussed above, in contexts where there is a clear primary task (such as cycling or driving), a cognitive model could have been used to model the cognitive demands of the task. However, in the study described by Publication I, this depth of analysis was not taken. Nevertheless, for the sake of completeness both social and cognitive demands of the environment should be taken into account in the environment analysis, even though in the future this analysis can be done in more detail than in Publication I.

The key to using the framework is in the mapping between the environment space and the resource space. This mapping tells us how different resources are affected by different aspects of the environment. The measured or observed values of the environment variables then tell how loaded a certain resource is in the studied context. Selecting the modalities to be used in that context then depends on which resources are less burdened. The type of information to be transferred between the user and the device also affects that selection (Sutcliffe & Faraday 1994, Lemmelä 2008) - but that is beyond the scope of Publication I.

In context-aware computing, efforts have been made to model the context of use, but so far the relevance of different characteristics for the HCI has been mostly shown implicitly. It has been pointed out by Barnard *et al.* (2007) that currently it is becoming possible to measure many kinds of potentially relevant data, but so far the research community has been unable to define how to use it in a manner relevant for the user. Barnard *et al.* state: “what is missing is an understanding of how changes in context affect the user.”

In the framework described in Publication I, relevance of the characteristics of the environment is explicit. The connections between a characteristic of the environment and the available capabilities of the user define how that characteristic is relevant for user interaction. In other words, the mapping from the environment space to the resource space defines this relevance. For example, a characteristic of the environment can be that there is background noise, mostly caused by people speaking. This characteristic can have several links to the capabilities of the user and the mobile device. The noise can prevent the user from hearing sounds that the device may play. It can also affect the user's ability to concentrate on device usage. Moreover, it can disturb the speech recognition capability of the mobile device. If the definitions of these links are valid then the findings are always relevant. They started from real capabilities of the user, affecting device use. This approach has two benefits. First, the analysis can be reasonably complete. There is a limited set of capabilities that are taken as the starting point for the analysis. The number of analysed characteristics of the environment can be limited to those affecting the selected capabilities. Second, the findings can be easily interpreted as user interface requirements, leading to design ideas. If the system were to concentrate on recognising, e.g., that the user is walking, then the question of how the walking affects his capability to operate a mobile device would still remain open.

Another novel approach in Publication I is that by clustering different environments according to how they affect the user, it may be possible to find environments that are not similar at first sight, but pose similar restrictions on the user. Flanagan *et al.* (2002) have proposed unsupervised clustering of environment data. However, their concept is based on clustering similar sensor data together – not cases where similar restrictions would be placed upon the user's resources. These two can be the same, but not necessarily. For instance, restrictions on the usage of sight can be caused by lighting conditions (which can be detected easily with sensor technology) or typical tasks in the context (which is difficult to detect by using sensors). Publication I presents a novel approach of clustering environments based on their effects on the user's resources, not based on the characteristics of the environments themselves. This information can then be used for creating products that would serve the user in several environments.

Finally, Publication I presents an example of analysis conducted by using the framework. Five different usage contexts were analysed and an example of clustering three of them is presented. Publication I provides an example mapping – that can still be expanded – from the points in the environment space, related to the analysed environment, into the resource space related to the user. Publication I also presents three user interaction concepts that have their roots in the analysis described in Publication I. Two of these are presented in more detail in Publications II and IV of this thesis.

So far no previous research has taken the approach of systematically listing the resources of the user where corresponding UI technologies are available, and analysing one by one what in the environment affects those resources and how. The work by Lemmelä *et al.* (2008) has a similar underlying philosophy, but the method does not systematically analyse all the relevant capabilities and senses. What is more, the idea of clustering environments based on similar effects on the user's resources is novel in Publication I.

As the framework presented in Publication I starts from the viewpoint of the user's senses and capabilities, it provides means for a much more complete analysis of a context than previous efforts. It is, after all, those senses and capabilities that define what the user can do in a certain context. Previous approaches of starting by trying to describe the user's context as accurately as possible have run into the problem described by Sears *et al.* (2003). The relation between a characteristic of a context, and what kind of a situational impairment it leads to, has been implicit. In some cases it can be obvious – for instance in very noisy environments the user can't hear the device sounds unless those are played through headphones. But even in those cases it has been easy to miss, e.g., the fact that the noise also has other effects such as disturbing the user's concentration. Or, the background sounds as such might not make device sounds inaudible, but the user needs to concentrate on the sounds, which makes the auditory channel overloaded. Starting separately by listing on one hand what the user is able to hear and on the other hand what he is able to concentrate on reveals these problems.

Let us now compare the framework of Publication I against the requirements listed in Section 3.2. The framework has originated from practical purposes of product design, so it aims at creating easily understandable and practical descriptions of how the environment affects mobile device use. The mapping between the environment space and the resource space provides a clear way of describing the effects of the environment. The framework starts by thinking of the foreseeable UI techniques, and listing the resources needed for utilising those. In Publication I it is presumed that no UI technologies utilising the senses of taste and smell will be available in the near future, so those are also left out of the design. From this point of view, the framework is complete – all the resources that are presumed to be relevant are analysed.

The completeness of the analysis in the sense of going through everything that affects a resource, is always debatable. The underlying philosophy of the framework is aimed at completeness, but in practice it was limited how far this analysis could be taken. Especially in the case of the cognitive resources and the social situation the study described in Publication I takes a rather shallow approach. For certain environments, things like dual task performance were measured. In practice, cognitive models could have provided more detailed analyses of where interaction becomes too difficult, regardless of the available UI technologies. Especially when evaluating the newly created interaction concepts, cognitive models can provide valuable tools for evaluating their usefulness based on their cognitive demands.

4 User interface design for ultra-mobile usage contexts

In previous chapters, pointing out and analysing situationally-induced impairments has been discussed. This chapter concentrates on ways for building out solutions for them. Special emphasis is put to designing solutions that would be feasible and acceptable on consumer devices. User interface designs presented in Publications II – VII will be discussed, and evaluated against the analysis framework described in Publication I.

Multimodal user interfaces have often been proposed as a solution for enabling system use in varying mobile usage contexts (Brewster *et al.* 2003, Kernchen *et al.* 2005, Lumsden & Brewster 2003, Oviatt 1999, Roessler *et al.* 2001, Sears *et al.* 2003, Serrano *et al.* 2006, W3C 2010). However, there are several kinds of multimodality, some of which are better suited for ultra-mobile usage contexts. Multimodality has been argued to lead into more natural user interfaces (Turk & Robertson 2000, Martin *et al.* 2001). While this may be true, it is a long-term goal. On short term, there are pressing issues in that the situationally-induced impairments often prevent usage of a mobile device at all in many usage contexts.

Therefore, the most beneficial types of multimodality for ultra-mobile use are those that allow using which ever modalities are available at a certain time. Systems that aim at increased user interfaces efficiency by allowing the user e.g. to point and speak simultaneously are probably better suited to office environments. In ultra-mobile situations neither pointing nor speaking is often suitable. Pointing typically requires the use of two hands (one to hold a device, another to point) and the typical background noise in mobile contexts can decrease the accuracy of the speech recognition system.

4.1 Multimodal UI in ultra-mobile usage contexts

In this thesis, the human-centred definition of multimodality (Baber & Mellor 2001) in user interfaces is used. The definition is based on the five sensory modalities through which people can receive information, and the at least two response modalities through which they can provide information. Multimodality can also be approached from the viewpoint of multisensory processing (Patton *et al.* 2002) where the emphasis is on studying the brain mechanisms behind simultaneous use of the human senses. However, that is beyond the scope of this thesis.

The human-centred definition (Baber & Mellor 2001) of multimodal HCI is based on the five sensory modalities through which people can receive information, and the at least two response modalities through which the user can present information. A modality can be seen as an information channel between the device and the user. The sensory modalities are related to the user's five senses. Response modalities are related to ways through which the device can sense user's input. In this thesis, input modalities are generalised into the manual and spoken inputs. These can be further divided. Means of manual input can include the usage of buttons, gestures or a touch screen. Means of spoken input can include speaking, but also whistling or making other non-speech sounds. Baber and Mellor also describe a technology-centred definition of multimodality, which is based on the different modes (e.g., a visual display or sound) through which a system can present information and the interaction devices (e.g., a keypad) through which it can receive user input. In practice there is a clear link between the two. A system can only interact through its different input and output modes. When designing a new device, there must be some knowledge of the technologies that are feasible to use. It is therefore fruitful to concentrate on those modalities of the user that can exploit the available technology. Typically the senses of smell and taste can be omitted from the design since technologies capable of producing those sensations are not available for mainstream product design.

Nigay and Coutaz (1993) define multimodality as follows: “Multimodality is the capacity of the system to communicate with a user along different types of communication channels and to extract and convey meaning automatically.” The definition emphasizes that the presentation of information in different channels contains semantic meaning in the interaction – as opposed, e.g., to watching a movie where the presentation clearly is multimodal, but the contents of the movie do not mean anything to the device presenting it.

However, multimodal interaction can be implemented in many ways. The modalities may be combined into a single communication act, as in the famous “put that there” problem. On the other hand, the system can allow inputs from several modalities to mean the same, leaving it to the user to decide which modality to use. In a similar manner, device output can be provided using multiple modalities either simultaneously (as when a visual note is accompanied by a sound) or separately (as when an incoming call is alerted either by a sound or as a tactile pattern, depending on the current mode of the device).

4.1.1 Types of multimodality

Nigay and Coutaz (1993) present a classification of different types of multimodal systems, concentrating on whether the input from different modalities happen simultaneously or in sequence, and whether they need to be combined in order to produce a single communication act (as in the “put that there” case) or not. The four classes they present, related to how events from different modalities are related to each other temporally and semantically, are as follows.

- Alternate (sequential use of modalities, combined into a single communication act)
- Synergistic (parallel use of modalities, combined into a single communication act)
- Exclusive (sequential use of modalities, events separate from each other)
- Concurrent (parallel use of modalities, events separate from each other).

In the paper, Nigay and Coutaz also present a software architecture for creating synergistic systems.

The classification presented by Nigay and Coutaz does not explicitly take into account the case where two events from separate modalities may both carry the same information. For example, a person might perform a hand gesture for “Stop” and say “stop” simultaneously. These could belong into the “synergistic” class, but on the other hand, in this example either modality could be left out of the communication act without changing its meaning, so not both are needed to comprise a communication act. The “concurrent” class could incorporate these cases as well. However, also cases where a user would, e.g., say “switch the light on” to a smart room while simultaneously adjusting the room temperature with a turn knob – i.e. the events would be completely unrelated – would belong to that class.

Martin (1997) presents a different classification that takes this issue into account. The classification, presenting various ways of different modalities cooperating, is as follows.

- Transfer (an event from one modality is used by another modality; e.g., a spoken search command leads to a presentation of a visual image)
- Equivalence (events from different modalities mean the same communication act; e.g., choosing a command in a drawing program either by clicking on an icon, or saying its name)

- Specialization (certain information is always presented through a certain modality, or a certain modality is used only to present certain type of information)
- Redundancy (events from different modalities occur simultaneously, and present the same information; e.g., clicking on a town's name on a map and simultaneously saying its name)
- Complementarity (events from different modalities need to be combined to form a single communication act; e.g., saying "put that there" while pointing to what is referred to as "that" and "there").

When considering an ultra-mobile usage context, it is clear that the user's resources are typically limited by the environment. The user may have to observe the behaviour of other people, vehicles, animals etc. in the environment. The background noise level may restrict the amount of sounds the user will hear from a mobile device carried in the pocket. The hands may be occupied, e.g., with carrying things, or the clothing required by the weather conditions restricts the natural movement of the body, limbs and most importantly, fingers.

If a multimodal user interface in this context requires that at least two of the user's available modalities are used simultaneously for operating it, the probability of these being available for use is reduced, compared to a unimodal system. This is contrary to what the user would need, and what has been claimed as a benefit with a multimodal user interface.

From the classification or Martin, the classes Equivalence and Redundancy would seem to have most potential for ultra-mobile use. In those, the user can perform a same action by using several modalities, or receive the same information through different modalities. They may also be somewhat easier in implementation, as they may not require as strict synchronization between the interpretations of inputs from different modalities.

4.1.2 Multimodal output

Research on multimodal interaction often concentrates on combining several input modalities. Multimodality has been seen as a way to enhance the naturalness of interaction (Turk & Robertson 2000). Multimodal output has often been seen as somewhat separate from multimodal input. Martin *et al.* (1998) point it out and list some potential benefits.

In practice, multimodal output is far more widely used in today's mobile devices than multimodal input. Practical design problems have led mobile device designers, e.g., to use vibration patterns for silent alerting of ringtones. In many cases a mobile device needs to catch the user's attention in order to inform of some event. The social situation may restrict the usage of some output modalities, so, e.g., a call alert needs to be performed with vibration instead of sound even though it is more difficult to present detailed information through the tactile channel. Sometimes it can also be the mobile context itself that requires using multimodal output. As Roto and Oulasvirta (2005) point out, non-visual output may need to be used just because the user's visual attention cannot be trusted to stay on the mobile device. Therefore, the field of multimodal output is at least equally important as that of the multimodal input. Especially in a divided-attention situation – as is often the case in ultra-mobile usage context – multimodal feedback from the user's actions has been proven beneficial (Lee & Spence 2008). Lee and Spence compared different feedback concepts related to a touch-screen-based mobile phone used in a simulated driving situation. The results showed that bi- and trimodal feedback showed significant benefits in both the driving task and the tasks performed with the mobile phone.

Solutions for the situationally-induced impairments of course need not be multimodal. For instance, a headset of a mobile phone is intended to solve call handling problems in eyes-free situations. Currently the output of a typical headset is unimodal, but it is not the visual but the auditory modality that is used. In practice, these cases still lead to a need for multimodal designs. The tasks the user can conduct using a headset using a manual/auditory UI, can obviously also be performed on the handset, using a manual/visual UI. Both must be designed so that the user understands what he is doing. This typically leads to questions of cross-modal presentations, i.e. how the same piece of information in the interaction flow should be presented in the visual and the auditory cases.

This notion has implicitly been presented when naming the auditory and tactile equivalents of visual icons. Blattner *et al.* (1989) presented the idea of *earcons*, i.e. short auditory patterns in which information could be encoded. Earcons are typically short musical themes that can be easily arranged into a hierarchy. For example, a melody played with a piano sound may represent the arrival of a new email message while the same melody played with a violin sound may represent a new multimedia message. The name “earcon” is derived from the word “icon” in a pun-like fashion. Gaver (1986) has presented the idea of *auditory icons* which are based on everyday listening. Gaver utilises the idea that when hearing a sound, humans typically think of its source rather than its acoustic parameters. This approach allows the generation of multimodal metaphors. For example, a data storage may be presented as a container that can look fuller or emptier. When interacting with the container, the system may produce a sound that in a similar manner resembles hitting a fuller or emptier container.

For the tactile modality, Brewster and Brown (2004) have presented a way to present information as tactile patterns. In their paper, they have named these patterns as *tactons*. In her PhD thesis, Brown (2007) presents the construction of tactons in a similar manner as earcons.

Hoggan and Brewster (2007) have studied the learnability of crossmodal icons by having users learn the meaning of tactons and equivalent earcons. In a similar fashion, Hoggan *et al.* (2008) have also studied the congruence between the visual style of touch-screen buttons and the tactile feedback from pressing them. Pirhonen (2007) has studied how the semantics of earcons and visual images are related.

Tuuri *et al.* (2007) present a hierarchical scheme of listening modes. The scheme starts from the very basic perceptional level – i.e. the reflexive mode of listening where a sound causes an affective reaction. For instance, a loud sound causes the listener to startle. The next, connotative mode is related to what kind of connotations the listener relates to the sound. The scheme lists listening modes in a rising order of abstraction level. The third, causal mode is related to what the listener suspects the cause of the sound to be. The scheme contains functional and semantic listening modes, where the listener has had some time to analyse the sound and its possible meaning. The highest, most abstract listening mode in the hierarchy is the reduced mode where the listener thinks of the properties of the sound itself.

In sound design, methods exist that utilise the listening modes proposed by Tuuri *et al.* (2007). The design paradigm described by Patterson (1982) **-** where sounds are designed to present a varying level of urgency – most probably relies on the reflexive and connotative listening modes. Earcon design – where the sound's properties such as the timbre or the rhythm are related to its meaning – work on the functional, semantic and even reduced modes. The auditory icons presented by Gaver rely mostly on the connotative and causal listening modes.

Tuuri *et al.* (2007) and later Mustonen (2008) point out that the earcon and auditory icon paradigms for sound design should not be seen as different types of UI sounds but as just design paradigms. One should not pick a paradigm first, and then continue to design sounds, but rather use different paradigms, mixing them whenever appropriate. However, in currently existing auditory UI concepts the different paradigms are hardly ever mixed. The roots of the auditory icons design paradigm are in the everyday listening, which allows finding familiar metaphors for sound design. Something that sounds empty can be used to represent emptiness. However, the structured nature of earcons is not easily applicable to auditory icon design. Furthermore, what is known of parameters related to the perceived urgency of sounds is rarely utilised in either earcon or auditory icon design.

Publication VI of this thesis addresses this issue. The aim in the publication was to find out sound parameters that would be associated with lower or higher importance of the event causing the sound. The parameters were selected so that they could be applied onto any sound. The sounds could first be created, e.g., by using the auditory icon or earcon paradigms, and the sounds could then be treated with the parameters related to the importance of the event.

It is interesting that a scheme similar to the listening modes presented by Tuuri *et al.* (2007) has not been proposed for the tactile modality. Currently, tacton design seems to be still concentrated on building tactons utilising multiple dimensions (e.g. rhythm and roughness), maintaining good recognition rates. In her PhD thesis, Brown (2007) mentions the semiotics of touch as one future research direction. Benali-Khoudja *et al.* (2005) present an attempt to create a tactile language where different tactile patterns are used for presenting different emotions. However, they do not analyse the factors why certain patterns and certain emotions would be linked.

4.1.3 Existing concepts for mobile multimodality

It has been shown that under increased cognitive load people tend to start utilising multiple modalities in a user interface, if they are given a free choice (Oviatt *et al.* 2004). For the ultra-mobile usage context this is an important fact since the context often increases the cognitive load, compared to a situation where the user is sitting down, fully concentrating on device use. It has also been shown that multimodal input can improve the performance of recognition-based systems – such as speech and gesture recognition – by allowing the system to mutually disambiguate the possible results of different recognition engines (Oviatt 2000).

As has been mentioned, the most potential use for multimodality in ultra-mobile usage contexts would be to build enablers for interaction; to create ways for interacting with devices in situations where it is difficult or impossible with today's visual-manual user interfaces. Examples of such situations are listed in Publication I.

It seems, however, that current research on multimodal systems has concentrated on building technical enablers for flexible multimodal interaction. Less research has been conducted on what *should* be possible and what the user benefit of a multimodal system would be in certain contexts of use.

Several architecture proposals exist for implementing flexible user interfaces that allow switching between modalities according to varying mobile usage situations and the user's preferences. A term called 'plasticity of user interfaces' has been proposed by Thevenin and Coutaz (1999) to describe model-based user interfaces that can be distributed over several SW platforms and devices and whose presentation can vary depending on the available devices and/or the user's context.

Bouchet and Nigay (2004) present a component-based framework for designing flexible, multimodal user interfaces. The framework has also been utilised in the mobile domain (Serrano *et al.* 2006). However, no evaluation is presented for what kind of situations would most benefit from the concepts created by the framework.

Roessler *et al.* (2001) describe a system aimed at offering multimodal Web-based services in mobile environments. They propose requirements for a multimodal mark-up language, MML, which would combine the advantages of visual-manual I/O typically used with HTML, and spoken I/O as used with VoiceXML. The paper addresses problems with synchronizing input from different modalities – solving the ambiguity in the “put that there” problem as described by Bolt (1980).

Kernchen *et al.* (2005) describe a concept where the system utilises the different devices that exist in the user's context, to create a representation of a multimodal user interface. In a similar manner, Eisenstein *et al.* (2000) have created a model and a modeling language for multimodal, multi-device user interfaces.

Mueller *et al.* (2004) compares various UI description languages (UIML, InkXML and VoiceXML) and considers their poor suitability for describing flexible, multimodal user interfaces. The paper presents the MMI (Multimodal Interaction) framework that was established in the W3C (World Wide Web Consortium), and proposes a new markup language called DISL for creation of general, platform-independent multimodal user interfaces.

The W3C multimodal interaction activity (W3C 2010) has resulted in yet another markup language for describing multimodal interfaces: EMMA (Extensible MultiModal Annotation markup language).

A stereotypical multimodal user-interface concept is that concentrating on combining manual input – typically pointing – and speech. Other modalities are often mentioned as requirements for future extensions, but in practice they are seldom utilised. This can be seen, e.g., in the W3C multimodal interaction use cases (W3C 2002), all of which concentrate on manual pointing and speaking. The same phenomenon can be seen also in studies where existing multimodal user interface concepts have been analysed (Oviatt *et al.* 2000, Martin *et al.* 1998).

There is nothing wrong in defining flexible systems that enable plasticity in the UI. However, it seems that the ways of, and motivation for, utilising them are still unclear. Commercial applications of multimodal user interfaces combining speech with pointing are rare, especially what comes to mobile applications. In fact, the design of an ultra-mobile system as described by Holzman (1999) shows an evolution away from the initial pointing-based interface into a system utilising spoken input and auditory output. It has also been mentioned that pointing and speaking works well within a user interface that is inherited from the world of computers, with windows and mice (Oviatt 1999). In the user interfaces of handheld devices, this is typically not the case. Furthermore, pointing often requires the use of two hands (one to hold the device, the other to point) which is often undesirable in a mobile context. The performance of speech recognition alone has been shown to degrade from an error rate of less than one percent to tens of percents when the background noise has been changed from a quiet position into a cafeteria (Das *et al.* 1993), which can be considered to be a quite normal mobile use case. So, especially if the inputs from two modalities are used independently so that mutual disambiguation is not possible, the poor performance of the speech recogniser and the requirement for two-hand use cause the combined speaking and pointing paradigm to be quite poor for mobile use!

In an ultra-mobile usage context, it is especially the visual channel that is often overloaded. Both the mobile device and the user's surroundings require the use of vision. Therefore, providing means for eyes-free interaction are an obvious choice for improving the mobile user interface. Concepts that utilise manual input combined with multimodal output exist. Brewster *et al.* (1995) has presented a system that provides both visual and auditory feedback from using touch-screen buttons. Brewster *et al.* (2007) present a system where touch-screen feedback is provided through the tactile and visual modalities. Lee and Spence (2008) have presented a system where feedback from touch-screen buttons is provided through a combination of visual, auditory and tactile modalities. Such concepts have also been recently introduced into commercial devices. At least Nokia, Samsung and LG have introduced touch-screen-based mobile phones that provide visual, auditory and tactile feedback from using the touch-screen objects.

Roto and Oulasvirta (2005) point out that the need for eyes-free usage can originate from the user's attention switching away from the mobile device during time-consuming operations such as when downloading a Web page. As a solution, they suggest multimodal output to be utilised for informing the user that the device is again ready for input. Tactile feedback is mentioned as an example of such output.

As an example of a hands- and eyes-free system, Brewster *et al.* (2003) present a system where the output is generated with sound, and detecting the user nodding his head is used as input. In the same publication, Brewster *et al.* also looked into a system where the user could draw 2-dimensional strokes on a touch-screen surface, receiving auditory feedback from the stroke.

One problem of manual input is indeed that manual input is typically manual/visual interaction. The desirable size of a mobile device is very small for easy portability. This leads into the used input techniques to be small as well. Hitting small buttons therefore becomes not only a problem for manual input, but also a search problem; the user often needs his vision to find where the buttons are and to read the button printings. On touch-screen user interfaces the problem is even clearer since the user cannot even feel around the keypad but the vision needs to be used just to find the buttons.

Pirhonen *et al.* (2002) compared users' performance on a stroke-based interface to a (virtual) button-based one. In the study, a user interface based on touch-screen gestures and auditory output for controlling a music player was studied and compared to a traditional user interface that offered virtual buttons to be pressed on screen. The study revealed significant improvement in the perceived workload and the task performance when the users were walking.

Raman and Chen (2009) present a similar system for text entry, aimed at eyes-free interaction. They point out that with stroke-based interaction the requirement of absolute pointing turns into that of where there is a fixed target the user must be able to hit, turns into that of relative pointing where the finger stroke can start almost on any point of the screen. The system of Raman and Chen originates from the world of universal accessibility, providing interaction methods for the visually impaired. Another example of such a system is the screen reader, which speaks out a description of the object the user is pointing at. Examples of screen readers are the Nuance TALKS (Nuance 2010) for the Symbian S60-based devices, and the VoiceOver for the Apple iPhone (Apple 2009). Raman and Chen point out that the same concept also works when the mobile usage context requires the user to look elsewhere than the mobile device. Arguably, the demands of a sighted user may be higher since the screen reader is not the only way to access the device. But nevertheless, Raman and Chen point out the same phenomenon – the situationally induced impairments and disabilities - as has been described by Sears and Young (2003).

Hand gestures can also be performed by holding the whole device and detecting its movement, e.g., by the use of an accelerometer. An example of such a system is the Shoogle concept (Williamson *et al.* 2007) where the user shakes the whole device to obtain information of its state. Presentation happens through auditory and tactile channels. The system still requires some level of manual input, but finger precision is not needed, nor any visual contact required to use the device. Other examples are the Sony Ericsson W910i phone which the user can shake in order to randomize the list of currently playing music, or the Samsung SCH-310 phone where the user can, e.g., draw digits in the air in order to dial a phone number. In a similar manner, on the Nokia 8800 Arte phone the user can silence an incoming call by turning the phone face down. A double tap gesture on the phone's surface can be used to switch on the display backlight and display a clock. On the Nokia 5500 Sport phone the double tap can be used, e.g., to make the phone read out a received text message or information about the current sports activity, by using speech synthesis. The same can be activated also by pressing and holding one of the hardware buttons on the UI. This feature has also been implemented into other Nokia phone models so that after receiving a text message the user can hold down a button to make the phone read the message out using synthetic speech.

Hand movement can be also recognised by using a camera embedded in the device. The Sony Ericsson W380 phone on the other hand uses its camera to detect a hand moving near the phone. The user can silence a ringtone or snooze an alarm by waving his hand by a phone lying, e.g., on a table.

Other commercial multimodal mobile user interfaces include the YapMe service (YapMe 2008) where the user can input text in a visual application by using speech recognition. Spoken commands have also been widely used in mobile phones to initiate phone calls in an eyes-free manner. Some phone models also allow using spoken commands as shortcuts to certain activities such as switching the phone's Bluetooth functionality on or off.

The OpenStream corporation has published a multimodal browser called CueMe (OpenStream 2008). The browser allows spoken and/or manual input and spoken/visual output. Based on the demo video at the OpenStream website (www.openstream.com/demo/cueme_demo.swf) it also allows simultaneous manual and spoken input, implementing the “point and speak” paradigm.

As multimodal output, many mobile phones offer the use of vibrotactile output in addition or replacing auditory alerts. For instance, an incoming phone call may be presented by using a tactile and/or auditory alert in addition to the visual presentation.

As mentioned, vibration has also been used to augment the visual display. On certain devices, vibration is also used as feedback from pressing hardware buttons. On the Nokia N95 (and most of the newer models) the device produces a short vibration pulse as feedback for switching the device on. A rather long keypress is required to switch the device on, so producing the feedback through the tactile modality allows the user to hold the device and, e.g., continue to walk without having to stare at the screen to know when the button can be released.

The auditory modality is in mobile devices typically used for catching the user's attention. Feedback sounds during use have been proposed,, e.g., by Leplâtre and Brewster (2000) for assisting in menu navigation, but so far the practical implementations are rare. Immediate feedback from pressing a button or interacting with the touch screen is common, as are sounds related to various notes such as warnings or error messages. For example dynamic interaction sounds similar to the SonicFinder concept presented by Gaver (1989) have so far not been utilised in commercial mobile devices. The SonicFinder augmented the Finder user interface existing in Apple computers at the time. Typically it would produce an auditory equivalent of visual feedback – e.g., a sound from tapping objects of different materials depending on what kind of an object the mouse click occurred. However, sometimes the auditory icons provided information only available through the auditory modality. For example, clicking on a file or an application sounded different even though the visual icons were the same for both.

Speech recognition is today a common way to initiate phone calls on mobile phones. It can replace the usage of manual input and visual feedback (although there typically is a button press required to initiate the speech recognition) with spoken input and output.

Table 3 presents the concepts and products described above, and lists which class in the classification by Martin (1997) they belong to. As can be seen, only the CueMe multimodal browser belongs to the Complementarity class., i.e. the one including the often studied “point and speak” paradigm. The SonicFinder includes some cases of the Complementarity class, such as the abovementioned example of the auditory icon providing more feedback from clicking on an object than what can be obtained from the visual presentation. However, the vast majority of the presented concepts belong to the Redundancy / Equivalence classes. The same information is presented through one or several modalities, or a same input can be provided through alternative modalities. Major causes for this may be that the currently existing systems for context awareness are not yet mature enough so that they would offer only the suitable modalities at a time. Instead, everything is made available simultaneously and it is left for the user's discretion which modalities to use. Another cause may be that the benefits from the Complementarity class are not clear.

Table 3. Existing mobile multimodal concepts and their type according to Martin's classification.

Device/concept	Feature	Martin's classification
Brewster *et al.* (1995)	Touch-screen input, auditory & visual output	Redundancy
Brewster *et al.* (2007)	Touch-screen input, tactile & visual output	Redundancy
Lee and Spence (2008)	Touch-screen input, tactile, auditory & visual output	Redundancy
Nokia 5800 XpressMusic, Samsung SCH-W559, LG Prada (and other touch-screen phones)	Touch-screen input, tactile, auditory & visual output	Equivalence/Redundancy (depending on user settings)
Roto and Oulasvirta (2005)	Visual & tactile output when a time-consuming task ends	Equivalence
Brewster *et al.* (2003)	Head & hand gesture input, auditory output	Specialization
Pirhonen *et al.* (2002)	Button or stroke input to control music player, auditory & visual output	Equivalence
Williamson *et al.* (2007)	Gesture input, auditory and/or tactile output	Equivalence / Redundancy
SonyEricsson w910i	Button or gesture to control music player	Equivalence
Samsung SCH-310	Button or gesture to input numbers	Equivalence
Nokia 5500	Button or gesture to initiate speech	Equivalence

Device/concept	Feature	Martin's classification
	synthesis	
Nokia 8800 Arte	Button or gesture to silence ringtone	Equivalence
Nokia message reader	Button input, spoken and/or visual output to present message contents	Equivalence
Sony Ericsson W380	Gesture or button input to silence ringtone	Equivalence
YapMe (YapMe 2008)	Speech or button input to enter text, visual output	Equivalence
CueMe (OpenStream 2008)	Speech & manual input, Auditory & visual output for browsing	Equivalence / Complementarity
Nokia N95 power key vibra	Button input, tactile & visual output	Equivalence / Redundancy
Mobile phone alerts and notifications	Visual, auditory and/or tactile output for alerts and notifications	Equivalence / Redundancy (depending on user settings)
Audio-enhanced menu navigation (Leplâtre & Brewster 2000)	Button input, visual and auditory output	Redundancy
SonicFinder (Gaver 1989)	Mouse input, visual and auditory output	Specialization & Redundancy & Complementarity (depending on UI event)
Mobile phone (voice-enabled) dialing	Button or speech input, auditory and/or visual output	Equivalence

4.2 Specific designs for ultra-mobile usage contexts

In this section, specific designs from Publications II – VII are presented. The designs aim at augmenting a graphical user interface with sounds (Publications II, V, VI and VII) in order to enable eyes-free operation or to create practical and industrially useful methods to use sound as an output technique in a user interface. The study presented in Publication III aims at revealing limits for utilising a time-out for automatically changing the state of the device. Such a time-out based user interface could be made very simple because the user would only need to provide an input at the correct time. However, a wrong timing could lead into error situations that are difficult to recover from. Publication IV studies various ways to use accelerometer-based input on the device which would relieve the user from having to be able to operate buttons that may be difficult to find when the eyes are occupied, or difficult to press when the hands are otherwise occupied or the device is not held in hand at all.

4.2.1 Speech assisted text entry

Publication II describes a system where the user can enter text using a handheld devices using multi-tap input, with little or no need to use his eyes. The system concentrates especially on multimodal output with no changes to the existing input modalities.

Multi-tap input is widely used in mobile devices. Its principle is that certain characters are entered by pressing a button multiple times within a time limit. For instance, in order to enter a letter 'c' on an ITU-T compatible (ITU 2001) keypad, the user needs to press the button '2' three times in a row. If the user waits too long between the keypresses, a so-called multi-tap time-out occurs and the next successive keypress is interpreted as the user wanting to enter a new character from the choice of characters associated to that button. For instance, if the user presses the button '2' twice, and waits for a longer period than what the device manufacturer has defined for the multi-tap time-out, the character entered is 'b' and the next press of the same button will bring a new character ('a' if no further button presses are made) onto the screen. The occurrence of the multi-tap time-out is typically presented with a visual cursor appearing on the screen again (after being hidden during the multi-tap entry of a character). The behaviour of the system is presented in Figure 2.

Fig. 2. Speech-assisted text entry concept (I, published by permission of Taylor & Francis).

As can be seen from the description above, text entry using multi-tap input is a very visual task. The time-out associated with the text entry requires the user to constantly follow the characters on screen so that errors due to waiting too long between the button presses can be avoided. Especially if there is a simultaneous visual task (such as walking in a crowded place), it can easily happen that the user's visual attention is drawn elsewhere and the device state is no longer clear.

On the other hand, for accessibility reasons of mobile devices state that the middle button (button '5') in the ITU-T compatible keypad is marked with tactile cues so that it can be found without having to look at the keypad (ITU 2001). The rest of the keys are expected to be found by the user feeling around the keypad. Therefore, the manual input does not necessarily require the usage of the eyes. It is only the visual output of the system that requires the constant visual attention to be on the device.

Publication II describes a concept where the visual output of the system is augmented with auditory output. Each keypress causes the letter currently on screen to be spoken out by the device. The cursor appearing after the multi-tap time-out is presented by a small non-speech sound. The concept therefore allows eyes-free text entry. In the classification by Martin (1997) the concept belongs in the Equivalence class, since the user can freely decide whether to utilise the auditory or visual output of the system.

A possible downside in the system that was identified is that the spoken output from the system may disturb the user when trying to formulate one's thoughts into text, and when keeping simultaneously in mind which characters of the current word have been entered, and what the next character should be. However, the extent of this problem was unknown before testing the application.

The application was tested both in the laboratory and in the field, in an office building where doors and staircases had to be taken into account when walking. The test was conducted in daytime conditions, with good illumination and negligible noise levels. The variables measured in the test were: Typing speed, Percentage of errors made, Percentage of uncorrected errors, Walking time and Glances away from the display. The latter two were only measured in the field condition. The walking times were also converted into percentages of the preferred walking speed for each user, as the absolute walking times could obviously vary between each user due to natural differences.

Table 4 presents the results from the walking conditions. The nominal performance in the typing speed mentioned in the table was obtained from the regular typing speed (with a visual/manual interface) in the laboratory condition.

In other words, the 100% typing speed for each user was the speed he could reach when sitting down in a laboratory, using the traditional UI for text entry. The nominal walking speed for each user was simply measured in the walking condition without any simultaneous typing tasks.

Table 4. Findings (% of nominal performance, or error % in case of typing errors) in different feedback (FB) types.

FB type	Typing speed (%)	Rem. Errors (%)	Tot. Errors (%)	Walk time (%)	Glances away from display
Silent	87.0	0.99	7.8	149.5	5
Speech	88.2	0.55	7.1	150.2	6.7

The results showed that in the laboratory setting, the spoken output from the system clearly did disturb text entry. For comparison, one expert user tried the system in the laboratory after the test, and did reach the same performance with and without the spoken output – but at the cost of greatly increased cognitive load.

Interestingly, in the field condition, there was an effect that the text typing speed did not differ between the audio-augmented and pure visual conditions. In other words, the disturbance caused by the mobile usage environment itself was evident in the users' performance (as the typing speed was slower than in the lab) but the effect caused by the speaking system did not further disturb the performance. The possible differences in the cognitive load were not measured. The system, however, did not bring a clear benefit for text entry either, at least with novice users. Two users (out of 12) tried if the system would enable them to spend more time looking at the environment than at the mobile device, which showed as a clear difference in the variable “Number of glances away from the display”. One of the users glanced away from the display 11 times when typing without the auditory feedback, and three times when the auditory feedback was active. For the other user, the number of glances away from the display was 12 times and one time, respectively.

The different feedback schemes did not lead to different walking times either. A similar finding has later been discovered by Mizobuchi *et al.* (2005). In their test, text entry as such slowed down walking, but text entry tasks differing in their cognitive and motoric demands did not lead to differences in the walking speed.

When evaluating this user interface concept against the framework of Publication I, one can see that the environments where the framework-based analysis shows that the user has limited (e.g., one-handed) but existing capability for manual operation, and very limited possibilities to use the vision, this concept could be useful. However, in its current form the problems with the cognitive load probably would not make it very useful in a real product. Even in the relaxed laboratory environment using the system clearly increases the cognitive load. Spoken input during the multi-tap text entry of each character would probably need to be removed, or, e.g., replaced with non-speech sound. Perhaps a more suitable place for providing spoken feedback would be after each typed word (i.e. whenever a space is entered) where it would still be possible to check for possible typing errors. Furthermore, the spoken word might even help the user to formulate the whole sentence.

The probable reason for why the system did not perform well was that the speech output it produces disturbs the inner speech of the user. This brings up an interesting point that a suitable cognitive model could have predicted this problem. However, expertise for creating and using such a model was not available during the study.

4.2.2 Time-out in user interface – case: mobile text input

Publication III handles the same subject area as Publication II, namely that of mobile text entry. However, the viewpoint in Publication III is completely different. The multi-tap time-out is studied from the point of view of how useful the concept of a time-out is in user interfaces. In small devices with restricted hardware capabilities (often simply due to the physical size of the device) the time-out is sometimes used as one user interface technology. For instance, the number of buttons that can be fit on a small device is dictated by the device size and by the ergonomic sizing and layout of the buttons. Therefore, it is sometimes lucrative to assign different modes to the buttons, i.e. that their meaning changes depending on the state of the UI. A similar approach is used also in traditional keyboards where holding the Shift key makes the keys to produce upper case letters. A Caps Lock key can be used to switch between permanent upper case and lower case modes.

In mobile user interfaces, using multiple keys to switch modes is typically unattractive because of the limited number of available buttons, and because of demands for one-handed use. Pressing two buttons simultaneously typically requires the usage of two hands. Therefore, the mode is sometimes changed by using timings; when a pre-set time-out occurs, the system changes the UI mode automatically.

Examples of such an automatic mode change are the time-out in multi-tap text entry as described earlier, or in the case of locking and unlocking the keypad in Nokia phones. Time-out is used also in desktop systems, e.g., for deciding how two successive mouse clicks shall be interpreted – as a double click, or two successive single clicks.

The research questions in Publication III were about whether the users remember the length of a time-out they are used to, and whether they are able to learn the length of a new one during usage. Three different feedback types were compared (visual feedback, auditory feedback and indirect feedback where the user's typing errors inferred that the time-out is different from what he thought).

Ideas behind the study were that it could be possible to adapt the time-out for a user, if learning effects would not be too strong. For instance, multi-tap text entry could be sped up by shortening the time-out length automatically, if the user would seem like a fast typist. But if a user would have learned the original length so well that any changes to it would disturb his performance, then this obviously could not be done.

Another use for the study would be when a device utilises different time-out values in different places of the UI. For instance, during text entry a relatively short time-out for a long-press could be used when the user holds down the 'Delete' key to erase the whole text. On the other hand, when the device is locked, a long-press of a certain key combination could be used to wake it up. In order to prevent accidental opening of the keys, a longer press could be required in this case. Now, if the user would have learned the shorter length during text entry, he might be inclined to release the press after the shorter duration when unlocking the device. It might then become irritating if the keys would not open when expected.

Another aspect that is shown in Publication III is the extent to which users rely on the feedback from the device, with regard to time-outs. If the users would clearly learn the time-out well, then it would not be so important to present the occurrence of the time-out. In environments where the user's senses are loaded, it could be a benefit if the user did not necessarily have to be explicitly informed about the occurrence of the time-out. Different modalities (auditory and visual) for presenting the time-out occurrence are also compared in Publication III. If the time-out length was learned more easily, e.g., when using an auditory feedback, then obviously this information could be used in user interface design.

The results in Publication III show that the users clearly do rely on the feedback they receive. This can be seen, e.g., in the fact that they reacted more quickly to auditory feedback than to visual feedback. This is congruent with results of simple reaction time tests where an auditory stimulus provides a quicker response on the average (Woodson *et al.* 1992: 630–631).

The results did not show evidence of long-term learning of the time-out length. This was evident at the beginning of the test where the used time-out length was the same as in the participants' own phones, but the feedback was of the indirect type – i.e. no visual or auditory presentation was made at the moment when the time-out expired. The users' estimation of when the time-out had occurred was far off the real length.

Results did show short-time learning results. When the time-out was kept constant for a while and users got feedback from its occurrence, removing the feedback again led to clearly better estimations of the time-out than at the beginning of the test. This can be seen in Figure 3 where the conditions labeled "none" are those where no direct feedback was given from the time-out occurrence.

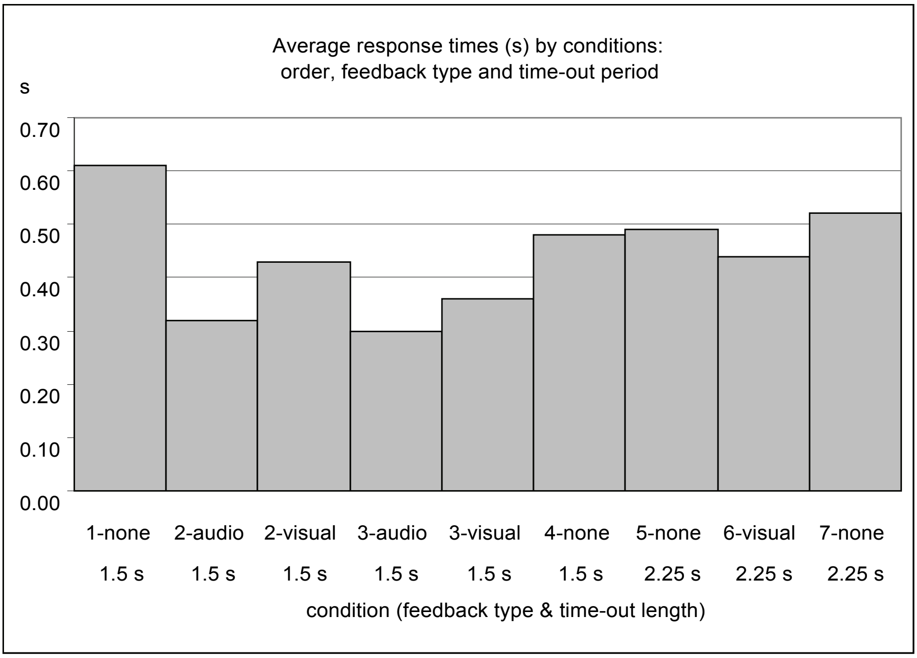

Fig. 3. Mean response times by condition (III, published by permission of Springer).

The feedback in the first text block typed in the test was simply the possible mistakes the users made when typing too late. No visual or auditory feedback was given from the occurrence of the time-out. The length of the time-out (1.5s) was familiar for the test participants, as it was the same as in their own mobile phones.

The response time was measured from the occurrence of the time-out, i.e. the time when the user could have typed the next successive character from the same button, to the time the user actually entered the next character. One can see that with feedback (text block 2) this time decreased dramatically. This is no surprise, but it does tell us that users had not learned to estimate the familiar time-out length accurately. However, in the next block where no visual or auditory feedback was presented (block 4) one can see that the response times are faster than in the first one (block 1). This suggests that short-term learning had occurred during the text blocks where feedback had been available.

When the users are presented with a new, increased time-out length in text block 5, it appears as if the users would react almost as quickly to the new time-out as to the previous one. However, this is a byproduct of the used measurement method. If a user presses a key before the time-out occurs, it is not registered as a reaction to the time-out but as a so-called anticipation error. Therefore we must look at both the reaction times and anticipation errors in order to understand the results of the test.

Figure 4 presents error percentages related to all conditions. The errors consist of anticipation errors, i.e. the user pressing a key before the time-out has occurred. The percentages are related to the number of successive same-key characters being entered (e.g., an 'a' followed by an 'a', 'b' or 'c') – not to the overall amount of entered text. In other words, in the first text block, where no feedback was provided, more than half of the successive same-key characters led to the user pressing the key too early for the latter character.

In text block 5, the users made significantly more errors than in block 4. In other words, the real time-out length was longer than they expected, but since there was no feedback from the time-out occurring, the users simply pressed too early. This backs up the finding that users indeed had learned the shorter time-out length in the previous text blocks.

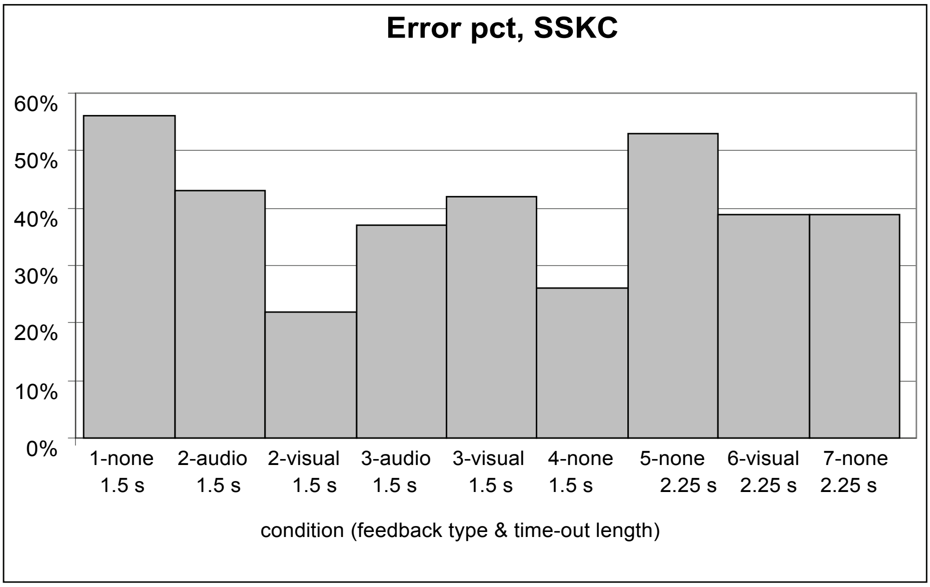

Fig. 4. Mean error rate ratio to typed SSK characters in each condition (III, published by permission of Springer).

Another interesting finding is that in text block 6 the error rate dropped, staying in the same level when the visual feedback was removed again. Apparently the users were able to learn the new, longer time-out quite quickly.

When considering the findings in Publication III in the viewpoint of the analysis framework of Publication I, it is evident that the concept of automatic, time-out based mode change for buttons is a useful method to be used in a mobile UI, but not without limitations. First, if the users are repeatedly subjected to a certain time-out length, then introducing a new length is likely to cause errors. In a system that has random delays, it becomes problematic if these delays affect the length of the time-out to which users are repeatedly subjected. Second, the users cannot be expected to remember a time-out length over a period of time.

In the test, users clearly relied on the received feedback. Auditory feedback led to quicker reaction times than visual feedback. Therefore it is important that the feedback of the system mode change is presented at all, and the modality through which it is presented, is available for the user in the current usage environment.

4.2.3 Gesture input in ultra-mobile usage contexts

Publication IV discusses the concept of utilising hand gestures that the user performs by moving the mobile device, either by holding it in hand or by tapping it with fingers when the device is in its storage location such as a pocket. Typically these gestures can be recognised by utilising an accelerometer or a camera embedded in the mobile device, detecting its movement. Possible uses for this kind of interaction are illustrated by the analysis framework when situations for limited manual and visual use are encountered. A hand gesture allows device manipulation without requiring the delicate accuracy of finger operation. Instead, the user can, e.g., hold the device and wave it, shake it, or move it in a circular motion. One hand obviously must be available for device use (at least for the duration of the gesture) but the user need not, e.g., be able to locate the buttons on the device surface or to press them.

With a new technology whose usage is potentially visible to the other people in the usage environment, the question of public acceptance is very relevant. This issue was addressed in Publication IV. A number of gestures were performed by actors in public places, and these were videotaped. The short videos of gesture usage were then placed on a Web page with a questionnaire about how likely a person answering the questions would see him using the gesture, and why. The choices in the questionnaire were deliberately aimed to reflect the social acceptance ("No, it looks silly") or overall utility of the feature ("No, it is not useful"). However, in practice it was often difficult to say exactly what the reason for using or not using a certain gesture was. Hence, the authors did not try to utilise statistical methods for analysing the questionnaire.

The results showed that only rather discreet gestures were accepted among the participants of the questionnaire. In fact, tapping the device cover to silence a ringing phone was the only one clearly leading to positive answers among the participants. Using the same gesture for controlling in-call volume led to users not objecting the feature because of its social acceptance, but almost 40% of the participants simply would not see value in using it.

Drawing patterns in the air was unacceptable – the most objected features because of the social context were drawing an 'X' or an 'O' in the air with the device. The results are presented in Figure 5.

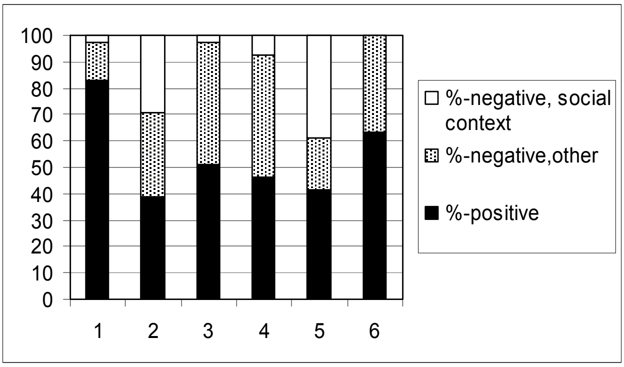

Fig. 5. Positive/negative answers for each video (IV, published by permission of ACM).

The gestures presented in each video were as follows:

1. User enters a café. Phone rings. User taps the phone through the pocket to silence the ringtone.
2. User enters a library. When walking, she draws an 'X' in the air with the device, to make it enter the silent mode.
3. User is cycling. He shakes the phone to initiate a call to a pre-defined number.
4. User is walking on a quiet street. She receives a text message. She swings the phone from side to side in front of her, to cause the message to be displayed.
5. User is walking on a crowded street. She draws an 'O' in the air with the device in order to open the calendar application.
6. User is talking on the phone. She taps the sides of the phone to adjust the in-call volume.

Other discreet gestures than tapping the device cover appeared to be acceptable for the users, but the applications they were used for, were not seen valuable. Neither shaking or swinging the device were objected because of the social context, but the features they were controlling (shaking the device to initiate a call to a predefined number, swinging the device sideways to see the contents of a received text message) did not gain very positive remarks among the participants.

A discreet gesture, namely that of tap input, was selected for further development. Two user tests were arranged in order to study the usability of certain features being controlled by tap input, and to study the way the users would prefer to tap the device cover, given no strict instructions for how to do it.

Perhaps the most interesting finding in the test was that the force, with which the users tapped the device cover, varied a lot, which should be taken into account when designing a user interface relying on tap input. Another interesting finding was that the feedback from the tap gesture needs to be immediate, confirming that the gesture was recognised. Yet another finding was that single taps and double taps were not useful as separate interaction methods, as users had difficulties remembering the meaning of each in cases where both were used in a single application.

The findings presented in Publication IV influenced the Nokia product creation. The double tap has been used, e.g., in the 5500 and the 8800 Arte models. In each, it can be used, e.g., to silence a ringtone, which was the most widely accepted gesture in the questionnaire presented in Publication IV.

Some of the concepts in Publication IV were direct results from applying the environment analysis framework of Publication I. In the study presented in Publication I, certain usage contexts such as when doing outdoor activities in winter, were found to be problematic from the point of view of the manual input. The mobile device is often buried in a pocket, and there are constraints to using the hands, because of things being carried or because handwear to be worn because of the weather conditions. The analysis also revealed that there are certain contexts, such as when commuting in a train in Tokyo, where social restrictions are high and the user would benefit from a quick way of silencing the phone when it rings. The double tap gesture was found to be useful in this kind of situations, as it can be quickly performed while keeping the phone in its storage location.

4.2.4 Aesthetics of audio-enhanced graphical buttons

When considering multimodal output in a user interface, sound is one natural candidate. In Publication V, results of a test concerning the usage of sound for augmenting a graphical touch-screen-based UI are presented. The study originated from the fact that auditory augmentation of the graphical feedback in pressing touch-screen buttons (successfully or unsuccessfully) proved out to improve usability in an earlier study (Brewster *et al.* 1995). However, the actual sounds used in the earlier study – even though the test participants did not find them irritating – appeared rather intrusive for a commercial device.

In the earlier study, the enhanced usability probably affected the acceptance of auditory user interfaces. Furthermore, as the sounds were only heard in a test setting, it was unclear how well the same sounds would be accepted in real products. Therefore, a repeat of the original test was conducted so that the sound design from the original test was compared against a new design with the only difference being that the new sounds were much shorter, quieter and overall more discreet than the original ones. The results were published in Publication V. The research questions were whether the users would be able to gain the same usability benefit also from the smaller sounds, and which of the three compared systems (quiet/original sounds/new sounds) they would prefer.

The results showed that the usability benefit did not disappear when the sounds were made more discreet. Furthermore, the preference was clear – 9 users out of 10 preferred the new sound design over the original one, and over the quiet situation.

The results of Publication V fit into the field of social requirements in the environment analysis framework presented in Publication I. It has already been shown by Brewster *et al.* that enhancing graphical buttons with sound improves the overall usability of small touch-screen buttons. However, the social context often requires the sounds to be used to be as small as possible, while still maintaining the usability benefit. The results presented in Publication V show that it is possible to make the sounds less intrusive, yet equally useful as more prominent sounds.

4.2.5 Perceived importance of sounds

In addition to enhancing direct feedback from the UI with sounds, sounds are often used to catch the user's attention and draw it to a certain event. The loudness of sound has traditionally been considered as a way to control how strongly the sound catches the user's attention. However, it has been pointed out (Patterson 1982) that this is in the long run not a good approach. If many devices are trying to compete for the user's attention, the overall noise level would rise. Furthermore, at some level the sounds will become irritating just because of their loudness. In the study by Patterson (1982) concerning the alert sounds used in aircraft, it was found that some alert sounds are even so loud that they disturb communication inside the cockpit. A further problem pointed out in the study was that a loud sound can easily draw the pilot's attention to shutting the sound off instead of concentrating on the problem that caused the sound. Too loud sounds are also annoying and disrupt thought and communication.

Therefore, alternative methods have been suggested for designing sound so that the perceived urgency of the event related to the sound is correct. Patterson developed a way to design sounds that consist of bursts of pulses so that the perceived urgency related to the sound can be controlled. Control is implemented by altering various characteristics of the single pulses in a burst, and the characteristics of the burst. In Patterson's original work (Patterson 1982) the urgency levels were discussed, but the relationship between the acoustic characteristics of a pulse burst and the perceived urgency of it were not tested. This work has later been continued by Edworthy *et al.* (1991) and Hellier *et al.* (1993) who have found out that certain characteristics indeed do correlate between the perceived urgency in a predictable manner. These characteristics were the rate of the pulses in the burst, the pitch of the pulses and the irregularity of the harmonics within a pulse. This design paradigm has also been used in hospital settings (Haas & Edworthy 1996), and it appears to be the only widely enough studied way to affect the perceived urgency.

However, it is difficult to apply sounds consisting of pulse bursts widely in a product user interface. As pointed out earlier in the Multimodal Output section, it is often necessary to present a same piece of information through different modalities. Crossmodal design - in the sense described by Hoggan and Brewster (2007) - is needed so that the user understands that, e.g., a visual icon is related to the same UI object as, e.g., a sound. The design method presented by Patterson does not support designing icons that way, as it requires that the sounds consist of bursts of pulses. If a sound designer wants to use a visual icon representing a trash can, it is easy to use a clanking sound for events related to the same object. However, presenting urgency using Patterson's method cannot be easily applied to the clanking sound.

Furthermore, pulse bursts are hardly ever aesthetically pleasing. It is difficult to utilise them in household products. It is also difficult for a product manufacturer, e.g., to build a brand around sounds consisting of pulse bursts. There is not enough freedom in the sound design to enable creation of a common sonic identity for a product.

In order to tackle these problems, a set of new variables for altering the perceived urgency of a sound has been subjected to user tests and published in Publication VI. The publication uses the word “importance” instead of “urgency” in the title, as that was the one the test participants were instructed to evaluate. However, it is discussed in the publication that in practice it most likely was the urgency that the participants were evaluating.

The sound parameters were selected so that they could be used to modify an existing sound, instead of creating a new one from scratch. This would leave more space for a sound designer as the urgency parameters could be “superimposed” on an existing sound. For instance, one could design a sound for alerting the user of an upcoming event in his calendar relatively freely. Changing the way how the alert is played could then be used for showing how soon the upcoming event is going to occur.

Two new parameters were found in the test. Altering the cutoff frequency of a low-pass filter used for filtering the sound was found to affect the perceived urgency of the sound in a statistically significant manner. The speed of a vibrato effect applied on a sound was also found to do the same. Furthermore, it was also interesting (but obviously inconclusive) that varying sound length did not have a clear effect, nor did altering the amount of reverberation applied to the sound.

The idea in Publication VI has been continued by Guillaume *et al.* (2007) by utilising the perceived distance of a sound source as a method for controlling the importance of the event the sound relates to. Moreover, Nicol (2005) has presented a way to modify the perceived urgency without using the pulse burst paradigm. In his study, the timbre of a sound was modified based on the effect of harmonics on the perceived urgency. This effect was first pointed out by Edworthy *et al.* (1991). Nicol's method could also prove useful when creating sounds for commercial devices, as it is based on modifying an existing sound.

4.2.6 Sound design for commercial handheld products

Publication VII presents a summary of the author's experiences in sound design for commercial handheld products. It draws together findings of the author's earlier publications. The publication focuses on the acceptance of sound use and things affecting it. Three basic principles for sound design are presented: Utility and Usability, Disturbance to the Environment, and Aesthetics.

Han *et al.* (2001) point out the importance of the impression that a user carrying a mobile device presents to people around him. Sounds can easily draw the attention of others, so it is important that the sounds are not disturbing for others, and the style of them matches with what the user wants to express to others.

Leplâtre and McGregor (2004) have pointed out that auditory displays, even though they have been studied for decades, still have not become widely used. One reason that Leplâtre and McGregor present is the lack of design quality in auditory displays. They also point out that design quality in practice is often related to the aesthetic quality.

Utility and usability related to usage of sound can greatly affect its acceptance. This issue has been pointed out both in Publication V and in the original study by Brewster *et al.* (1995). When a sound improves the usability of a system in a user-perceivable manner, the acceptance of the sound will be better.

In Publication VII, issues related to catching the user's attention are discussed as well. Often a sound draws the user's attention to it, and if this is done for a thing that the user already knew about, then the sound is less likely to be accepted. The sound is also typically heard by the people around the user. It can draw their attention to it, which can be even more irritating since the event makes no sense to the passers-by. This can then cause the user to feel embarrassment. One proposed solution (Brewster 2002) has been to rely on the user wearing headphones. Headset and headphone usage has become more common with the legislation in many countries requiring hands-free kits to be worn when driving and speaking on the phone, and with the increasing memory sizes in mobile phones having enabled the storage of a larger music library to be carried on the user. However, in Publication VII the current author argues that that the sounds themselves should be designed to be comparable to the sounds naturally heard in the environment. People do make sounds, e.g., by body movements and talking, so device sounds whose attention catching qualities would not be much higher than those, would be more likely to be accepted also by others. It is suggested that the principle of minimalism is utilised when designing product sounds. However, it is pointed out that it is currently unknown what the limits for making the sounds as small as possible while maintaining their usability benefits are.

The aesthetics of sounds in a commercial product is also discussed in Publication VII. It is pointed out that since people naturally relate qualities of the imagined sound source to a sound they hear, it would be possible this in creating the overall identity of a product. The contradiction of keeping the sounds small and simple, and on the other hand enabling the user to hear qualities – such as the material – of a sound source is discussed.

As a real world case study, the sound design process for the Nokia Internet Tablet product family is presented in Publication VII. The three sound design principles are discussed in relation to the issues in real world product creation and design decisions are explained. Some research problems are raised that have become evident during the sound design process of the Internet Tablets. The publication also discusses the relation of research of auditory user interfaces in relation to how real world product sounds are created.

The publication discusses several issues related to utilisation of sound in the user interface of a real, commercial mobile product. When analysing a usage environment according to the analysis framework of Publication I, these should be taken into account. The framework provides means to analyse the usage of sound specifically in a certain environment, and also to analyse the indirect effects that the typical social demands cause in it.

5 Discussion

When looking at today’s mobile user interfaces, one cannot help but feel that they have not come very far from being just miniaturized computers. Based on Apple's iPhone OS, Google's Android, Nokia's Symbian or MeeGo or any other modern mobile phone OS, a common denominator in mobile devices is that their output is mostly restricted to using small screens with amazingly high pixel densities. Content can be presented in high resolution, but user input is restricted to finger-size objects. Simple speech input may be possible, but no natural language dialogue can be held with the device. Some server-based dictation services exist, but they are not very common yet. Visual output can be augmented with haptic feedback from the touch screen. Sounds and haptics are used to draw the user’s attention to alerts. But it seems that everything more elaborate is still waiting in research laboratories.

In this thesis, it has been proposed that a major cause for not utilising all the possible modalities available in mobile contexts is that their benefits are not clear. Concepts that have been implemented are the most obvious ones. Voice dialing while driving is a clear example. Usability testing in mobile environments may reveal usability problems related to the mobile context of use, but testing as such does not suggest solutions to the problems. The solutions therefore tend to be the obvious ones. What is needed is a tool for performing a more comprehensive analysis of the situation, to help the designer think more broadly of what the solutions could be.

In this thesis it has been shown that currently no comprehensive analysis tools exist that would clearly and systematically show which of the users capabilities are overloaded in a certain context, and which could be exploited more than they are today. An analysis framework has been proposed as one solution to the problem. The starting point in the framework is indeed in going through the user’s capabilities one by one, and analysing how the environment around the user affects them.

The presented framework has been used to analyse several mobile usage environments. During the analysis, new interaction concepts have been created and their usefulness has been tested. The double tap gesture was identified in the analysis. Its usefulness as well as social acceptance was also studied in more detail in this thesis, and the study was extended to cover other gestures performed using the whole device as well. Another concept resulting from the analysis was using spoken output to augment the visual feedback in a text entry task. This concept has also been analysed in more detail in this thesis. The framework has also been used for evaluating an existing concept, namely that of haptic feedback from touchscreen interaction. The concept was known at the time, but the framework helped in showing the contexts where the concept would be most useful. This served as a trigger to study the concept more thoroughly.

In order to build new interaction concepts, one often needs also new interaction techniques. In this thesis, a concept of utilising a time-out to change the system state has been studied. If users were able to remember accurately the time-out length, it would have the effect that the length should remain constant, and that it would be difficult to adapt it dynamically. On the other hand, the time-out could then be safely used for changing modes in the UI since users could be trusted to have learned the pace of the system. If short-term learning would occur during use, it would cause an effect that variance in the time-out length should be small. However, dynamic adaptation of the feedback length could be utilised if the provided feedback would be enough to teach the users the development of the time-out. The results showed no evidence of long-term learning. However, evidence of short-term learning was found. Furthermore, it was seen that a new time-out length could quickly be learned based on the provided feedback. Adaptation therefore seems to be possible without ruining the usability.

Another interaction technique studied in this thesis is that of UI sounds. Details of UI sounds have been studied so that they would be both useful and commercially viable. The results showed that it is possible to bring the same usability benefit in touch screen interaction by using quite small sounds instead of longer ones. Another study revealed two new methods of modifying UI sounds to control the perceived importance of the event related to them. Existing studies had concentrated on creating new sounds by controlling the parameters used to synthesize them, which restricts the creativity of the sound designer and is problematic for maintaining a consistent audio brand of a company. The new methods found in this thesis were filtering out the high frequencies in a sound, and modifying the speed of a vibrato effect applied to the sound. These can be more easily applied over sounds synthesized using which ever means suitable for product sound design.

The last part of the thesis presents an analysis and a case study of applying sound design techniques into commercial handheld products. The study presents a synthesis of how different sound design paradigms can in practice be applied to the UI sounds of a commercial device. It also presents ways of combining the UI sound design with the visual style of the product. The study discusses the practical problems of utilising sound in user interaction, and points out several currently unresolved key questions for the academic community.

Why are multimodal user interfaces for ultra-mobile contexts important right now? An old and often repeated caricature of a mobile phone is presenting it as a modern Swiss army knife, with all imaginable functionalities being embedded into a single device. A mobile phone could have a razor, a cigarette lighter, a bottle opener, and so on. A common factor for those functionalities typically is that they are all local and it is the user who starts using them.

The number of applications used in today's mobile phones is currently exploding. The mobile phone of today is no longer a stand-alone device with mainly local applications running on it, with the capacity of communicating through messages and phone calls. Instead, it is developing into a device connecting the user to online communities and services. The one-to-one communication is vastly expanded by things happening in the community, which are then communicated to everyone in it. One user may upload a video to an online service while another writes a new chapter to his blog. The weather forecast can updated regularly during a day, and a news service can keep feeding news to the world.

This change leads to a situation where the user is no longer always personally contacted when there is something interesting happening. In contrast to friends sending text messages to each other, an updated blog entry will not be personally informed to the user. But the mobile device can keep track of what is happening and inform the user. The online communities also increase the number of times when the user wants to communicate something to the outside world. Furthermore, the increasing number of available services leads into more frequent situations when the user wants to access some information on the mobile device.

The mobile device must in the future be capable of informing much more and more complex information for the user even when it is being stored away from the user's sight. This causes new requirements for utilising other senses than the user's sight for presenting this information. Another cause for a need of more multimodal output is that many internet services have been designed for PC-size screens. Rendering the same content onto a small screen of the mobile device obviously works, but can lead to an excessive amount of scrolling and difficulties to get an overview of where everything is. Placing more information into a screenful on the other hand leads to difficulties to see what is on the screen, and especially on a touch-screen device the touchable targets may become too small to be hit comfortably. Multimodality could be used to present some of the information, e.g., as sound instead of placing it on screen, as discussed by Brewster *et al.* (2003). On the other hand, multimodality can aid the user in interacting with small objects on a touch screen as presented by Brewster *et al.* (1995) and discussed also in Publication V.

In multimodal input, a clear benefit would be gained with systems that do not require finger operation, or even hand operation in ultra-mobile usage contexts. The hands are often occupied with activities related to the context, and the device can be stored inside a pocket. Recognition of head movements as presented by Brewster *et al.* (2003), finger taps through the pocket as discussed in Publication IV, wrist movements as described by Fallman (2002) or maybe even eyeblinks could be used as input methods in the future. Using the whole device as an input method as described by Williamson *et al.* (2007) would relieve the user from precise operation of device buttons or the visual search problem related to finding the correct button to press. Camera-based recognition of the user's hand or head gestures could also be used in contexts where the device is placed so that its camera can see the user.

Open research questions for multimodal user interfaces clearly still remain. Presenting a same piece of information will be important when creating alerts and notifications that can be presented through different modalities. In the case of using multimodal output to save screen space, the same piece of information needs to be presented sometimes using graphics, sometimes using haptics or sound. Hoggan and Brewster (2007) show that often it would be beneficial to be able to switch between different unimodal interaction methods. They also show that in order to, e.g., present an alert through several possible modalities, it is important to design the presentations so that people recognise them as representing the same event. In order to do this, research on crossmodal perception is needed. Hoggan *et al.* (2008) have also shown that crossmodal congruence is important also in the case of providing tactile feedback from visual buttons – where the user does not have to guess or learn what the tactile feedback means. In their study, users expected a certain look of a button also would have a certain feel.

Typically with new interaction concepts come issues related to acceptance of new interaction methods. The visual sense is still very strong when the user needs to search for a piece of information or when the information to be presented is itself visual. Furthermore, people are used to visual displays. Therefore, the question remains whether people would rather just stop what they are currently doing and start operating a mobile device, or whether excellent new interaction methods would lead them into using devices while moving. It can be assumed, though, that the more always active online services will be used through mobile devices, the greater will become the likelihood of the user wanting to receive a piece of information and to react to it even in ultra-mobile usage contexts.

More research is therefore needed on the requirements for interaction in ultra-mobile usage contexts. The framework presented in Publication I provides a tool for doing this research – but the research obviously needs to be conducted.

This research and the methods for multimodal interaction would probably first be in the interest of manufacturers of mobile devices or mobile software platforms. After all, the tools typically need to be there first: only then can application developers start using them. This work has started already with software platforms offering application programming interfaces for auditory and haptic output as well as accelerometer-based motion input.

Optimally, a software platform would offer the tools for multimodal interaction automatically based on the user's context, without an application developer having to be an expert on multimodality. Eisenstein claims: "We believe that user interface modeling will be an essential component of any effective long term approach to developing UIs for mobile computing" (Eisenstein *et al.* 2000). This may well be true for future mobile software platforms.

6 Conclusions

In this thesis, user-interface design for ultra-mobile usage contexts is discussed. It is pointed out that current usability testing methods are not always adequate for finding out usability problems in ultra-mobile usage contexts. Furthermore, those methods that are still lack the capability of suggesting the use of alternative modalities for solving the problems caused by the mobile environment. It is shown that ways to analyse mobile contexts to come up with detailed requirements for ultra-mobile use have currently been lacking. As a solution, an analysis framework that takes the senses and capabilities of the user as well as possible recognition capabilities of the mobile device as a starting point is proposed. Two novel interaction concepts are proposed for solving certain interaction problems in ultra-mobile contexts. Three interaction techniques are studied for using as building blocks when creating new interaction concepts. Finally, a case study of applying design methods for creating the UI sounds for commercial handheld devices is presented.

References

Adipat B & Zhang D (2005) Interface design for mobile applications. Proc 11th Americas Conference on Information Systems. Omaha, NE, USA: 2281–2290.

Apple (2009) Press release (June 8 2009): Apple announces the new iPhone 3GS - the fastest, most powerful iPhone yet. URI:https://www.apple.com/uk/pr/library/2009/06/08iphone.html.

Baber C & Mellor B (2001) Using critical path analysis to model multimodal human–computer interaction. International Journal of Human-Computer Studies 54(4): 613–636.

Baillie L & Schatz R (2005) Exploring multimodality in the laboratory and the field. Proc 7th International Conference on Multimodal Interfaces ICMI 2005. Torento, Italy: 100–107.

Barnard L, Yi J, Jacko J & Sears A (2007) Capturing the effects of context on human performance in mobile computing systems. Personal and Ubiquitous Computing 11(2): 81–96.

Benali-Khoudja M, Hafez M, Sautour A & Jumpertz S (2005) Towards a new tactile language to communicate emotions. Proc IEEE International Conference on Mechatronics and Automation. Niagara Falls, Ontario, Canada: 286–291.

Berg S, Taylor A & Harper R (2003) Mobile phones for the next generation: device designs for teenagers. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2003. Ft. Lauderdale, Florida, USA: 433–440.

Beyer H & Holzblatt K (1997) Contextual Design: Defining Customer-Centered Systems, 1st edition. San Francisco, USA, Morgan Kaufmann.

Blattner M, Sumikawa D & Greenberg R (1989) Earcons and icons: their structure and common design principles. Human Computer Interaction 4(1): 11–44.

Bolt R (1980) Put that there: voice and gesture at the graphics interface. ACM Computer Graphics 14(3): 262–270.

Bouchet J & Nigay L (2004) ICARE: a component-based approach for the design and development of multimodal interfaces. CHI '04 Extended Abstracts on Human Factors in Computing Systems. Vienna, Austria: 1325–1328.

Brewster S (2002) Overcoming the lack of screen space on mobile computers. Personal and Ubiquitous Computing 6(3): 188–205.

Brewster S & Brown L (2004) Tactons: structured tactile messages for non-visual information display. Proc AUIC 2004. Dunedin, New Zealand: 15–23.

Brewster S, Chohan F & Brown L (2007) Tactile feedback for mobile interactions. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2007. San Jose, California, USA: 159–162.

Brewster S, Lumsden J, Bell M, Hall M. & Tasker S (2003) Multimodal 'eyes-free' interaction techniques for wearable devices. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2003. Ft. Lauderdale, Florida, USA: 473–480.

Brewster S, McGookin D & Miller C (2006). Olfoto: designing a smell-based interaction. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2006. Montréal, Québec, Canada: 653–662.

Brewster S, Wright P, Dix A & Edwards A (1995) The sonic enhancement of graphical buttons. Proc IFIP Interact'95. Lillehammer, Norway: 43–48.

Brown L (2007) Tactons: structured vibrotactile messages for non-visual information display. PhD thesis. Univ Glasgow, Dept Computing Science.

Brumby D, Howes A & Salvucci D (2007a) A cognitive constraint model of dual-task trade-offs in a highly dynamic driving task. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2007. San Jose, CA, USA: 233–242.

Brumby D, Salvucci D, Mankowski W & Howes A (2007b) A cognitive constraint model of the effects of portable music-player use on driver performance. Proc Human Factors and Ergonomics Society 51st Annual Meeting. Baltimore, MD, USA: 1531–1535.

Card S, Moran T & Newell A (1983) The Psychology of Human Computer Interaction. Mahwah, NJ, USA, Lawrence Erlbaum Associates.

Cockton G (2008) Revisiting usability's three key principles. CHI '08 Extended Abstracts on Human Factors in Computing Systems. Florence, Italy: 2473–2484.

Crease M (2001) A toolkit of resource-sensitive, multimodal widgets. PhD thesis. Univ Glasgow, Dept Computing Science.

Crossan A, Murray-Smith R, Brewster S & Musizza B (2008) Instrumented usability analysis for mobile devices. In: Lumsden J (ed) Handbook of Research on User Interface Design and Evaluation for Mobile Technology. Idea Group Inc, Hershey, PA, USA: 927–944.

Das S, Bakis R, Nadas A, Nahamoo D & Picheny M (1993) Influence of background noise and microphone on the performance of the IBM Tangora speech recognition system. Proc IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-93. Minneapolis, MN, USA: 71–74.

Dey A, Abowd G & Salber D (2001) A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. Human-Computer Interaction 16(2): 97–166.

Driver J & Spence C (1998) Crossmodal attention. Current Opinion in Neurobiology 8(2): 245–253.

Duh H, Tan G & Chen V (2006) Usability evaluation for mobile device: a comparison of laboratory and field tests. Proc 8th International Conference on Human-Computer Interaction with Mobile Devices and Services MobileHCI 2006. Helsinki, Finland: 181–186.

Dumas J (2007) The great leap forward: the birth of the usability profession (1988–1993). Journal of Usability Studies 2(2): 54–60.

Dunlop M & Brewster S (2002) The challenge of mobile devices for human-computer interaction. Personal and Ubiquitous Computing 6(4): 235–236.

Edworthy J, Loxley S, Dennis I (1991) Improving auditory warning design: relationship between warning sound parameters and perceived urgency. Human Factors 33(2): 205–231.

Eisenstein J, Vanderdonckt J & Puerta A (2000) Adapting to mobile contexts with user-interface modeling. Proc Third IEEE Workshop on Mobile Computing Systems and Applications: 83–92.

Fallman D (2002) Wear, point, and tilt: designing support for mobile service and maintenance in industrial settings. Proc 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques DIS '02. London, UK: 293–302.

Fitzmaurice G (1993) Situated information spaces and spatially aware palmtop computers. Communications of the ACM 36(7): 39–49.

Flanagan J, Mäntyjärvi J & Himberg J (2002) Unsupervised clustering of symbol strings and context recognition. Proc 2nd IEEE International Conference on Data Mining ICDM 2002. Maebashi City, Japan: 171–178.

Gaver W (1986) Auditory icons: using sound in computer interfaces. Human-Computer Interaction 2(2): 167–177.

Gaver W (1989) The SonicFinder: an interface that uses auditory icons. Human-Computer Interaction 4(1): 67–94.

Gawron V (2008) Human Performance, Workload, and Situational Awareness Measures Handbook, 2nd Edition. Boca Raton, FL, USA, CRC Press.

Goodman J, Brewster S & Gray P (2004) Using field experiments to evaluate mobile guides. Proc 3rd International Workshop on HCI in Mobile Guides (workshop at MobileHCI 2004). Glasgow, Scotland. URI:http://www.comp.lancs.ac.uk/~kc/mguides04/mGuidesHCI-papers.html.

Guillaume A, Rivenez M, Andéol G & Pellieux L (2007) Perception of urgency and spatialization of auditory alarms. Proc 13th International Conference on Auditory Display ICAD 2007. Montréal, Canada: 160–166.

Haas E & Edworthy J (1996) Auditory warnings. Designing urgency into auditory warnings using pitch, speed and loudness. Computer & Control Engineering Journal 7(4): 193–198.

Hackos J & Redish J (1998) User and Task Analysis for Interface Design. New York, NY, John Wiley & Sons.

Han S, Yun M, Kwahk J & Hong S (2001) Usability of consumer electronic products. International Journal of Industrial Ergonomics 28: 143–151.

Hart S & Staveland L (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock P & Meshkati N (eds) Human mental workload. Amsterdam, Elsevier: 139–183.

Hellier E, Edworthy J & Dennis I (1993) Improving auditory warning design: quantifying and predicting the effects of different warning parameters on perceived urgency. Human Factors 35(4): 693–706.

Hoggan E & Brewster S (2007) Designing audio and tactile crossmodal icons for mobile devices. Proc 9th International Conference on Multimodal Interfaces ICMI 2007. Nagoya, Japan: 162–169.

Hoggan E, Kaaresoja T, Laitinen P & Brewster S (2008) Crossmodal congruence: the look, feel and sound of touchscreen widgets. Proc 10th International Conference on Multimodal Interfaces ICMI 2008. Chania, Crete, Greece: 157–164.

Holzblatt K (2005) Customer-centered design for mobile applications. Personal and Ubiquitous Computing 9(4): 227–237.

Holzman T (1999) Computer-human interface solutions for emergency medical care. Interactions 6(3): 13–24.

Hughes J, Randall D & Shapiro D (1992) Faltering from ethnography to design. Proc ACM Conference on Computer-Supported Cooperative Work CSCW 1992. Toronto, Ontario, Canada: 115–122.

ISO/IEC (International Organization for Standardization) (1998) Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on usability. International standard, ISO 9241–11.

ISO/IEC (International Organization for Standardization) (1999) Human centred design processes for interactive systems. International standard, ISO 13407.

ITU (International Telecommunication Union) (2001) ITU-T Recommendation E.161 (02/01) Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network.

Jastrzebowski W (1857) An outline of ergonomics, or the science of work based upon the truths drawn from the science of nature. English version In: (2001) Karwowski W (ed) International Encyclopedia of Ergonomics and Human Factors, Volume 1. Taylor & Francis, London: 129–141. Originally published in Nature and Industry: A weekly devoted to accessible presentation of all branches of natural sciences, their practical application to living, and the latest discoveries and inventions, Poznan, Poland (1857). No. 29, pp. 227—231; No. 30, pp. 236—238; No. 31, pp. 244—247; No. 32, pp. 253—255.

John B & Kieras D (1996) The GOMS family of user interface analysis techniques: comparison and contrast. ACM Transactions on Computer-Human Interaction 3 (4): 320–351.

Kaikkonen A, Kallio T, Kekäläinen A, Kankainen A & Cankar A (2005) Usability testing of mobile applications: a comparison between laboratory and field testing. Journal of Usability studies 1(1): 4–16.

Kernchen R, Mossner K & Tafazolli R (2005) Adaptivity for multimodal user interfaces in mobile situations. Proc Autonomous Decentralized Systems ISADS 2005. Chengdu, China: 469–472.

Kieras D & Meyer D (1997) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction 12(4): 391–438.

Kieras D, Wood S & Meyer D (1997) Predictive engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task. ACM Transactions on Computer-Human Interaction 4(3): 230–275.

Kim H, Kim J, Lee Y, Chae M & Choi Y (2002) An empirical study of the use contexts and usability problems in mobile Internet. Proc 35th Annual Hawaii International Conference on System Sciences HICSS 2002. Big Island, HI, USA: 1767–1776.

Kjeldskov J & Graham C (2003) A review of mobile HCI research methods. Proc 5th International Conference on Human Computer Interaction with Mobile Devices & Services MobileHCI 2003. Udine, Italy: 317–335.

Kjeldskov J & Stage J (2004) New techniques for usability evaluation of mobile systems, International Journal of Human-Computer Studies 60 (5-6): 599–620.

Kjeldskov J, Skov M, Als B & Høegh R (2004) Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. Proc 6th International Conference on Human Computer Interaction with Mobile Devices & Services MobileHCI 2004. Glasgow, UK: 529–535.

Korpipää P (2005) Blackboard-based software framework and tool for mobile device context awareness. PhD thesis. Univ Oulu, Technical faculty, Dept Electrical engineering.

Lee J & Spence C (2008) Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. Proc 22nd British HCI Group Annual Conference on HCI 2008: People and Computers Xxii: Culture, Creativity, interaction - Volume 1. Liverpool, United Kingdom: 185–192.

Lemmelä S (2008) Selecting optimal modalities for multimodal interaction in mobile and pervasive environments. Proc Third International Workshop on Improved Mobile User Experience - IMUx 2008. Sydney, Australia: 208–217.

Lemmelä S, Vetek A, Mäkelä K & Trendafilov D (2008) Designing and evaluating multimodal interaction for mobile contexts. Proc 10th International Conference on Multimodal Interfaces ICMI 2008. Chania, Crete, Greece: 265–272.

Leplâtre G & Brewster S (2000) Designing non-speech sounds to support navigation in mobile phone menus. Proc 6th International Conference on Auditory Display ICAD 2000. Atlanta, USA: 190–199.

Leplâtre G & McGregor I (2004) How to tackle auditory interface aesthetics? Discussion and case study. Proc 10th Iternational Conference on Auditory Display ICAD 2004. Sydney, Australia. URI:http://www.icad.org/Proceedings/2004/LeplaitreMcGregor2004.pdf

Lewis C & Rieman J (1994) Task-Centered User Interface Design: A Practical Introduction. Shareware book available at ftp.cs.colorado.edu/pub/cs/distribs/ clewis/HCI-Design-Book.

Lin M, Goldman R, Price K, Sears A & Jacko J (2007) How do people tap when walking? An empirical investigation of nomadic data entry. International Journal of Human-Computer Studies 65(9): 759–769.

Lumsden J & Brewster S (2003) A paradigm shift: Alternative interaction techniques for use with mobile & wearable devices. Proc 2003 Conference of the Centre For Advanced Studies on Collaborative Research. Toronto, Ontario, Canada: 197–210.

MacKay B, Dearman D, Inkpen K & Watters C (2005) Walk'n'scroll: a comparison of software-based navigation techniques for different levels of mobility. Proc 7th International Conference on Human Computer Interaction with Mobile Devices & Services MobileHCI 2005. Salzburg, Austria: 183–190.

Martin J (1997) TYCOON: theoretical framework and software tools for multimodal interfaces. In: Lee J (ed) Intelligence and Multimodality in Multimedia Interfaces (CD-ROM book). Menlo Park, CA: AAAI Press.

Martin J, Grimard S & Alexandri K (2001) On the annotation of the multimodal behavior and computation of cooperation between modalities. Proc Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents in Fifth International Conference on Autonomous Agents. Montreal, Canada: 1–7.

Martin J, Julia L & Cheyer A (1998) A theoretical framework for multimodal user studies. Proc 2nd International Conference on Cooperative Multimodal Communication, Theory and Applications (CMC 1998). Tilburg, The Netherlands: 104–110.

Maynes-Aminzade D (2005) Edible bits: seamless interfaces between people, data, and food. ACM CHI 2005 Extended Abstracts: 2207–2210.

Microsoft (2006). Q&A: Microsoft unveils details for ultra-mobile personal computers. URI: http://www.microsoft.com/presspass/features/2006/mar06/03-09Mobile.mspx

Millen D (2000) Rapid ethnography: time deepening strategies for HCI field research. Proc 3rd Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques DIS 2000. New York City, New York, USA: 280–286.

Mizobuchi S, Chignell M & Newton D (2005) Mobile text entry: relationship between walking speed and text input task difficulty. Proc 7th International Conference on Human Computer Interaction with Mobile Devices & Services MobileHCI 2005. Salzburg, Austria: 122–128.

Morgan D (1997) Focus Groups as Qualitative Research, 2nd edition. Sage Publications.

Mueller W, Schaefer R & Bleul S (2004) Interactive multimodal user interfaces for mobile devices. Proc 37th Annual Hawaii International Conference on System Sciences. Hawaii, USA: 10–20.

Mustonen, M-S (2008) A review-based conceptual analysis of auditory signs and their design. Proc 14th International Conference on Auditory Display ICAD 2008. Paris, France. URI:http://www.icad.org/Proceedings/2008/Mustonen2008.pdf

Nicol C (2005) Development and exploration of a timbre space representation of audio. PhD thesis. Univ Glasgow, Dept Computing Science.

Nielsen C, Overgaard M, Pedersen M, Stage J & Stenild S (2006) It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. Proc 4th Nordic Conference on Human-Computer Interaction NordiCHI 2006. Oslo, Norway: 272–280.

Nigay L, Coutaz J (1993) A design space for multimodal systems: concurrent processing and data fusion. Proc INTERACT '93 and CHI '93 conference on Human Factors in Computing Systems. Amsterdam, Netherlands: 172–178.

Nuance (2010) Press release (March 30 2010): Nuance announces TALKS 5.0 for touchscreen phones. URI:http://www.nuance.com/news/pressreleases/2010/20100330_talks5.asp.

Nunes L & Recarte M (2002) Cognitive demands of hands-free-phone conversation while driving. Transportation Research Part F: Traffic Psychology and Behaviour 5(2): 133–144.

Obrenovic Z, Abascal J & Starcevic D (2007) Universal accessibility as a multimodal design issue. Communications of the ACM 50(5): 83–88.

OpenStream (2008) Press release (Aug 19 2008): Openstream® announces Cue-me™ multimodal browser for mobile phones. URI:http://www.openstream.com/press_release0808.htm.

Oulasvirta A, Tamminen S, Roto V & Kuorelahti J (2005) Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2005. Portland, Oregon, USA: 919–928.

Oviatt S (1997) Multimodal interactive maps: designing for human performance. Human-Computer Interaction 12(1): 93–129.

Oviatt S (1999) Ten myths of multimodal interaction. Communications of the ACM 42(11): 74–81.

Oviatt S (2000) Multimodal system processing in mobile environments. Proc 13th Annual ACM Symposium on User interface Software and Technology UIST 2000. San Diego, California, USA: 21–30.

Oviatt S, Cohen P, Wu L, Vergo J, Duncan L, Suhm B, Bers J, Holzman T, Winograd T, Landay J, Larson J & Ferro D (2000) Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. Human Computer Interaction, 15(4): 263–322.

Oviatt S, Coulston R & Lunsford R (2004) When do we interact multimodally? Cognitive load and multimodal communication patterns. Proc 6th International Conference on Multimodal Interfaces ICMI 2004. State College, PA, USA: 129–136.

Pascoe J, Ryan N & Morse D (2000) Using while moving: HCI issues in fieldwork environments. ACM Transactions on Computer-Human Interaction 7(3): 417–437.

Paternò F (2005) Model-based tools for pervasive usability. Interacting with Computers 17(3): 291–315.

Patterson R (1982) Guidelines for auditory warning systems on civil aircraft, Civil Aviation Authority Paper 82017. London, Civil Aviation Authority.

Patton M (2002) Qualitative Research & Evaluation Methods, 3rd edition. Thousand Oaks, CA, USA, Sage Publications.

Patton P, Belkacem-Boussaid K & Anastasio J (2002) Multimodality in the superior colliculus: an information theoretic analysis. Cognitive Brain Research 14(1): 10–19.

Pirhonen A (2007) Semantics of sounds and images – can they be paralleled? Proc 13th International Conference on Auditory Display ICAD 2007. Montréal, Canada: 319–325.

Pirhonen A, Brewster S & Holguin C (2002) Gestural and audio metaphors as a means of control for mobile devices. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2002. Minneapolis, MN, USA: 291–298.

Preece J, Rogers Y, Sharp H, Benyon D, Holland S & Carey T (1994) Human-Computer Interaction. Harlow, England: Addison-Wesley.

Price K, Lin M, Feng J, Goldman R, Sears A & Jacko J (2006) Motion does matter: an examination of speech-based text entry on the move. Universal Access in the Information Society 4(3): 246–257.

Raman T & Chen C (2009) Looking beyond the screen: text-to-speech and eyes-free interaction on Android. Google I/O developer conference. San Francisco, USA. URI:http://dl.google.com/io/2009/pres/Th_0115_EyesFreeUserInteraction.pdf.

Ritter F, Baxter G & Jones G (2000) Supporting cognitive models as users. ACM Transactions on Computer-Human Interaction 7(2): 141–173.

Roessler H, Sienel J, Wajda W, Hoffmann J & Kostrzewa M (2001) Multimodal interaction for mobile environments. Proc International Workshop on Information Presentation and Natural Multimodal Dialogue. Verona, Italy. URI:http://i3p-class.itc.it/ipnmd-proceedings.html.

Roto V & Oulasvirta A (2005) Need for non-visual feedback with long response times in mobile HCI. In Special Interest Tracks and Posters of the 14th international Conference on World Wide Web (WWW 2005). Chiba, Japan: 775–781.

de Sá M & Carriço L (2008) Defining scenarios for mobile design and evaluation. In CHI '08 Extended Abstracts on Human Factors in Computing Systems. Florence, Italy: 2847–2852.

Salvucci D (2001) Predicting the effects of in-car interface use on driver performance: an integrated model approach. International Journal of Human Computer Studies 55(1): 85–107.

Schmidt A, Aidoo K, Takaluoma A, Tuomela U, Laerhoven K & Velde W (1999a) Advanced interaction in context. In Proceedings of the 1st international Symposium on Handheld and Ubiquitous Computing (Karlsruhe, Germany, September 27 - 29, 1999). H. Gellersen (ed.) Lecture Notes In Computer Science, vol. 1707. London, Springer-Verlag: 89–101.

Schmidt A, Beigl M & Gellersen H (1999b) There is more context than location. Computers and Graphics Journal 23(6): 893–902.

Sears A, Lin M, Jacko J, Xiao V (2003) When computers fade...pervasive computing and situationally-induced impairments and disabilities. In: Stephanidis C (ed.) Human–computer interaction: theory and practice (Part II). Hillsdale, Lawrence Erlbaum Associates: 1298–1302.

Sears A & Young M (2003) Physical disabilities and computing technologies: an analysis of impairments. In: Jacko JA & Sears A (eds) Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Mahwah, New Jersey, Lawrence Erlbaum Associates: 482–503.

Serrano M, Nigay L, Demumieux R, Descos J & Losquin P (2006) Multimodal interaction on mobile phones: development and evaluation using ACICARE. Proc 8th International Conference on Human-Computer Interaction with Mobile Devices & Services MobileHCI 2006. Helsinki, Finland: 129–136.

Sigman M & Dehaene S (2005) Parsing a cognitive task: a characterization of the mind's bottleneck. PLoS Biol 3(2): e37. doi:10.1371/journal.pbio.0030037.

Sodhi M, Reimer B, Cohen J, Vastenburg E, Kaars R & Kirschenbaum S (2002) On-road driver eye movement tracking using head-mounted devices. Proc 2002 Symposium on Eye Tracking Research & Applications ETRA 2002. New Orleans, Louisiana, USA: 61–68.

Sutcliffe A & Faraday P (1994) Systematic design for task related multimedia interfaces. Information and Software Technology 36(4): 225–234.

Tamminen S, Oulasvirta A, Toiskallio K & Kankainen A (2003) Understanding mobile contexts. Proc International Conference on Human Computer Interaction with Mobile Devices & Services MobileHCI 2003. Udine, Italy: 17–31.

Tarasewich P, Bhimdi T & Dideles M (2004) Increasing the effectiveness of notification cues in mobile environments. Proc 10th Americas Conference on Information Systems AMCIS 2004. New York, NY, USA: 3201–3204.

Thevenin D & Coutaz J (1999) Plasticity of user interfaces: framework and research agenda. Proc 7th IFIP Conference on Human-Computer Interaction INTERACT 1999. Edinburgh, Scotland: 110–117.

Turk M & Robertson G (eds.) (2000) Perceptual user interfaces [Special issue]. Communications of the ACM 43(3).

Tuuri K, Mustonen M-S & Pirhonen A (2007) Same sound – different meanings: a novel scheme for modes of listening. Proc Audio Mostly Conference. Ilmenau, Germany: 27–28.

Wagner D & Schmalstieg D (2003) ARToolKit on the PocketPC platform. Proc IEEE Augmented Reality Toolkit Workshop. Tokyo, Japan: 14–15.

Wexelblat A (1997) Research challenges in gesture: open issues and unsolved problems. Proc International Gesture Workshop. Bielefeld, Germany. Wachsmuth I & Fröhlich M (eds) Lecture Notes In Computer Science, vol. 1371. London, Springer-Verlag: 1–11.

Williamson J, Murray-Smith R & Hughes S (2007) Shoogle: multimodal excitatory interaction on mobile devices. Proc SIGCHI Conference on Human Factors in Computing Systems CHI 2007. San Jose, California, USA: 121–124.

W3C (World Wide Web Consortium) (2010) Multimodal interaction activity. URI:http://www.w3.org/2002/mmi/.

W3C (World Wide Web Consortium) (2002) Multimodal interaction use cases. URI:http://www.w3.org/TR/mmi-use-cases/.

Wobbrock J (2006) The future of mobile device research in HCI. CHI 2006 Workshop Proceedings: What is the Next Generation of Human-Computer Interaction? Montréal, Québec, Canada: 131–134.

Woodson W, Tillman B & Tillman P (1992) Human Factors Design Handbook, 2nd edition. New York, NY, USA, McGraw-Hill.

YapMe (2008) What is Yap®?. URI:http://www.yapme.com.

Original publications

I Ronkainen S, Koskinen E, Liu Y & Korhonen P (2010) Environment analysis as a basis for designing multimodal and multi-device user interfaces. Human-Computer Interaction (in press).

II Ronkainen S, Häkkilä J & Hexel R (2004) Speech-assisted mobile text entry. Proc 10th International Conference on Auditory Display ICAD 2004. Sydney, Australia. URI: http://www.icad.org/Proceedings/2004/RonkainenHakkila2004.pdf.

III Marila J & Ronkainen S (2004) Time-out in user interface: the case of mobile text input. Personal and Ubiquitous Computing 8(2): 110–116.

IV Ronkainen S, Häkkilä J, Kaleva S, Colley A & Linjama J (2007) Tap input as an embedded interaction method for mobile devices. Proc 1st International Conference on Tangible and Embedded Interaction TEI 2007. Baton Rouge, Louisiana, USA: 263–270.

V Ronkainen S, Häkkilä J & Pasanen L (2005) Effect of aesthetics on audio-enhanced graphical buttons. Proc 11th International Conference on Auditory Display ICAD 2005. Limerick, Ireland: 121–126.

VI Häkkilä J & Ronkainen S (2003) Dynamic auditory cues for event importance level. Proc 9th International Conference on Auditory Display ICAD 2003. Boston, MA, USA: 233–237.

VII Ronkainen S (2008) User interface sound design for handhelds – case: internet tablet. In: COST Action IC0601 – Proceedings of the CHI SID Workshop, Florence, Italy: 71–76

Reprinted with permission from Taylor & Francis (I), Springer (III), and ACM (IV).

Original publications are not included in the electronic version of the dissertation.

339. Ojansivu, Ville (2009) Blur invariant pattern recognition and registration in the Fourier domain
340. Suikkanen, Pasi (2009) Development and processing of low carbon bainitic steels
341. García, Verónica (2009) Reclamation of VOCs, n-butanol and dichloromethane, from sodium chloride containing mixtures by pervaporation. Towards efficient use of resources in the chemical industry
342. Boutellier, Jani (2009) Quasi-static scheduling for fine-grained embedded multiprocessing
343. Vallius, Tero (2009) An embedded object approach to embedded system development
344. Chung, Wan-Young (2009) Ubiquitous healthcare system based on a wireless sensor network
345. Väisänen, Tero (2009) Sedimentin kemikalointikäsittely. Tutkimus rehevän ja sisäkuormitteisen järven kunnostusmenetelmän mitoituksesta sekä sen tuloksellisuuden mittaamisesta
346. Mustonen, Tero (2009) Inkjet printing of carbon nanotubes for electronic applications
347. Bennis, Mehdi (2009) Spectrum sharing for future mobile cellular systems
348. Leiviskä, Tiina (2009) Coagulation and size fractionation studies on pulp and paper mill process and wastewater streams
349. Casteleijn, Marinus G. (2009) Towards new enzymes: protein engineering versus Bioinformatic studies
350. Haapola, Jussi (2010) Evaluating medium access control protocols for wireless sensor networks
351. Haverinen, Hanna (2010) Inkjet-printed quantum dot hybrid light-emitting devices—towards display applications
352. Bykov, Alexander (2010) Experimental investigation and numerical simulation of laser light propagation in strongly scattering media with structural and dynamic inhomogeneities
353. Kannala, Juho (2010) Models and methods for geometric computer vision
354. Kela, Lari (2010) Attenuating amplitude of pulsating pressure in a low-pressure hydraulic system by an adaptive Helmholtz resonator

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/

UNIVERSITY OF OULU P.O.B. 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-951-42-6178-7 (Paperback)
ISBN 978-951-42-6179-4 (PDF)
ISSN 0355-3213 (Print)
ISSN 1796-2226 (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO