UNIVERSITY of OULU
OULUN YLIOPISTO

Faculty of Technology

Department of Process and Environmental Engineering

Master's Thesis

Cross-Cultural Usability Testing of Mobile Multimedia: A Case Study with Four Cultures

Oulu 18.5.2010

Author _____
Sara Belt

Supervisor _____
Seppo Väyrynen
Professor

Instructor _____
Jonna Häkkilä
Dr, Research Team Leader

UNIVERSITY OF OULU **Abstract of thesis**

Faculty of technology

Department	Laboratory
Department of Process and Environmental Engineering	
Author	**Supervisor**
Belt, Sara Elina	Väyrynen, S., Professor

Name of the thesis

Cross-Cultural Usability Testing of Mobile Multimedia: A Case Study with Four Cultures

Subject	Level of studies	Date	Number of pages
Safety, Ergonomics and Industrial Maintenance	Master's Thesis	May 2010	101 p., 6 p., 5 appendices

Abstract

Cross-cultural usability has emerged as an important topic within mobile HCI due to the rapid increase in mobile phone penetration across the world. The objective of this research was to chart the impact of culture on usability of mobile multimedia and to evaluate cross-cultural usability practices within mobile industry. The state-of-the-art research in the field of cross-cultural usability was reviewed to outline the theoretical foundation for the interrelationship between culture and usability, to explore frameworks for approaching cultural accommodation in the industry, and to identify the gaps in the body of knowledge from the viewpoint of practitioners in mobile domain. In the empirical part of this research, a think-aloud usability test of a mobile multimedia application was conducted in Singapore, Finland, Canada and India. Cultural differences in the fulfillment of usability targets (task completion rates and SUS scores) and in usability problem detection (the type and frequency of problems) were analyzed.

Regarding the fulfillment of usability targets, compelling differences between cultures were not detected. However, differences in the identification of usability problems were practically significant. The verbal protocol revealed distinct verbalization patterns in different cultures and qualitative observations showed differences in interaction style and approach to testing. The findings were in alignment with, however, not comprehensively predicted by, frameworks of cultural cognition and Hofstede's cultural dimensions applied in HCI.

This research provides evidence on the importance of cross-cultural usability testing and suggests that testing in a single culture does not yield a comprehensive evaluation on the usability of a user interface. Rather than in the level of usability, quantitatively, differences were found in how quality of use is constructed across cultures; different elements on the interface and attributes of usability are emphasized. This research suggests that, in order to yield valid and reliable evaluations on usability, usability testing methodology should be localized to support differences in cognitive and interaction styles. Furthermore, established usability metrics and targets employed in product development should be further developed from cross-cultural usability perspective, as effectiveness, efficiency and error focused metrics are prone to cultural bias in the analyzing phase and may not be comprehensive indicators of usability across cultures.

Library location

University of Oulu, Science Library Tellus

Additional information

Keywords: Usability, Cross-cultural usability, User interface design

OULUN YLIOPISTO **Tiivistelmä opinnäytetyöstä**

Teknillinen tiedekunta

Osasto	Laboratorio
Prosessi- ja Ympäristötekniikan Osasto	
Tekijä	**Työn valvoja**
Belt, Sara Elina	Väyrynen, S., Professori

Työn nimi

Cross-Cultural Usability Testing of Mobile Multimedia: A Case Study with Four Cultures

Oppiaine	Työn laji	Aika	Sivumäärä
Turvallisuus, Ergonomia ja Tehdaspalvelu	Diplomityö	Toukokuu 2010	101 s., 6 s., 5 liitelehteä

Tiivistelmä

Mobiililaitteiden nopean maailmanlaajuisen yleistymisen seurauksena käyttäjäpopulaatioiden kulttuurin ja käytettävyyden suhde on noussut tärkeäksi tutkimuskohteeksi ihmisen ja teknologian välisen vuorovaikutuksen tutkimuksessa. Tämän työn tavoitteena oli selvittää kulttuurin vaikutusta mobiilimultimedian käytettävyyteen ja arvioida mobiililaitteiden tuotekehityksessä sovellettuja monikulttuurisen käytettävyystutkimuksen ja -suunnittelun käytäntöjä. Kirjallisuuskatsauksessa tarkasteltiin vallitsevaa näkemystä kulttuurin ja käytettävyyden suhteen teoreettisesta perustasta ja kartoitettiin tapoja huomioida kulttuuri tuotekehityksessä. Tieteenalan nykytilaa arvioitiin kriittisesti sovellettavuuden näkökulmasta. Työn empiirisessä osassa tutkittiin mobiilisovelluksen käytettävyyttä ääneen ajattelua soveltavalla käytettävyystestillä Singaporessa, Suomessa, Kanadassa ja Intiassa. Kulttuurin vaikutusta arvioitiin vertaamalla käytettävyystavoitteiden (SUS-arvo ja tehtäväkohtainen suoriutumisaste) täyttymistä ja tunnistettuja käytettävyysongelmia.

Merkittäviä kulttuurien välisiä eroja käytettävyystavoitteiden saavuttamisessa ei havaittu. Sen sijaan tunnistettujen käytettävyysongelmien suhteen erot olivat käytännöllisesti merkittäviä. Ääneen ajattelussa, vuorovaikutustavoissa ja tehtävien suorittamiseksi omaksutuissa strategioissa havaittiin eroja kulttuurien välillä. Ihmisen ja teknologian välisen vuorovaikutuksen tutkimuksessa sovellettujen kulttuurisen kognition teorian ja Hofsteden kulttuurimallin todettiin tukevan löydöksiä. Teoriat eivät kuitenkaan ennustaneet kaikkia empiirisiä havaintoja.

Tutkimuksen tulokset puoltavat monikulttuurisen käytettävyystestauksen tärkeyttä ja osoittavat, että kokonaisvaltaista arviota käyttöliittymän käytettävyydestä ei voida esittää yhdessä kulttuurissa tehdyn käytettävyystestin perusteella. Kulttuurien välisiä eroja havaittiin siinä, miten käytettävyys rakentuu eri kulttuureissa, vaikka erot käytettävyyden kokonaistasossa olivat vähäiset. Havaittiin, että käyttöliittymän ominaisuuksien ja käytettävyyden attribuuttien painoarvo vaihtelee kulttuurien välillä. Johtopäätöksenä todettiin, että testausmetodologia tulee sovittaa kulttuurien toisistaan poikkeaviin kognitiiviseen tapaan ja vuorovaikutustapaan, jotta valideja reliaabeleita arvioita käytettävyydestä voidaan tehdä. Lisäksi todettiin, että vakiintuneet käytettävyysmetriikat, jotka korostavat vaikuttavuutta, tehokkuutta sekä virheitä, ja tuotekehityksessä sovelletut käytettävyystavoitteet ovat alttiita kulttuurisidonnaiselle väärintulkinnalle, eivätkä välttämättä mittaa käytettävyyttä kokonaisvaltaisesti eri kulttuureissa. Käytettävyyden mittausta ja käyttäjäkeskeisen suunnittelun tavoitteenasettelua tulee kehittää monikulttuurisen käytettävyyden näkökulmasta.

Säilytyspaikka

Oulun Yliopisto, Tiedekirjasto Tellus

Muita tietoja

Asiasanat: Käytettävyys, Monikulttuurinen käytettävyys, Käyttöliittymäsuunnittelu

Preface

This research was conducted as a part of a project aiming at improving cross-cultural usability research practices within Nokia R&D. The objective of this Master's thesis was to study the importance of cross-cultural usability testing in the case of mobile multimedia, and to support the iteration of usability research methodology, process and integration in product development. The project was started in the Multimedia Experiences unit at the Oulu site in November 2006 and completed in the Services unit during spring 2010. The project was related to my duty as user researcher, and the writing of this Master's thesis was done alongside my work.

I would like to thank Professor Seppo Väyrynen for supervising this thesis, providing me with an extensive perspective on the interactions between humans and machines during my studies, and encouraging and guiding my academic pursuits. I would also like to thank my instructor Jonna Häkkilä for insightful and invaluable guidance on this research, and introducing me to the fascinating world of mobile HCI when I started my work in Nokia.

Sincerest thanks to my past and present colleagues and managers in Nokia, I am privileged to have worked with such talented and inspiring people. Special thanks to Jussi, Jukka, Tatu and Esa; for mentoring, support and encouragement – what fun times we have had on both sides of the glass, traveling the world to understand end users, and spending long hours with post-its.

Most of all, I want to express my greatest gratitude to my family, for always believing in me and providing me with endless support, and to my friends, for everything.

Oulu, May 12th 2010
Sara Belt

Table of Contents

1 Introduction

1.1 Background and Motivation

Nowadays, modern technology crosses national and cultural borders with ease. The western world no longer has an exclusive right to technical innovations. Especially in the field of information and communications technology (ICT), products are targeted at increasingly international and cross-cultural audiences. Despite the globalization of markets, cultural differences remain a reality. In order to satisfy the global user base, there is a demand for making cultural adaptation a part of product design. In addition to the business goals of culture-sensitive design, there is an ongoing discussion within the field of Human-Computer Interaction (HCI) around the social and moral responsibilities and global opportunities of design (Marcus 2002, Rosenzweig 2006, Winschiers & Paterson 2004, Shen et al. 2006). Themes closely linked to cultural adaptation, for instance a universal right to usable products, the preservation of cultural identity and tradition, sustainability, bridging the digital divide and improving global communication to avoid conflicts, are emerging as a topic of discussion within the HCI community.

While spreading geographically, modern technology is also penetrating deeper into everyday life. Information and communications technology is described as pervasive and ubiquitous (Preece et al. 2002). Technical development has enabled increasingly complex functions in ICT products, which has resulted in more versatile and frequent usage encompassing communication, entertainment, commerce, education, health care and several other everyday activities. For instance, mobile handheld devices include features ranging from cameras and GPS to social networking clients and music players. ICT products, mobile phones in particular, are also becoming increasingly important emotionally, *"helping people to feel safer, less lonely, more human"* (Jones & Marsden 2006). Due to the proliferation of devices and diversification of users and their tasks, usability engineering, in general, has undoubtedly become more challenging and more vital than ever.

Two current trends, the internationalization of the user base and the fast-paced development in both technical features and services, affect the mobile communications industry perhaps more than any other domain. By the end of 2009, there were 4.6 billion

mobile cellular subscriptions in the world. Mobile phone penetration is growing even faster than Internet use, and this growth is largely driven by developing countries. (International Telecommunication Union 2010, ix) The usage of mobile handheld devices is expanding to activities traditionally performed by personal computers (PC) and by specialized devices (Wobbrock 2006). Mobile devices are becoming universal appliances, gaining on both desktop systems in computing power and on specialized devices such as digital cameras in picture quality and MP3 players in sound quality. Because of the fast growth of mobile penetration, mobile phones are also becoming the first and primary experience of the Internet for many users in developing countries (Joshi and Avasthi 2007). While becoming powerful everyday appliances, mobile devices continue to have distinctive features that pose challenges to their design and usability. Ever-smaller mobile phones with increasing capabilities must adapt to a variety of contexts and tasks. Mobile technology is thus an increasingly interesting and important research area from the point of view of HCI.

Due to challenges posed by the diversification of users, rapid technical development and ever-growing pervasiveness of technology, the scope of HCI, and mobile HCI in particular, is widening. The increasing awareness of cultural aspects in interaction design and usability engineering is one of the rising areas of interest. Although first publications on designing and the usability testing of international interfaces date back to 1990 (Nielsen 1990a), the field is still relatively novel. The number of publications on the impact of culture on usability in major forums of HCI remains moderate (Kamppuri et al. 2006). Nonetheless, there is a relatively wide agreement within the field of HCI on the importance of cross-cultural usability and design (Smith & Yetim 2004), also referred to as *cultural usability* (Sun 2002a) and *culturability* (Barber & Badre 1998). However, in addition to the fragmentation of underlying theoretical background, cross-cultural usability and user research methodology, and concepts used within the field, there are gaps in understanding how and to what extent culture relates to usability, and a lack of established processes for incorporating culture successfully in product design (Kamppuri et al 2006, Smith & Yetim 2004).

Essentially, the research in the field of cross-cultural HCI derives from three assumptions:

- National and ethnic cultures have an impact on the usability of products.

- Cross-cultural usability can and should be taken into account in different phases of user interface (UI) design and usability engineering.
- Cultural adaptation beyond translation is required in order to create successful products to international audiences (e.g. Smith et al. 2004, Shen et al. 2006, Sun 2004).

However, despite the intuitiveness of the assumptions, persuasiveness of qualitative evidence, and the axiomatic status that culture, as an attribute to usability, has gained in the minds of several HCI researchers, the assumptions remain to be supported only by limited quantitative evidence. There are many unsolved issues regarding the assumptions, and defining a framework for studying and designing cross-cultural usability. Moreover, the existing research is, for the most part, limited to specific application or service domains and specific cultures. (Smith & Yetim 2004, Sun 2002a).

The field of cross-cultural HCI builds upon predominantly qualitative case studies on cultural differences in user and usability attributes: cognitive style (Clemmensen et al. 2009), interpreting representations on a UI (Bourges-Waldegg & Scrivener 1998), performance (Ford & Gelderblom 2003), UI preferences and acceptance (Downey et al. 2005, Evers & Day 1997), and context (Kim et al. 2002), for example. Another prominent topic in the field is cross-cultural usability and user-research methodology (Clemmensen et al. 2009). The existing literature also includes an abundance of specific design guidelines for designing cross-cultural user interfaces (e.g. Fernandes 1995, del Galdo 1990, Gould et al. 2000, Russo & Boor 1993, Mayhew & Bias 2005). Finally, some publications have suggested overarching frameworks for understanding or designing cultural usability (Clemmensen 2009, Sturm et al. 2005, Shen et al. 2006) and developed approaches for integrating cultural usability into product development (Smith et al. 2004, Sun 2004). Thus far, however, none of them can be considered established, particularly from the view point of ICT industry.

A mobile phone consists of dozens of applications which are all localized to dozens of languages and cultures. Mobile phones and mobile applications are designed and developed in a complex and time-to-market driven product-development process. Typically, design and development happens in global organizations. Parallel and sequential projects exist side by side, and are integrated into a complete mobile product late in the product development cycle. (Ketola 2002, 2) Localization is typically carried

out by external agencies late in the product development cycle, while concepting and user interface design, including user interface internationalization, are typically conducted by internal teams (Sun 2002b). Compared to many other technological areas within ICT, particularly to the Web domain, mobile phone development is complex both from the process and technology perspectives. In addition to the characteristic restrictions to design posed by the small size and limited computing power, user interface design must also take into account the restrictions resulting from the development process. User interfaces of mobile applications must follow the platform restraints (input paradigm, screen size and resolution), and UIs across the whole device follow a consistent UI style (defined set of UI components).

From the view point of practitioners in mobile domain, the existing knowledge on cross-cultural usability has several unsolved issues. The field lacks conclusive evidence on the importance of cross-cultural UI accommodation beyond (traditional) localization. The possible variation in significance of cross-cultural usability in different application domains, and in different experience groups, for instance experienced multimedia users and novice users of mobile phones, are rarely discussed. Furthermore, approaches suggested for cultural accommodation of user interfaces pay insufficient attention to compatibility with the product-development cycle. Apart from a few exceptions (e.g. Sun 2002b), the fact that internationalization and localization are typically carried out by different organizations and different phases of product development is not discussed in the existing literature. Feasible approaches and methodology on cross-cultural user research and usability testing are still under development. Most importantly, it remains unsolved which of the current practices of both research and design in mobile application development are effective and successful, and which of them should be further developed or replaced.

This Master's thesis discusses cross-cultural usability testing in the case of mobile applications. Mobile user interfaces and mobile user interface development have unique characteristics that need to be taken into account, even though most principles and methods in HCI are applicable to user interfaces in general. There are only a limited number of (partially inconclusive) results related to cultural usability which support a usability practitioner, even in the field of desktop systems. Evidently, further study is needed to supplement general knowledge and especially to consider the cultural challenges from the viewpoint of mobile applications.

The mobile industry is committed to improving usability practices. Awareness of the need to avoid cultural design exclusion has risen among practitioners. Research evaluating the current cross-cultural usability practices, from which practical implications can be derived, also in short term, is required. In other words, there is a need for research which is naturally tied into the mobile application UI development characteristics and which, to some extent, accepts and considers the limitations always present in industrial settings. To establish the practices in industry and to create a systematic process for cross-cultural usability research, it is important to determine whether there are differences in findings of evaluative usability research conducted according to industry practices. This Master's thesis contributes to the still relatively novel field of cultural usability by adopting an industrial mindset.

1.2 Research Aim

The aim of this Master's thesis research is to study the impact of users' cultural diversity on the usability of mobile multimedia applications. This thesis evaluates the cultural differences in think-aloud usability test findings between four different countries employing methodology that follows industry practices. Finally, the aim of the thesis is to evaluate the current cross-cultural usability testing practice; the impact and necessity of cross-cultural usability testing, and to outline areas for development.

1.3 Research Scope

This study focuses on the evaluative phase of the user interface design and usability engineering lifecycle, involving assessment of the usability of the UI against usability targets in different target markets and identifying the usability problems that need to be addressed before product launch. Informative user research in the UI concepting phase, and consequently the study of cultural differences in users' goals, tasks and overall user requirements, is limited outside the scope of this research. User evaluations in the later phases of UI design lifecycle are crucial due to the nature of mobile product development. Mobile user interface concepts are typically innovative and productization is a relatively lengthy process. Consequently, requirements may change even in the late phases of development. The final outcome of the process, a mobile application user interface, is shaped by changing requirements and unexpected technology constraints

emerging throughout the design and also largely defined by the UI style and integration to other applications and the device. Thus, the role of usability testing, even towards the late phases of product development, is central in mobile application design.

This study focuses on usability testing of globalized mobile multimedia. The research employs only non-translated, non-localized user interfaces. Traditionally localizable elements, such as language, date and time formats and icons, which mainly accommodate the user interface to the superficial level of target culture, are not the primary focus in this research. In line with the mobile UI development process, this Master's thesis is focused on studying (by cross-cultural usability testing) the globalized product of user interface design process, a UI into which considerations of objective culture can be attached in the localization phase. In other words, what is assumed to be the universal layer of UI is examined.

The field of usability research is extensive. Different methods for usability research have been widely studied, and there is an agreement that including the *context of use* is an important consideration in usability research. The consideration of the context of usage adds to the validity of usability findings (e.g. Beyer & Holtzblatt 1998, 36-37), and culture most likely affects the contexts of use (Sun 2004). However, these elements are not in the scope of this research and this research employs laboratory testing. For cross-cultural usability testing, evaluators who do not represent the target cultural are easier to involve in laboratory testing without introducing cultural bias and practical challenges that may result from field testing (Blom et al. 2005) Moreover, formative thinking-aloud usability testing in laboratory settings is an established method within the practitioners, and testing in a real context when evaluating industry practices would be an artificial approach. Furthermore, there are indications that laboratory testing does not impact the identification of usability problems, even though it might impact the frequency and thus the severity of detected problems (Kaikkonen et al. 2005).

2 Literature Review

This chapter outlines the building blocks of studying cultural usability by defining HCI, usability and culture. Characteristics of mobile user interfaces and user interface design are presented to introduce the practical framework related to cultural usability research and design in the industrial context of mobile phone development. Finally, the existing knowledge on cultural usability is presented. The chapter provides a theoretical background for the research, and points out the demand for studying cultural usability and cross-cultural usability testing in an industrial setting.

2.1 Usability

2.1.1 Human-Computer Interaction

The ACM Special Interest Group on Computer-Human Interaction (SIGCHI) defines HCI as:

> *A discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them*. (Hewett et al. 1996, 5)

Although not as established as the ACM definition, Carroll's (2002) view emphasizes the industry goal of creating successful products:

> *HCI is about understanding and creating software and other technology that people will want to use, will be able to use, and will find effective when used*. (Wania et al. 2006)

HCI is an interdisciplinary field building upon the field of human factors and ergonomics, especially cognitive ergonomics, which is broader in scope (Human Factors and Ergonomics Society 2010, International Ergonomics Association 2010). To understand the human side, practice and study of HCI builds upon the theories of psychology (cognitive processes and the empirical analysis of user behavior), sociology and anthropology (interactions between technology, work, and organization). Applied theories include, for instance, cognitive science (Clemmensen et al. 2009), activity

theory (Kuutti 1995), and lately, models of culture (Smith & Yetim 2004). From the computer side, HCI is part of computer science and industrial design. (Hewett et al. 1996, 5)

Human-computer interaction is an umbrella for several partially overlapping design and research philosophies, approaches and processes related to designing interactive technology. Some established approaches and constructs are presented in Table 1.

Table 1. Some approaches to designing interactive technology.

Interaction Design	Method for designing interactive products to support people in their everyday and working lives. (Preece et al. 2002)
User-Centered Design	Philosophy based on the needs and interests of the user, with an emphasis on making products usable and understandable. (Norman 2002)
Usability Engineering	Usability engineering is a process. The usability engineering lifecycle includes the following stages: know the user, competitive analysis, setting usability goals, parallel design, participatory design, coordinated design of the total interface, apply guidelines and heuristic analysis, prototyping, empirical testing, iterative design and collect feedback from field use. (Nielsen 1993)

HCI has a process focus. The approaches summarized in Table 1 all emphasize the process aspect and iterative nature of designing usable interactive products. User research and usability testing that inform and evaluate the user interface design are in the core of process descriptions. HCI encompasses the cycle of studying users, designing for usability and evaluating usability (e.g. Preece et al. 1994, Nielsen 1993). Even though presented as closely related, it has been argued that in practice *evaluation* and *design* are separate communities within HCI, and understanding of how research and design relate is somewhat poor within the field. (Wania et al. 2006, Fallman 2003)

HCI is an innovation and design oriented field, both historically and by definition (Fallman 2003). Consequently, HCI has an integral interrelationship with practitioners and products in the industries developing interactive products. Regardless of the inherent connection between HCI and industry, it has been argued that user-centered design is currently not fully integrated into product development in practice (Venturi & Troost 2004).

The field of HCI has evolved and become increasingly complex due to computerization of everyday life (McNamara & Kirakowski 2005). Up until three decades ago, computer systems were nearly exclusively used in work places by trained operators in specific organizational and environmental contexts. The goals and task the operator had for the product were limited and easy to itemize. Nowadays, technology is embedded into an increasing number of everyday appliances used by the masses. Therefore, the key elements of human-computer interaction; the context of use, characteristics and goals of a user, and characteristics of the computer system, have become more diverse and more challenging to outline. The shift from designing highly specific tools to designing consumer electronics and ubiquitous technology has had a significant impact on HCI as a discipline. Furthermore, from the industry perspective, HCI practice has also become more complex due to changes in product and software development processes. For instance, in the case of mobile phones, design and evaluation have to be adapted to fast-paced concurrent engineering development process within complex organizational settings (Ketola 2002).

2.1.2 User Interfaces

A user interface is a property of a computer system and includes the input and output devices, all the information and feedback presented to the user, information elicited from the user, control sequences by which the user controls the system, and the system's behavior. (Preece et al. 1994, Sarjanoja 2008) Information, feedback and controls may be in different modalities including textual, graphical, auditory and haptic.

In essence, user experience design and the product of this design, a user interface, consist of three overlapping areas: *form*, *content* and *behavior*. The form of a user interface consists of the graphical appearance and physical shape of a product. The content of a user interface includes the information architecture, user interface texts and sounds. The behavior of the system is the core of interaction design and closely relates to content and form of the interface. (Cooper et al. 2007)

A mobile phone user interface includes software and hardware components. The elements of a mobile phone user interface are presented in Figure 1. User interface software platform includes UI software libraries and UI components. Software

platforms are reusable: several mobile phone models with varying applications and functions are built on top of the same user interface software platform. Each software user interface platform defines an *interaction style* and a *presentation style*. The interaction style depicts the interaction paradigm, or information architecture, and includes product-wide conventions for input and output dialogues. The presentation style refers to the 'look and feel', defining for instance the windows, layouts, colors, icons, fonts, sounds. (Kiljander 2004) Consequently, from the viewpoint of mobile applications, user interface design is confined by the style determining the user interface components, navigation paradigm and graphical style.

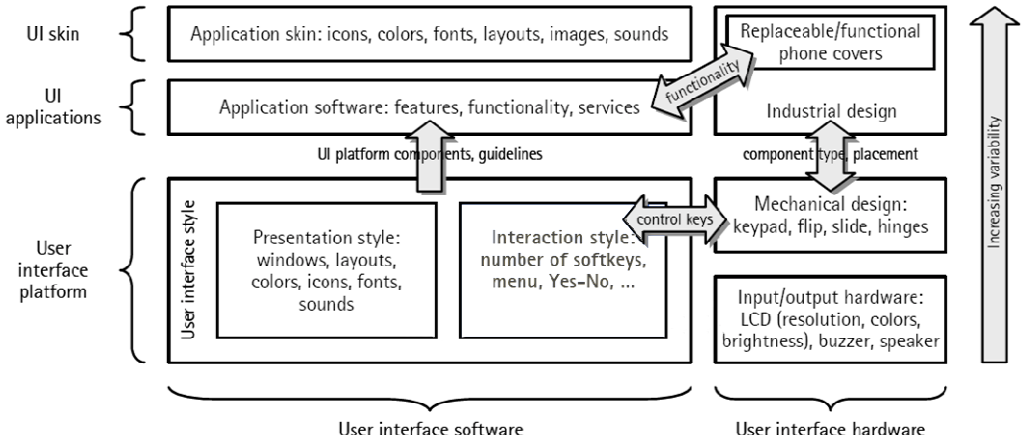


Figure 1. Mobile Phone User Interface Elements (adapted from Kiljander 2004, 80).

Categorization of user interface layers to platform, applications and skin is not the only way to discuss the layered structure of a user interface. Products that are targeted at international audiences include a *globalized* user interface and *localized* variants. Globalization (also referred to as *internationalization*) represents "*the creation of code and a UI design foundation that will work internationally or on which local designs can be implemented*". Localization refers to the "*process of making changes to a globalized product to make it usable and viable in a particular market*". (Fernandes 1995)

From the viewpoint of mobile phones, it is important to note that the user interface is not the only aspect that the user encounters when interacting with a device. According to Ketola & Röykkee (2001), mobile interaction is not limited to the *user interface*, but the usability of mobile devices extends to usability related to *external interface* (enhancements, user support materials, PC connectivity, downloadable software) and

usability related to *service interface* connecting the device to service provider's network. Ketola & Röykkee (2001) emphasize integration and interoperability as important attributes of mobile phone usability.

2.1.3 Defining Usability

There are several perspectives, both theoretical and practical, within HCI on how to conceptualize usability. The ISO standard 9241-11 and Nielsen's definitions are probably the most widely used references. (McNamara & Kirakowski 2005, Jokela et al. 2003)

> *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*. (ISO 9241-11 1998)

> *Usability is about learnability, efficiency, memorability, errors and satisfaction*. (Nielsen 1993)

The terms used in the definitions are explained in Table 2.

Table 2. Terms used in usability definitions.

	ISO 9241-11	**Nielsen (1993)**
Effectiveness	The accuracy and completeness with which users achieve specified goals	-
Efficiency	The resources expended in relation to the accuracy and completeness with which users achieve goals	Once users have learned the design, how quickly can they perform tasks?
Satisfaction	Freedom from discomfort, and positive attitude to the use of the product	How pleasant is it to use the design?
Context of use	Characteristics of the users, tasks and the organizational and physical environments	-
Goal	Intended outcome	-
Task	Activities required to achieve a goal	-
Learnability	-	How easy is it for users to accomplish basic tasks the first time they encounter the design?

Memorability	-	When users return to the design after a period of not using it, how easily can they reestablish proficiency?
Errors	-	How many errors do users make, how severe are these errors, and how easily can they recover from the errors?

Currently, dominant perspective within HCI is perceiving usability as *quality of use* (Hornbæk 2006), which ISO 9241-11 aligns with (McNamara & Kirakowski 2005). Usability is not a static *product property*, but dependent of the user characteristics, goals and context (Bevan et al. 1991). From the perspective of software development, usability is viewed as a software quality attribute (ISO/IEC 9126-1 2001). Nielsen's definition, on the other hand, emphasizes the product perspective; usability is the attributes of the user interface that make it easy to use (Bevan 2008). Both definitions support operationalization of usability as they include measurable dimension that constitute usability. Even though rarely emphasized, both definitions also include the subjective dimension of usability. (McNamara & Kirakowski 2005)

The biggest debate on usability definitions is perhaps related to their scope and to which extent the definitions truly depict the quality of interactive products and predict product acceptance (Bevan 2008, Hassenzahl 2008, McNamara & Kirakowski 2005). Even though Nielsen (1993) introduced the concept of system acceptance, presented in Figure 2, alongside with the definition for usability, the theory is criticized because it does not include utility aspect into usability definition. According to Nielsen, product can be usable even though it has no utility (Bevan 2008).

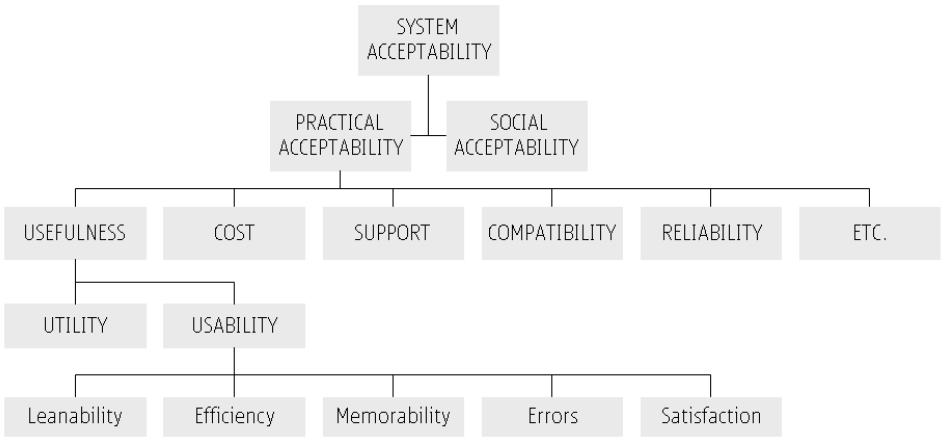


Figure 2. Nielsen's System Acceptability (Nielsen 1993).

Davis (1989) developed a *technology acceptance model* (TAM) according to which perceived usefulness may have a bigger impact on whether product will be used or not than usability or perceived ease of use. Moreover, recently risen interest in *user experience* and the discussion on paradigm shift from *quality of use* to *quality of experience* has contributed to the concerns that established definitions for usability are not comprehensive enough and are limited to perceptual and cognitive processes and the study of performance. Furthermore, critics argue that the established definitions for usability pay too little attention on subjective aspects of technology use and neglect to view technology as a source of "insight, pleasurable stimulation and social exchange". However, criticism only rarely involves replacing existing usability definitions but rather recommends complementing the field of HCI research and practice with structured ways of studying acceptance or experience. (Hassenzahl 2008, McNamara & Kirakowski 2005)

2.1.4 Designing Usability

Usability is the desired outcome of interaction design process (Preece et al. 1994, 14). Several processes and methods have been suggested for designing and evaluating user interfaces (e.g. Nielsen 1993, Cooper et al. 2007). Proposed processes for creating usable systems typically include user research phases and design phases. Research charting end-user characteristics, contexts of use, goals, and tasks of the intended users of the system, as well as actively involving users in the evaluation of the produced design solutions, is presented as requirements for creating usable interfaces (e.g. Cooper et al. 2007, Preece et al. 2002, ISO 13407). The ISO 13407 (1999) standard provides guidance for designing usability and depicts a process of four user-centered design activities. The process is presented in Figure 3.

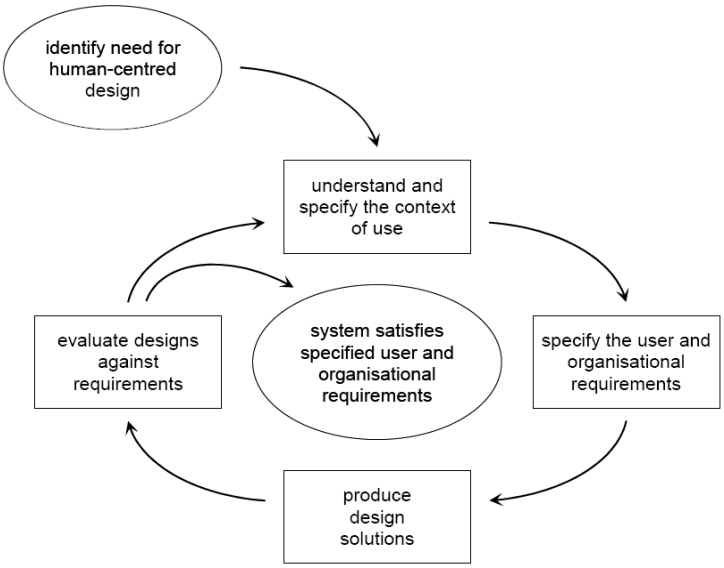


Figure 3. Activities of user-centered design (ISO 13407 1999).

An abundance of user and usability research methods are applied within HCI, and the different methods can be categorized in different ways. Table 3 presents an overview on some ways of categorization and examples on methods that are commonly employed within HCI.

Table 3. Examples on common categorizations of usability and user research methods.

Category	Examples on methods
Qualitative methods	Rapid ethnography (Millen 2000) Contextual inquiry (Beyer & Holtzblatt 1998) Focus groups (Kuniavsky 2003, 201-257)
Quantitative methods	Questionnaires and Surveys (Kuniavsky 2003, 303-366) Usability testing with performance measurement (Nielsen 1993, 192-195) Log analysis (Jones & Marsden 2006, 148)
Field studies	Shadowing (Blom et al. 2005) Contextual inquiry (Beyer & Holtzblatt 1998)
Laboratory studies	Think-aloud usability testing (Nielsen 1993, 195-198) Coaching method in usability testing (Nielsen 1993, 1999)
User studies	Participatory design (Kuniavsky 2003, 468-469) Card sorting (Kuniavsky 2003, 192-199) Self-reporting methods such as probes and diary studies (Hagen et al. 2005)
Non-user methods	Heuristic evaluation (Nielsen & Mack 1994, 5) Competitive research (Kuniavsky 2003, 419)

Background research informing design	Interviews (Preece et al. 1994) Personas and scenarios (Cooper et al. 2007)
Evaluative research	Usability testing (Nielsen 1993) Dramatization (Jones & Marsden 2006, 153)
Prototype testing	Low-fidelity paper prototyping (Ketola 2002) High-fidelity flash prototyping (Sarjanoja 2008)
Testing with product software	Summative evaluations (Preece et al. 1994, 603) Beta testing (Kuniavsky 2003, 393)

From the view point of practitioners, the key factors that impact the choice of research methods include the phase in the design cycle and resources available for the research (Preece et al. 2002, 210-216) *Investigative methods*, for instance contextual inquiry, naturalistic observations and in-situ interviews, can be used to inspire and inform the design and to specify user requirements in early phases of design cycle (Jones & Marsden 2006, 121; Preece et al. 2002, 210-216, Cooper et al. 2007, 58). *Evaluative methods* such as laboratory user testing with prototypes (from low to high fidelity) and heuristic evaluations can be used iterate and evaluate existing designs (Jones & Marsden 2006, 195; Nielsen & Mack 1994, 5-6). Time and cost constraints within the industry have a significant impact on research method selection. For instance, ethnographic research and field evaluations are still relatively rare within the industry due to the fact that they are very time consuming, even though the ability to produce valuable data on context of use, elicit user requirements that users are likely to fail articulating, and the ecological validity issues related to laboratory evaluations, are recognized. (Kaikkonen et al. 2005, Millen 2000, Hagen et al. 2005) Usability research methods employed in this research are introduced in detail in Chapter 3.2.

In addition to the involvement of end-users in the design cycle to understand their needs and to iterate the design solutions, user interface design is supported by several design guidelines. Literature published in the field of HCI includes general interaction design principles, detailed guidelines discussing specific interface platforms or components, as well as case studies (e.g. Cooper et al. 2007, Preece et al. 2002, 20-27). The four categories of design principles outlined by Cooper et al. (2007) are presented in Table 4. Many of the *behavioral principles* and *interface-level principles* are platform-specific. There are several distinct principles, for example, related to designing embedded systems and handheld devices. Moreover, detailed user interface styles, for instance established by device manufacturers, both guide and limit the user interface design

(Häkkilä 2006). Style guides aim at consistency across different parts of the user interface, including different applications within a mobile phone, which is considered an important usability principle (Nielsen & Mack 1994).

Table 4. Design principles (Cooper et al. 2007, 150).

Design values	Describe imperatives for the effective and ethical practice of design. These principles inform and motivate lower-level principles.
Conceptual principles	Help define what a product is and how it fits into the broad context of use required by its users.
Behavioral principles	Describe how a product should behave, in general, and in specific situations.
Interface-level principles	Describe effective strategies for the visual communication of behavior and information.

Mobile interaction design and related user research is a distinct domain within the field of HCI. Specific considerations arise for instance from the ubiquitous context of use, global and heterogeneous user group, limited input and output capabilities, and integration of several interfaces within the mobile interaction system.

The first action in user-centered design process, presented in Figure 3, is to understand and specify the context of use. Hagen et al. (2005) describe the challenge of context of use in the case of mobile phones: *"The very nature of mobile devices means that we carry them with us, and we use them in a number of ways and situations both professionally and personally."* Consequently, the environment of use, which is often limited to a workplace or other stationary location for desktop systems and variety of other interfaces, is ubiquitous in the case of mobile phones and applications.

From the viewpoint of user research, the challenges posed by diverse contexts of use are related to the significant resources required to conduct studies in real contexts, and problems related to privacy considerations and practicalities of fieldwork. For instance, observing the use of mobile phones might require accessing private locations, which participants consider disruptive or uncomfortable. Moreover, as mobile devices are used within personal body space, observing or recording the interactions in real context is very difficult. (Hagen et al. 2005) Blom et al. (2005) have conducted several contextual studies within mobile domain and describe practical limitations researchers have faced during *shadowing* participants: following participants that are using bicycles, being questioned by authorities when recording video inside public buildings, trying to record participants actions in a dark nightclub, following a subway-bound participant in Tokyo

during rush hour and safety of the research team when traveling in a car with participants. From the viewpoint of cross-cultural research, the authors argue that social acceptance related to foreign researchers visiting participants' homes and shadowing them on the street, and on the other hand validity and reliability of data when hiring local researchers to conduct the fieldwork pose challenges.

The design considerations resulting from the diverse context of use are substantial, impacting both the hardware and software. Mobile phones are used in both public and private spaces, while running and when driving a car, in meetings and when exercising, in Brazilian favelas and in Chinese metropolises. According to Schilit et al. (1994)

> *Context encompasses more than just the user's location, because of things of interest are also mobile and changing. Context includes lighting, noise level, network connectivity, communication cost, communication bandwidth and even the social situation.* (from Jones & Marsden 2006, 136)

Even though evidently an important design consideration for mobile phones in general, it has been argued that in the case of some applications, the contextual considerations are over-emphasized. Kim et al. (2002) found that contexts of use for mobile Internet are not widely diverse, but the application use was dominated heavily by few contexts. Furthermore, Kaikkonen et al. (2005) found that the same usability problems were identified both in laboratory and contextual usability research, even though there were differences in the frequency of encountered problems.

According to ISO 13407 standard, context also includes the characteristics of intended users and tasks that users are to perform. Mobile phones are global consumer products and thus, the user groups of different mobile hardware and software platforms are highly heterogeneous in terms of culture, age, skill level and several other attributes (Häkkilä 2006, 31).

From the usability and user research perspective, heterogeneous user group poses challenges related to both collecting user data and external validity of the findings. Cooper et al. (2007) argue that an entire range of user behaviors regarding a product should be captured during investigative research phase, which is a significant challenge, considering that one mobile phone model might have more than 100 million users.

Proposed variables that impact users' goals and behaviors, and thus groups that should be included in representative research samples include for instance:

- Novices and experts (Nielsen 1993)
- Different levels of domain expertise (Cooper et al. 2007)
- Physical characteristics (Preece et al. 2002, 173)
- Lifestyle (Ketola & Röykkee 2001)
- Product buying or technology adoption behavior: late and early adopters (Ketola 2002)
- Demographics such as age, gender, and income (Cooper et al. 2007)
- Culture. (Sun 2006)

Ketola (2002) proposes an adapted model of Nielsen's (1993) *experience dimensions*, a key user group categorization approach, in the case of mobile phones. Ketola's model for mobile domain is presented in Figure 4.

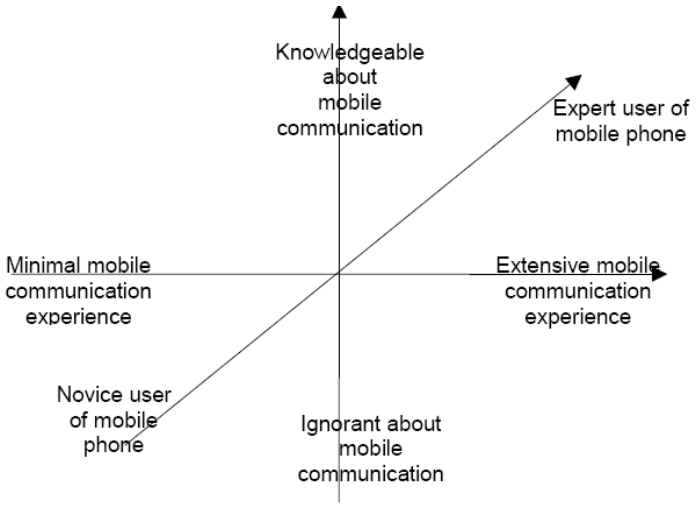


Figure 4. The three experience dimensions for mobile phone use (Ketola 2002, 69).

From the viewpoint of a specific application, service or device, however, it is challenging to evaluate the impact of specific variables on users' goals or behaviors (e.g. Smith et al. 2004).

From a design perspective, heterogeneous user group means balancing between the needs of different user groups of the user interface or creating variants of the interface for different user groups. Mobile interfaces are, for instance, *localized* for different cultures. However, this does not typically involve creating different UI concepts for

different markets but rather adjustment of the user interface to target locales, for instance adapting to different order of presentation (orientation in writing style) and introducing locale specific content (e.g. Chinese music in Chinese variant of music store). There are indications that in some cases designing variants does not meet user needs even though there are special requirements posed by the user group that might justify creating separate versions. For instance, according Ketola & Röykkee (2001) elderly users and users with disabilities do not want to be stigmatized and may want to use the same mobile devices as users with no disabilities. Furthermore, Onibere et al. (2001) found that Batswana users preferred to use standard Windows icons instead of localized ones.

A complex product development environment to which user-centered design must adapt to also introduces important considerations for the mobile user interface design and user research. Innovation and creation of mobile applications and integrated mobile phone products starts with UI concepting which precedes the actual development phase. Concepts are built upon extensive research efforts (including for instance end user research, market research, study of trends and fashion and investigation of technologies), as well as extensive company knowledge and memory. Requirement analysis, or requirement negotiation, follows the concepting phase. In addition to the aforementioned research efforts, requirements are obtained from several sources such as network operators, often being the main customers, competition information, company roadmaps and technical capabilities. Requirements are turned into individual applications and products in design phase (including industrial design, mechanical design, hardware design, software design and UI design) followed by the implementation and integration phases. (Ketola 2002) Software development may work in agile mode, meaning that requirements, design and implementation phases are largely parallel and iterative (Ambler 2010). For mobile phone products, times to market are long, but the product development cycles are fast-paced. User requirements and product requirements evolve and change even during the design phase due to, for instance, unexpected dead-ends and performance problems in implementation. (Ketola 2002)

In the mobile domain, iterative design is limited by the fact that fully integrated and localized variants of the final product are available only in the late phases of user interface development. Thus, evaluative research is often conducted with paper prototypes and of non-localized UIs, which entails some inherent constraints.

Furthermore, collecting longitudinal user feedback (from real end users) and iterating and updating launched products based on the feedback is only just starting to emerge as a paradigm for high-end smartphones. Software updates for devices on the market rarely reach average consumers compared to, for instance, web domain where websites are typically launched as beta versions and frequently updated.

2.1.5 Measuring Usability

The ability to measure usability is central in HCI. The user-centered design process presented in Section 2.1.4 includes evaluating the produced design solutions against the requirements set in an earlier phase. The requirements described by ISO 13407 include *measureable criteria against which the emerging design can be tested*. Furthermore, ISO 9241-11 states that usability is measured in terms of *user performance and satisfaction*. Also Nielsen (1993, 80) accentuates the importance of measureable usability goals as a part of usability engineering process.

The evaluative research phase, that includes usability measurement, typically means *usability testing* of produced designs. Preece et al. (2002, 323-324) describes usability testing from a summative viewpoint: "*usability testing involves measuring the performance of typical users on typical tasks. In addition, satisfaction can be evaluated through questionnaires and interviews*. […] *unlike research experiments, variables are not manipulated and the typical number of participants is too small for much analysis.*"

Usability cannot be measured per se. Operationalization of usability, presented in Section 2.1.3, allows aspects of usability, such as *effectiveness*, *efficiency* and *satisfaction*, to be measured. Research in the field of HCI lacks established approach to measuring usability, and choosing the appropriate measures is problematic for the researchers in the field. The biggest problems are related to the validity and reliability of usability measures: "*whether they actually measure usability, if they cover usability broadly, how they are reasoned about, and if they meet recommendations on how to measure usability*". (Hornbæk 2006)

Hornbæk (2006) analyzed 180 studies published in prominent HCI journals and proceedings. Table 5 presenting *effectiveness*, *efficiency* and *satisfaction* measures applied in HCI research summarizes his findings.

Table 5. Effectiveness, efficiency and satisfaction measures (adapted from Hornbæk 2006).

Effectiveness	
Binary task completion	Number or percentage of tasks that users successfully complete
Accuracy	The accuracy with which users complete tasks, that is some quantification of error
Error rate	Errors made by the user during the process of completing a task or in the solution to the task. Including, e.g., the number of hints given and attempts needed to complete a task
Spatial accuracy	Users' accuracy in pointing to or manipulating user interface objects
Precision	The ratio between correct information and total amount of retrieved information
Recall	Users' ability to recall information from the interface
Completeness	The extent or completeness of users' solutions to tasks
Quality of outcome	Measures of the quality of the outcome of the interaction
Understanding	Understanding or learning of information in the interface, e.g., standardized tests of learning
Experts' assessment	Experts' assessment of outcomes of the interaction, e.g., expert grading on the work products
Users' assessment	Users' assessment of the outcome of interaction
Efficiency	
Time	The duration of tasks or parts of tasks
Task completion time	The time users take to complete a task
Time in mode	The time users spend in a particular mode of interaction, e.g., on part of a task or in part of the interface
Time until event	Time elapsed until users employ a specific feature or perform a particular action
Input rate	Rate of input by the user, e.g., using mouse or keyboard
Mental effort	The users' mental effort when using the interface, e.g., NASA's Task Load Index questionnaire, physiological measures, ratings of mental effort by users
Usage patterns	Measures of how users make use of the interface to solve tasks
Usage frequency	The frequency of function use or actions, e.g., number of key strokes, number of times help consulted
Information accessed	The amount of information users access or employ
Deviation from optimal	The ratio between actual behavior and an optimal method of solution
Communication effort	Resources expended in communication processes
Learning	Users' learning of the interface, e.g., changes in task completion times over sessions

Satisfaction	
Standard questionnaires	Measure satisfaction by using a standardized questionnaire or by building directly on previous work, e.g. QUIS
Preference	Measures satisfaction as the interface users prefer using, e.g., choosing between interfaces, ranking or rating interfaces
Satisfaction with the interface	Users' satisfaction with or attitudes towards the interface, e.g., broad measures of users' overall satisfaction, satisfaction with specific features
Users' attitudes and perceptions	Users' attitudes towards and perceptions of phenomena other than the interface, e.g., attitudes towards the content of the interface, users' perceptions on the interaction.

Based on his analysis, Hornbæk (2006) concludes that more emphasis should be given on measuring the outcome of the interaction and quality of the interaction in order to make comprehensive claims about usability. He found that also learning is infrequently addressed and evaluated in research even though it is emphasized in the existing literature. Furthermore, he argues that distinction between subjective and objective measures of usability is unclear and they are inaccurately mixed together and used synonymously. Finally, Hornbæk criticizes the use of satisfaction questionnaires that are not standardized or validated.

The ability to quantitatively evaluate the produced design solutions against competitor products, other versions of the design and against the requirements set for the product is important. Moreover, empirical data from end-users is considered a persuasive tool to drive design changes in the product development. However, in order to support the goal of improving usability, effective research should provide information on the problems users encountered, why and where the design failed, and on the priority of fixing the problems (Ebling & John 2000). Lewis (2006) makes a distinction between two different usability testing approaches based on their primary focus. Summative usability testing focuses primarily on task-level measurements and formative usability testing focuses on problem discovery. Preece et al. (2002) make the distinction based on the phase in the product development: formative methods are used during design phase, aiming at evaluating the designs against user needs, and summative methods are used with a finished product, assessing the success of a product. According to Carroll (1997), formal experiments are not flexible or rich enough to support iterative design paradigm. Thus, the focus has shifted from summative to formative. Carroll concludes that "*thinking aloud [has become] the central empirical, formative evaluation method in HCI.*" Furthermore, Nielsen (1993, 194) refers to think-aloud usability testing as

perhaps the most valuable usability engineering method. In practice, qualitative and quantitative measures and formative and summative research are often combined. Ebling & John (2000) conducted a usability study where they used quantitative usability attributes and verbal protocol (thinking aloud) side by side. Altogether 65 usability problems were identified, out of which 61 problems were identified with verbal protocol and 40 problems only with verbal protocol. However, they also found that less critical problems were disproportionately represented.

The problem identification approach to usability testing and qualitative data collection has also some demerits. One important aspect is making the distinction and correctly interpreting *subjective* and *objective* measures. Hornbæk (2006) defines subjective measures as "*users' perception of or attitudes towards the interface, the interaction, or the outcome*." and objective measures as "*aspects of the interaction not dependent on users' perception; on the contrary these measures can be obtained, discussed, and validated in ways not possible with subjective measures*." Based on the analysis by Hornbæk (2006), in usability research, subjective measures are sometimes used inappropriately instead of objective measures. For instance, he challenges the validity of measuring learnability by directly asking users.

The e*valuator effect* causes challenges when collecting and analyzing qualitative data, for instance related to think-aloud usability testing and interpreting usability problems participants encounter while completing tasks. Jacobsen et al. (1998) found considerable individual variability between evaluators on identification and rating of usability problems related to same test sessions. Moreover, even though very effective in producing constructive and actionable feedback on the user interface, qualitative methods may fail to predict the severity, frequency or impact, of usability problems (Ebling & John 2000).

In product development, the challenge of usability research that yields qualitative usability problem descriptions is the fact that sufficient information for correcting the problems, from a viewpoint of software projects, is not provided or the data is in a format difficult to interpret or apply (Keenan et al. 1999). Hvannberg & Law (2003) parallel usability problems to software defects. Consequently, they argue that from the viewpoint of software development and quality assurance, defect tracking and analysis is crucial. Several frameworks for classify usability problems as usability problems

have been proposed, for instance the *Usability Problem Taxonomy* (UPT) by Keenan et al. (1999) and *Classification of Usability Problems* (CUP) by Hvannberg & Law (2003). The UPT Framework is presented in Table 6.

Table 6. Usability Problem Taxonomy (Keenan et al. 1999).

Starting points	Primary categories	Sub-categories	
Artifact component	Visualness	Object (screen) layout	
		Object appearance	
		Object movement	
		Presentation of information / results	
		Non-message feedback	
	Language	Naming/labeling	
		Other wording	Feedback messages
			Error messages
			Other system messages
			On-screen text
			User-requested information / results
	Manipulation	Cognitive aspects	Visual cues
			Direct manipulation
		Physical aspects	
Task component	Task-mapping	Interaction	
		Navigation	
		Functionality	
	Task-facilitation	Alternatives	
		Task/function automation	
		User action reversal	
		Keeping the user task on track	

When analyzing usability problem descriptions with UPT, each usability problem detected is placed within one of the mutually exclusive categories presented on the table, starting from primary category and continuing to subcategories. Both starting points, *artifact component* and *task component*, are used for the same usability problem. Keenan et al. (1999) conclude that applying UPT to classify usability problems will enhance the developers' ability to examine the usability data and systematically include that to product development.

2.2 Culture

The concept of *culture* is central for this research as the objective is to chart the effect of culture, as an attribute to users of technology, on the results of a usability study. In social sciences, especially anthropology, there is a long tradition for studying culture

and this section is founded on the understanding existing in that domain. However, in this thesis, the cultural theories are discussed to the extent that they are relevant from the HCI perspective. Thus, an in-depth examination of the concept of culture and cultural theories is outside the scope of this research. The research does not make a study of culture per se, but uses existing definitions of culture to help explain how users around the world might differ from one another in terms of interacting with computers. Moreover, an overview on definitions, models and metamodels is needed to understand the existing knowledge in the field of cross-cultural usability.

2.2.1 Defining Culture

Holden (2002, 21) refers to the concept of culture as *a Babel of definitions*. There is no commonly accepted and established definition for the concept (Evers 2001, 20; Hoft 1996). The Encyclopædia Britannica (2006) defines culture as follows:

> *Behavior peculiar to Homo sapiens, together with material objects used as an integral part of this behavior. Thus, culture includes language, ideas, beliefs, customs, codes, institutions, tools, techniques, works of art, rituals, and ceremonies, among other elements.*

Hall (1989) emphasizes the non-verbal and unstated layers of culture in his characterization:

> *Culture is a man's medium: there is not one aspect of human life that is not touched and altered by culture. This means personality, how people express themselves (including shows of emotion), the way they think, how they move, how problems are solved, how their cities are planned and laid out, how transportation systems function and are organized, as well as how economic and government systems are put together and function.*

According to Hall (1989), even though there are different approaches and definitions for culture, anthropologists agree on some characteristics of culture: that it is learned, different facets of culture are interrelated and that it is shared and defines the boundaries of different groups.

To form an understanding of culture and its importance in computer human interaction, the extensive and general definitions are useful. The generality, however, makes them challenging to utilize in practice in applied disciplines including HCI. Thus, within HCI, researchers often adopt definitions that link culture to the way users interact with computers (Honold 2000). However, significant part of cross-cultural HCI research does not employ or build upon theories or definitions of culture but is confined to naming one definition or even leaving the concept without any definition (Smith & Yetim 2004, Honold 2000, Kamppuri et al. 2006). Some definitions and approaches adopted in the field of cross-cultural HCI are presented in Table 7.

Table 7. Definitions of culture adopted in HCI literature.

Author	Definition of Culture
Hofstede (1991, 5)	Software of the mind…the collective programming of the mind which distinguishes the members of one group or category of people from another.
Honold (2000)	- Culture defines members of a group as distinct from members of other groupings. - Culture creates an orientation system and a field of action for these members. - Culture manifests itself in cultural models. These may be internal cognitions or external artifacts and institutions. - Cultural models may differ in their scope and therefore in their significance to a culture. - Cultural models are acquired through interaction with the environment. Action and experience on the other hand and cultural models on the other affect one another through the process of accommodation and assimilation. - Culture does not determine the behavior of individuals but it does point to probable modes of perception, thought, and action. Culture is therefore both a structure and a process.
Nistbett (2003) (from Clemmensen et al. 2009)	Culture [provides a] means of distinguishing between regional differences in cognitive style, that is, empirically well defined differences in the perception and thinking of people with a background from majority cultures in different regions.

The fact that culture is shared, a construct related to a group of people, implies that in order to study cultural effects, the group needs to be defined. In HCI, cultural effects are typically studied from national or ethnic perspective (Kamppuri et al. 2006) likely due to the fact that nationality is easy to establish. Moreover, there are convincing evidence to support the notion that culture, for instance shared values, varies between nations (Hofstede 1991, Trompenaars 1993, Hall 1989). However, human behavior is not determined by culture at an individual level and there are several other factors that

impact human behavior (Hofstede 1991). The national culture approach is criticized as overly simplistic, because insufficient attention is paid to different ethnic groups within one country (Myers & Tan 2003). Moreover, culture as a construct also applies to subcultures (Sun 2004). Hofstede (1991) identifies different layers of culture which include for instance, religious group, social class and occupation, and organizational culture. In this thesis, cultural effects are studied at a national level.

2.2.2 Cultural Metamodels

Steward and Bennett formed the theory of *objective* and *subjective* culture, based on which culture can be divided into two layers. The inner layer, subjective culture, includes values, assumptions and patterns of thinking. Objective culture is the externalization of subjective culture and includes institutions and artifacts such as economic system, social customs and arts. The externalization can be seen and easily pointed out. Subjective culture is the core of the concept of culture and encompasses the psychological features and, as such, is more difficult to examine (Hoft, 1996).

The *Iceberg model* is based on a perception similar to the key insight in Steward's and Bennet's model, namely the idea that some aspects of culture are visible and some, a notably larger portion, are hidden "under the surface". The Iceberg model is depicted in Figure 5.

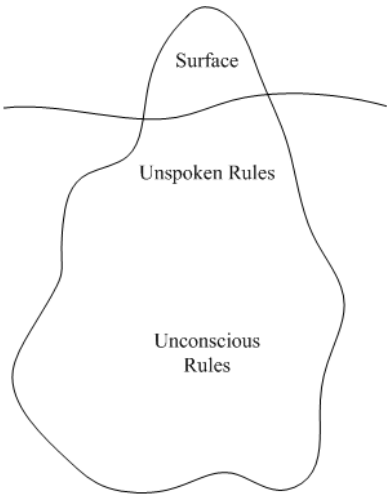


Figure 5. The Iceberg model (Hoft 1996).

The Iceberg model makes the division of culture into three layers. The *surface* layer corresponds to objective culture and includes easily accessible elements of culture, such as currency, time, number and date formats and language. In the Iceberg model, subjective culture is further divided into *unspoken* and *unconscious* rules. Unspoken layer includes rules that often relate to a specific context, such as business etiquette. Unconscious layer includes rules that the group members are not consciously aware of, such as nonverbal communication, and the rate and intensity of speech. (Hoft 1996)

Hofstede's (1991) metamodel of culture is the *Pyramid model* describing the "human mental programming". Rather than discussing the layers of culture, it demonstrates how culture is related to other factors affecting human behavior. Hofstede distinguishes *culture* from two other layers in the human mental programming: *human nature* and *personality*. Human nature is inherited and universal; it is what all humans have in common. Human nature determines the physical and basic psychological functioning. On the other hand, personality is unique to an individual and is partly inherited and partly learned. In the Pyramid model, the layer between human nature and personality is culture. Culture is learned and specific to a group.

2.2.3 Cultural Models

While the metamodels discuss the scope and focus of culture as a construct, cultural models provide a tool for understanding how cultures can be studied and classified based on *cultural dimensions* (Evers 2001). Hofstede (1991) determines cultural dimensions as *"aspect of culture that can be measured relative to other cultures"*.

Hofstede's (1991) cultural dimensions are widely used in HCI research (e.g. Marcus & Gould 2000, Ford & Gelderblom 2003, Smith et al. 2004). Compared to other cultural models, his five-dimensional model is based on probably the most comprehensive sample of cultures and provides an extensive tool for creating hypothesis for comparative cross-cultural studies (Soares et al. 2007). Hofstede's variables are related to subjective culture and focus on explaining cultural patterns of thinking feeling and acting (Marcus & Gould 2000).

The cultural dimensions identified by Hofstede (1991) are:

- *Power distance*

  The extent that people accept large or small distances of power in social hierarchies (Marcus 2001). Power distance reflects the consequences of power inequality and authority relations in society. It influences hierarchy and dependence relationships in the family and organizational contexts. (Soares et al. 2007)

- *Uncertainty avoidance*

  "The extent to which people feel threatened by uncertainty and ambiguity and try to avoid these situations." (Hofstede 1991)

- *Individualism* versus *Collectivism*

  The orientation to individual or group achievements (Marcus 2001). In individualistic societies, individuals look after themselves and their immediate family only whereas in collectivistic cultures, individuals belong to groups that look after them in exchange for loyalty. (Soares et al. 2007)

- *Masculinity* versus *Femininity*

  Dominant values in masculine countries are achievement and success and in feminine countries are caring for others and quality of life (Soares et al. 2007).

- *Long-term orientation* versus *short term orientation*

  "Stands for the fostering of virtues oriented towards future rewards, in particular perseverance and thrift." (Hofstede 2001)

Hofstede's research findings include index scores and rankings for 53 countries regarding each of the dimensions. For example, in terms of uncertainty avoidance, Denmark has the rank 51 (out of 53 countries) indicating weak uncertainty avoidance and Japan has the rank 7 indicating strong uncertainty avoidance. The results indicate that for Japanese, 'time is money' and that the members of the society have 'emotional need for rules, even if these will never work'. In Denmark, on the other hand, 'time is a framework for orientation' and that 'there should be no more rules than is strictly necessary'. (Hofstede 2001, Hofstede 1991)

In addition to Hofstede's dimensions, several other models and cultural variables have been proposed (Hoft 1996). Hall (1989) found cultural variables related *high-context* versus *low context* (whether meaning draws upon context, or if instead, information must be stated explicitly) and *polychromic time* versus *monochromic time* (whether

individuals would prefer to do 'more things at once' or 'one thing at a time') (Evers 2001, 22-23). Trompenaars (1993) explains cultural differences from the perspective of how a group of people solves problems and has created a set of 7 cultural variables. (Hoft 1996)

2.3 Cross-Cultural Usability

This section outlines the existing body of knowledge related to cross-cultural usability within the field of HCI. Different theoretical frameworks and approaches for examining and designing cross-cultural usability are introduced, alongside with the criticism and challenges associated with the current practices related to cross-cultural accommodation of user interfaces. Finally, this section discusses the existing empirical evidence related to the relationship between culture and usability as well as the applicability and validity of usability research methodology in different cultures.

2.3.1 Background of Cross-Cultural Usability

Cross-cultural considerations emerged in HCI community in the middle of the 1990s (Kamppuri et al. 2006). Pioneering books and papers include *Designing User Interfaces for International Use* edited by Nielsen (1990b), *How Fluent is Your Interface*? by Russo and Boor (1993) and *International User Interfaces* edited by del Galdo & Nielsen (1995). The interest in creating UIs that were intended to be used internationally arose from increasing technical sophistication in many countries around the world, end of the Cold War and the consequent larger world trade. (Nielsen 1990a)

Many of the early publications within the field of cross-cultural usability discuss adaptation of user interfaces to *objective* culture: character sets in different languages, translation challenges, date and time formats, currency formats, interpretation of symbols, acceptability of image and cultural associations of color, for example. The focus of the discussion was on creating guidelines and checklists for internationalization of user interfaces and designing and evaluating localized user interface variants. (del Galdo 1990, Fernandes 1995, Russo & Boor 1993). Process and practices were established for creating globalized user interfaces that can either be used internationally

or on top of which localized designs can be implemented. Characteristics of globalized user interfaces include:

- User interface is free of all cultural elements.
- Internationally recognizable, acceptable and usable elements (e.g. iconography) are preferred so that the least possible number of elements needs to be localized.
- User interface is designed and programmed so that it is feasible to localize; all cultural context is extracted (e.g. preferring direct manipulation and gestures over commands, removing text from icons, supporting language directions and different layouts, selection of character encoding schemes). (Taylor 1990, Fernandes 1995)

Nowadays, globalization and localization processes and methods are standardized and the localization industry is an established part of ICT industry (Localization Industry Standards Association 2010). The supply of universal software has contributed to globalization – in the more common meaning of the word – and has obvious compatibility advantages (Shen et al. 2006). However, the approach and scope of prevalent localization practice, building on the early cross-cultural usability work, and the current state of cross-cultural HCI have been strongly criticized. Clemmensen (2009) argues that *"until recently the basic assumption among HCI researchers was that cultural issues could be treated as a practical matter of occasional and peripheral interest […] Most of HCI was regarded implicitly as non-cultural, and something that easily could be transferred across different cultural settings."* Sun (2006) criticizes the narrow scope and surface level view of culture. She argues that concrete cultural realities with their inherent complexities are not taken into account in localization. Sun (2004) draws an analogy between the iceberg metamodel of culture and the layers of user interface currently included in cultural considerations of localization industry: *"enthusiasm for the forms of information products—the tip of the iceberg—usually results in ignorance of the huge underwater iceberg—the broader cultural context where information products are situated, and where products are designed, produced, distributed, and consumed"*.

2.3.2 Approaches to Cross-Cultural Usability

The motivation for cross-cultural usability has become more evident and wider in scope than in the 1990s. Globally available systems such as mobile phones and the Internet have become common, localized ICT technology is used by local-language populations all over the world and countries are becoming increasingly multi-cultural due to the increased mobility of people. Moreover, product development in ICT industry has evolved; outsourcing as well as software development and design in global teams have become common. (Clemmensen et al. 2009)

Cross-cultural usability is still approached from a variety of theoretical backgrounds, but in general, increasing attention is paid on also on subjective level of culture. Cultural models, namely the dimensions proposed by Hofstede, Hall and Trompenaars, and cross-cultural communication approach are widely applied in the field (e.g. Massey et al. 2001, Khaslavsky 1998, Marcus & Gould 2000, Smith et al. 2004). Approaches are also inspired by and built upon cultural psychology (Clemmensen 2009) and activity theory (Honold 2000, Sun 2006).

The core of cross-cultural usability is the notion that user interface, or a computer product, represents the culture of its designer. According to Honold (2000), "*cultural orientation systems manifest themselves in artifacts that reflect the designer's culture*". Clemmensen (2009) argues that external artifacts have a built-in cultural model of technology use which users of the tool have to accept and adapt to in order to use the artifact. He proposes that in order for a system to be usable, user's internal model of technology use – the culturally determined way of doing – and the model of use embedded in the artifact must align. From the viewpoint of cross-cultural communication, analogy can be drawn between human-computer interaction and communication between people from different cultures. Humans apply their language and interaction protocols when interacting with a computer. When using a user interface designed for another culture, user may encounter the same problems or misunderstanding as when communicating with a person from another culture. (Khaslavsky 1998, Sacher et al. 2001)

Ford & Gelderblom (2003) examine the impact of culture on human-computer interaction from the perspective of human performance and related cognitive processes.

They propose that culture impacts each of the four sequential stages from stimulus to response: attention, identification, analysis and response. The suggested impacts of culture on the process, predominantly based on Hofstede's cultural dimensions, are presented in Table 8. Ford & Gelderblom build the framework for studying cultural differences in performance on top of two widely accepted models: stimulus-response model and Hofstede's theory of national culture. Information processing psychology has long been the foundation of HCI (Kuutti 1995) and Hofstede's model is an established theory of understanding cultural differences. However, the information processing psychology approach to HCI has been criticized for reducing the human component of HCI into collection of attributes of cognitive processors instead of active actors (Kuutti 1995). Moreover, there is a lack of explicit demonstration that Hofstede's dimensions, originally not intended for HCI, are transferrable to examination of usability (Smith et al. 2004)

Table 8. Impact of culture on cognitive processes in HCI (adapted from Ford & Gelderblom 2003).

Stage of Process	Description of the Stage	Cultural Impact
Attention	Stimulus, typically aimed at visual or auditory sensors, attract user's attention.	The cultural impact in these stages is dominated by objective culture. Stimuli that are in foreign language, or follow foreign patterns of using color, or employ foreign symbols and metaphors, will not attract attention or user may be unable to identify the signal.
Identification	User identifies the signal using her knowledge base and general context.	
Analysis	If signal is identified, it is classified, stored and analyzed. Decision making, problem solving and reasoning takes place.	Problem solving process and reasoning, consequently also the performance when interacting with a system, are affected by complexity of the problem, concentration and possibly experienced anxiety. Hofstede's cultural dimensions impact the perception of problem complexity, ability to concentrate and level of anxiety in given circumstances.
Response	Analysis triggers a decision to react which is typically done via movement when interacting with a computer system.	Hofstede's cultural dimensions may impact the way in which users will or want to respond to messages.

Ito & Nakakoyi (1996) also approach cross-cultural usability from the viewpoint of cognitive processes by studying the impact of culture in *perception*, *semantic association* and *logical reasoning* phases. According to the model they propose, the

listening mode of interaction – when user perceives and interprets the presentations on a user interface – is dependent on culture. The culture dependence may exist already in the perception phase, but increases towards semantic association and is at its greatest in the reasoning stage. Ito & Nakakoyi argue that in perception phase, user attributes should be primarily considered from the perspective of universal cognitive aspects because, for instance, psychology of color, size and shape are not dependent on language or culture. In the semantic association phase, however, culture has an important role for instance due to the fact that different cultures have different associations for color and symbols have different shapes. The reasoning phase is affected by more complex cultural phenomena such as social norms and values. The framework suggested by Ito & Nakakoyi has been later employed by Onibere et al. (2001). The criticism that relates to the approach suggested by Ford & Gelderblom (2003) presented above is also applicable for Ito & Nakakoyi's model. In addition, the model might include inaccurate assumptions related to the perception stage. Convincing evidence show that perception is in fact culture dependent. For instance, Ahmad (1978) found that perception of size color and shape varies according to native language (from Sukaviriya & Moran 1990), and according to Nisbett (2003) there are differences in perception between Westerners and Easterners, for instance Easterner attend more to environments and are more likely to detect relationships among events than Westerners (Clemmensen et al. 2009).

Bourgess-Waldegg & Scrivener (1998), based cross-cultural usability test findings, induce a theory according to which cultural differences, from the viewpoint of HCI, are representational differences. They argue that "*cultural differences affecting usability and design are mainly representational, and that a culturally determined usability problem can be characterised as the user's difficulty in understanding that representation R means M in context C"* (Bourgess-Waldegg & Scrivener 2000). For instance, a user from both southeastern Asia and from Europe likely recognizes an owl, but in the context of a computer interface, the European user might associate that with system training (in the Greco-Roman tradition, an owl represented wisdom and intelligence) whereas for the Asian, the representation might not make sense in the context (in Southeast Asia, an owl is considered a particularly stupid and brutal bird) (Horton 2005). The framework includes a notion that, even though representational systems are part of users' culture, being able to understand representations of another culture is possible and even common. Clemmensen (2009) extends this notion based on

social psychological studies of multiculturedness and their conclusion that users hold one or more cultural meaning systems as well as theories of perceiving and acting. He argues that a user chooses or implicitly applies the system or theory that is available in that situation and primed by a computer artifact (its cultural model of technology use represented by a product). Cultural Model theory of Usability (CM-U) proposed by Clemmensen is presented in Figure 6.

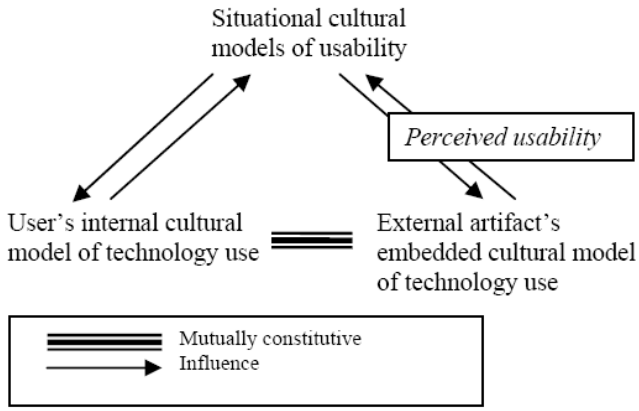


Figure 6. Cultural Model theory of Usability (Clemmensen 2009).

In addition to the frameworks that emphasize the impact of culture on cognitive processes, macro-level approaches, often employing Hofstede's variables, are used within the field to understand and design cross-cultural usability. Barber & Badre (1998) audited 168 websites from different countries and different genres and identified a set of *cultural markers* specific to a culture. For instance, they found that Lebanese websites employ light graphics and emphasize text and that Brazilian websites employ heavy graphics and make references to geographical location often. Based on studies of situated learning and their conclusion that environmental and contextual cues impact learning and memory performance, they argue that the presence or absence of cultural markers on a website may impact learning and performance. Marcus & Gould (2000) extend the examination of cultural differences on websites beyond visual level of a user interface studied by Barber & Badre. They apply Hofstede's dimensions to suggest guidelines for user interface design. Because culture is manifested in the choice of symbols, heroes/heroines, rituals and values, they argue that cultural dimensions can be used to develop design patterns for websites related to metaphors, organization of data, navigation, interaction and appearance. Table 9 presents some implications of dimensions to user interface design suggested by Marcus and Gould.

Table 9. Examples on UI design guidelines based on cultural dimensions (adapted from Marcus & Gould 2000).

High Uncertainty Avoidance	Low Uncertainty Avoidance
Simplicity, with clear metaphors, limited choices, and restricted amount of data.	Complexity with maximal content and choices.
Attempts to reveal or forecast the results or implications of actions before users act.	Acceptance (even encouraging) of wandering and risk, with a stigma on "over-protection".
Navigation schemes intended to prevent users from becoming lost.	Less control of navigation; for example links might open new windows leading away from the original location.
Mental models and help systems that focus on reducing "user errors".	Mental models and help systems might focus on understanding underlying concepts rather than narrow tasks.
Long-Term orientation	**Short-Term Orientation**
Relationships as a source of information and credibility.	Rules as a source of information and credibility.
Patience in achieving results and goals.	Desire for immediate results and achievement of goals.
High Power Distance	**Low Power distance**
Access to information highly structured.	Access to information less-highly structured.
Strong focus on expertise, authority, experts, certifications, official stamps and logos.	Weak focus on expertise, authority, experts, certifications, official stamps and logos.
Individualism	**Collectivism**
Controversial and argumentative speech, tolerance and encouragement of extreme claims.	Official slogans and subdued hyperbole and controversy.
Willingness to provide personal information.	Protection of personal data differentiating the individual from the group.
Prominence given to youth and action.	Prominence given to experienced, wise leaders and states of being.
Masculinity	**Femininity**
Traditional gender/family/age distinctions.	Blurring of gender roles.
Work, task, roles, and mastery, with quick results for limited tasks.	Mutual cooperation, exchange and support (rather than mastery and winning).
Attention gained through games and competitions.	Attention gained through poetry, visual aesthetics, and appeals to unifying values.

Smith et al. (2004) employ the guidelines proposed by Marcus and Gould (2000) to create a process for cultural adaptation of websites. They introduce a concept of *cultural fingerprint* that allows diagrammatical comparisons between cultural profile of the target country (Culture Fingerprint) and cultural profile of the site (Site Fingerprint). Culture fingerprint is developed based on Hofstede's dimension scores for the target

locale. Culture fingerprint can be enhanced by empirically studying the weighting or importance of each variable in the target culture. For instance in their case study, Smith et al. found that for Chinese users, power distance is by far the most significant dimension impacting subjective acceptance of sites and this should be emphasized in the culture fingerprint. Site fingerprint is developed by experts who assess the site against guidelines by Marcus and Gould and give the site a score related to each of the cultural variables. Smith et al. argue that matching the site fingerprint and the culture fingerprint impacts the usability and acceptability of the website.

Joshi & Avasthi (2007) view cross-cultural usability and design from the perspective of technology penetration and other factors external to definition of culture. They argue that when designing for the developing world, biggest cultural differences may not be explained by cultural models or cultural differences in cognition, but are due to the fact that users in developing countries do not have the conceptual model of Internet and are not familiar with the common metaphors used in PC domain. Onibere et al. (2001) also discuss the possibility that experience with computers (the familiarity of Batswana users with computer representations), rather than cultural factors, may have influenced their survey findings related to interface preferences and semantic associations with interface elements. Mrazek & Baldacchini (1997) introduce the concept of *cultural false positives* and argue that many differences in users' tasks and goals are often actually due to user type or user segment rather than geography or culture. Several other authors also emphasize that culture is not the only or primary design consideration. For instance Barber & Badre (1998), alongside with cultural markers, found genre specific markers in their website audits. This indicates that, e.g. banking websites, government websites and travelling websites all have distinct characteristics that are not determined by culture. Furthermore, Smith et al. (2004) discuss the implications globalization of Internet: "*although culture is generally agreed not to change very fast, there are views that, being a global phenomenon, issues related to the Internet may apply globally, thereby transcending local concern*s."

2.3.3 Designing Cross-Cultural Usability

Founded on the early work in the field of cross-cultural usability, there is an abundance of guidelines and checklists for adapting user interfaces to (primarily) objective culture

of different locales. Often, they are not based on theories or models of culture but pragmatically derived from interface audits or existing resources that list superficial differences in culture.

Fernandes (1995) introduces four categories of cultural considerations: language, visual communication, appropriateness of features and taste. Adaptation to different languages and translation poses several challenges to both design and development of user interfaces. For example, different languages have different character sets (and the significant difference in the number of characters in the set) and orientation systems (bidirectional and unidirectional). Different types of characters pose, for instance, inputting and outputting (e.g. resolution required to ensure readability of characters) challenges. Orientation system impacts, for instance, the window and dialog layouts. (Fernandes 1995, Sukaviriya & Moran 1990) Moreover, the translatability of the user interface texts needs to be taken into account. For instance, translating texts from English into other languages can expand the text length significantly, which is problematic especially in the case of mobile user interfaces employing small screens. The use of acronyms and stringing of nouns is not recommended due to difficulties they inflict on translation. Appropriateness of style or tone of the text varies between countries (e.g. addressing the user in the first person might be considered condescending). (del Galdo 1990) Furthermore, domain-specific terminology translations, particularly if metaphors are employed, are problematic because direct translations may not exists (e.g. zooming and panning) or direct translations are not understandable in the context. (Sukaviriya & Moran 1990)

Cross-cultural appropriateness and understandability of visual language is addressed widely in HCI body of knowledge. Horton (2005) introduces an extensive list of graphical design rules which include for instance:

- Make reading direction explicit or unimportant (e.g. if objects must be arranged from left to right, use and arrow to indicate direction).
- Avoid puns and verbal analogies (e.g. avoid using mouse icon, because controlling the screen pointer is not done with a device named after an animal in all countries).
- Avoid mythological and religious symbols (e.g. the Red Cross emblem is not globally recognized).

- Avoid totems (e.g. piggy bank is associated with savings account in many parts of the Western world, but Muslims and Jews consider pigs unclean and unholy).
- Avoid body parts (e.g. a single eye might be associated with examine/inspect but also interpreted as casting a curse).

Sturm et al. (2005) synthesize both objective and subjective levels of culture and propose a model for developing mobile devices for international markets. According to their TLCC model, there are four categories of cultural adaptation: *Technical level*, *Linguistic level*, *Cultural level* and *Cognitive level*. The technical level includes the accommodating the product to technical infrastructure and technical standards used in a target locale. The linguistic level involves translation of interface, manuals and other materials. Sturm et al. argue that cultural aspects related to languages, for instance differences in preference of grammatical structure, are often neglected. Furthermore, they argue that, typically, international adaptation of technical products is limited to technical and linguistic level. The cultural level consists of three areas: adaptation to the context of use, adaptation to the different meanings of technology or the device itself, and adaptation of symbols, graphics, colors and metaphors. Contextual adaptation, according to Sturm et al., is related to integrating the user interface or the product to existing practices and determines the required functionality and necessary elements. Sturm et al. argue that understanding the context of use, ethnographic research is needed. Even though the alternative of using cultural models in this phase is recognized, Sturm et al. consider the cultural dimension too general to fully understand the context of use related to a specific product and a specific country. Finally, the cognitive level determines how to present the functions and elements defined as necessary in the cultural level: menu structures, naming, referential systems, priorities of functions, cognitive styles, interaction styles and basic cognitive processes applied in human-computer interaction. The TLCC model accentuates the importance of end-user involvement in the design process and attaches little importance to general cross-cultural design checklists or guidelines. This emphasis is shared by several other researchers (e.g. Honold 2000, Sun 2002b, Shen et al. 2006), which has resulted in increasing criticism towards the current approach to cross-cultural usability dominated by guidelines and static view of culture.

Current localization approach and guideline driven view to cross-cultural adaptation have been challenged with several arguments. The static view of culture they

incorporate is considered incomplete; cultures are not ontologically objective but rather continuously developing. Cultures interact and there is dynamics between different sub-culture groups. National culture is by far not the only culture that determines a user. Drawing boundaries between cultures is difficult. For instance, should a user be characterized as Indian, Urdu speaking or Muslim? Bi- and multilingualism has become common due to localization and users hold several cultural theories according which to perceive and act. The long-term utility of guidelines is questionable and they may be used to legitimize bad designs. (Sun 2002b, Bourges-Waldegg & Scrivener 1998, Clemmensen 2009) Despite the criticism, proposals for systematic processes for cross-cultural adaptation, especially from the viewpoint of HCI industry, are few and far apart. According to Nielsen (1993, 242), "*an international user interface is a new interface and should in principle be subjected to exactly the same usability engineering process as any other interface*." Also Shen et al. (2006) propose an extensive and iterative involvement of users from representing (each) target locale. Sun (2002b) suggests changing the localization paradigm from exploration of presentation to inclusion of appropriate content in a specific context and involving localization in the design throughout the design lifecycle.

Arguments exist, however, also against the categorical demands of integrating culture deeply into the product development cycle and into user interfaces. Glocalization is defined by Friedman (2000) as follows:

> *The ability of a culture, when it encounters other strong cultures, to absorb influences that naturally fit into and can enrich that culture, to resist those things that are truly alien and to compartmentalise those things that, while different, can nevertheless be enjoyed and celebrated as different*. (from Shen et al. 2006)

Consequently, it may be deduced that the attempt to make all ICT products "local" is not always required, or even desired, by the target user group. Cockton (2009) points out that not all Western influences or cultural markers on a user interface are perceived negative by users outside the Western markets. The ideal combination of local and global is difficult to identify in advance for a specific user interface. Bourges-Waldegg & Scrivener (1998) criticize the hypocrisy related to presenting current practice of cultural adaptation as opposing cultural homogenization or Americanization. They argue that, in fact, undisguised foreignness may be less damaging to a target culture

than carrying Western values into the target cultures disguised as local. From more practical perspective, it has also been questioned whether the interaction paradigms already introduced, for instance in Asia, are so strong that paradigm change at this point would cause difficulty to the users. For instance, Ito & Nakakoyi (1996) argue that Japanese users are so accustomed to the Western-style word processing interfaces that they would find it awkward if the metaphor was changed.

2.3.4 Impact of Culture on Usability

Empirical studies within the field of HCI have demonstrated various cultural effects in different aspects of product acceptance, preferences and usability. However, conclusive quantitative evidence related to the importance of cultural effects on usability, compared to other factors, remain limited. (Smith & Yetim 2004) Characteristic to the research in the field is comparisons between Western users and users from other cultural backgrounds. Research typically focuses on specific technological domains and employs traditional usability and user research methodology, such as interviews, surveys, observation and formal experiments. Surprisingly, in a review of prominent HCI journals and proceedings conducted in 2006, only eight studies were found that reported findings from formal cross-cultural usability experiments. (Kamppuri et al. 2006)

Studies have been conducted which empirically demonstrate cultural differences in cognitive style in the case of user interfaces and, consequently, strategy selected to perform tasks on a user interface. In addition, there are convincing evidence in the field of cultural psychology that cognitive styles differ between Easterners (primarily Korean, Japanese and Chinese) and Westerners (primarily European, American and British), which can be applied in the field of HCI. For instance, differences have been found in the extent to which person's perception of the object is influenced by the environment in which the object has been set (Easterners paying more attention to the environment), and how person deals with seeming contradictions (Westerners following polarization approach and seeking for "one right way"). (Clemmensen 2009)

Already twenty years ago, Sukaviriya & Moran (1990) studied the impact of users' native language on selection tasks, more specifically preferences related to computer

command syntax. Participants were chosen from three cultural/linguistic groups: American, Thai and Indian subcontinent (including Urdu, Hindi, Persian and Telugu speaking users). Thai and English languages employ action-object pattern (e.g. "move chair", "select action") while Indian subcontinent languages employ object-action pattern. The results confirmed the hypothesis that participants representing Indian subcontinent group preferred different command syntax than Thai and English speaking users.

Kim & Lee (2007) examined user preferences related to mobile phone menu structures. The researchers hypothesized that a Western sample of users (Dutch) would follow taxonomy or rules when looking for an option/setting to change ringtone on an UI and that Easter sample (Korean) would find categorization based on relation and contextual information more intuitive. The hypothesis was based on the findings from studies in the field of cultural psychology showing that, generally, Westerners follow analytical cognitive style and Easterners follow holistic cognitive style. The research also involved examining the categorization strategy of each participant in a separate test prior to the actual task with the phone interface, because individual differences in cognitive style within cultural groups exist despite the general cultural differences between groups. The research did not reveal statistically significant differences between the Dutch and the Korean samples. However, differences in how the users approached the menu structure task were found between holistic and analytic categorizers (both categorization styles were represented in both cultural groups). Kim & Lee inferred that the absence of statistically significant difference between the Dutch and the Korean groups was due to their small sample (emphasizing the individual differences) and based their conclusions on the existing knowledge in the field of cultural psychology. They concluded that because users' categorization style impacts the preference of menu structure (holistic categorizers preferred thematic menu structure and analytic categorizers preferred functional similarity in menu structure), user interfaces for East should perhaps employ relational categorization (e.g. everything related to sound grouped together) and contextual options whereas user interfaces for West should employ functional and goal-oriented structures (e.g. all settings or downloads together).

Evidence supporting the existence of cultural differences in cognitive style, which have also been demonstrated to have an effect on strategies adopted when performing tasks on a user interface as described above, are convincing. In addition, cultural differences

have been empirically found, for instance, in how *affordances* (action possibilities that are readily perceivable by an actor) are interpreted (Oshlyansky et al. 2004). Interestingly, however, the body of knowledge related to differences detected in usability test findings, provides inconclusive evidence on influence of culture on usability, as it is currently measured, and little grounds for generalization across different ICT domains and different types of user interfaces.

Several researchers have used Hofstede's dimensions in the study of cultural differences in usability (detected by usability metrics). However, the studies have not conclusively shown that Hofstede's cultural model is eligible for predicting cultural differences in usability. Ford & Gelderblom (2003) hypothesized that performance of participants with different cultural profiles (Hofstede's dimensions) would depend on the cultural profile of websites (assessed based on guidelines proposed by Marcus & Gould (2000)). Namely, that if the cultural profiles of the participant and of the website align, performance, measured in terms of speed, accuracy and satisfaction, is increased. The results indicated, however, that usability of the interfaces was increased for all users, regardless of their cultural profile, with user interfaces employing high uncertainty avoidance, masculinity, collectivism and high power distance. Vöhringer-Kuhnt (2002) also studied the relationship between Hofstede's cultural variables and usability, namely perceived efficiency, effectiveness, satisfaction and attitude towards overall product usability. Statistically significant correlations between cultural variables and components of usability were not found. The correlation between individualism/collectivism and attitude towards overall usability of the system was, however, significant. Downey et al. (2005), on the other hand, found that participants representing high power distance and low uncertainty avoidance made more errors than others and that collectivistic participants showed greater satisfaction with the system than others.

Despite the lack of theoretical underpinning related to relationship between culture and usability metrics, formative usability studies have revealed that usability testing with participants representing different (national) cultures, yields into identifying different problems or the severity of problems may vary between testing locations. Ketola (2002) studied mobile voice mailbox application usability in four European locations (Finland, Italy, Germany and United Kingdom) with low-fidelity prototypes. Only approximately 30% of all usability problems identified were found in all the locations, and nearly 40%

of the usability problems were identified in just one of the locations. The Finnish participants (Finland also being the country of origin of the designers) encountered approximately 40% of all the identified usability problems while the Italian and British participants encountered nearly 70% of all the usability problems. Nielsen (1990b) and Bourges-Waldegg & Scrivener (1998) introduce several examples of culturally rooted usability problems, mostly related to textually conveying meanings in a specific context of the user interface, identified in usability tests. For instance, Nielsen (1990b) discusses the usability problems encountered in Danish version of *MacDraw*. The control/function-keys (for instance R for right-justify and L for left-justify) lose their mnemonic value for Danish users (for whom H for "højre" and V for "venstre" would be more intuitive). Nielsen (1990b) introduces an important consideration, also later addressed by others (e.g. Shen et al. 2006, Massey et al. 2001), related to consistency of interfaces across cultures. In the globalized world, many users are forced to, and thus used to, operating technology in English because many user interfaces the market are not localized. Furthermore, many technical appliances and shared tool in multilingual countries and in global virtual teams are in a single language to enable collaboration and communication. Translating the control/function-keys, in this case, or especially some parts of the underlying UI behavior, might be equally confusing as using foreign representations.

An interesting empirical finding related to cross-cultural usability is the relativism related to usability attributes and determinants of product acceptance. There are indications of cultural differences related to what attributes of usability (effectiveness, efficiency and satisfaction) are considered most important, as well as weight given to aspects like aesthetics and fun (Clemmensen 2009, Vöhringer-Kuhnt 2002, Tractinsky 1997). Moreover, Evers & Day (1997) found cultural differences in user interface preferences (e.g. related to use of color and menu styles) and in the importance of different attributes (e.g. perceived usefulness and ease of use) as predictors of product acceptance.

2.3.5 Cross-Cultural Usability Research Methodology

It is widely assumed within the field of HCI that usability research, from the view point of methodology, is not affected by culture even thought the methods have been

developed from a Western perspective (Clemmensen et al. 2009). However, several studies indicate that the methods commonly used in usability research, particularly usability testing, introduce a cultural bias and that culture impacts the applicability and validity of usability testing methodology. (Clemmensen et al. 2009, Vatrapu & Pérez-Quiñones 2006, Evers 2002). Clemmensen (2009) explains the findings based on his CM-U theory summarized in Figure 6. He argues that situational cultural models of usability included in usability research procedures and situation (e.g. environment in which the testing is conducted, testing materials, the moderator's cultural background) may bias the findings. For instance, usability tests traditionally focuses on a foreground object, the interface, and efficiency of conducting tasks, while it would be natural for Easterners to focus on the surrounding context; the work situation at large, and perhaps pay less attention to the efficiency of the system. A usability test conducted following the established (Western) procedure is intrinsically biased towards finding issues that Westerners would perceive usability problems while finding actual problems that users from other cultures will encounter when using the product in the future may or may not be discovered.

One important consideration related to cross-cultural usability testing is cultural background of the moderator and the overall relationship between a participant and the moderator. It has been shown that more usability problems and suggestions for improvement are found in structured interviews when the moderator and the participant have a shared cultural background, at least when usability evaluation is conducted in a high-power-distance country. (Vatrapu & Pérez-Quiñones 2006) According to Clemmensen (2009), shared knowledge of usability problems, and consequently, interpretation and communication of usability problems, depends on the mutual perception of group belonginess; "*The participants may ask themselves implicit questions about the appropriateness of the available knowledge, such as 'if I tell them about this usability problem, will they understand that this is a problem, or will they think that I am ridiculing them?*" Furthermore, in some cultures, the relationship between a participant and a moderator may be significantly more important than in others. Regardless of the cultural background of the moderator, socio-emotionally focused Easterners may find it important to strive for social harmony resulting in lack of negative feedback and poor correlation between subjective and objective usability measures. Westerners, on the other hand, tend to be task-focused and assign little importance to the interpersonal climate of the testing situation. (Clemmensen et al.

2009, Herman 1996) However, even though Westerners are generally considered more direct and willing to give negative feedback, Evers (2002) found the correlation between performance measures and subjective measures to be relatively poor within the North-American sample in her study. She interpreted the participants to be competitive when answering "quiz-like" questions and trying to give the right answer rather than stating their true opinions.

There are also several other factors that should be taken into account in all phases of cross-cultural usability testing. Guidelines proposed by Clemmensen et al. (2009) are presented in Table 10.

Table 10. Advice for think-aloud usability testing in cross-cultural settings (adapted from Clemmensen et al. 2009).

Explain the background of the test	Easterners want to know the broader context and background of a test; Westerners are less likely to focus on it.
Allow for more pauses when Easterners think aloud	Easterners have more difficulty in thinking aloud. Holistic thought (responding to a much wider array of objects and their relations, and making fewer sharp distinctions among attributes or categories) is less well suited to linguistic representation.
Thinking aloud might adversely affect Easterners' task performance	Thinking aloud might impair the performance of Easterners and enhance the performance of Westerners.
Rely less on expressions of surprise when Easterners are test participants	The extent to which people express surprise differs between cultures. Using surprise as a main marker of usability problems is thus problematic.
Be aware of and mitigate cross-cultural biases in analyzing TA results	The attribution of causes to behavior differs across cultures. Further, the grouping and perception of similarities among behaviors and usability problems may differ depending on the evaluators' cultural background.
Critique of interfaces is likely to seek a compromise and be indirect when users are Easterners	Easterners use conversational indirectness and often attempt to find a middle path.
Use evaluators and users with similar cultural backgrounds, if possible.	Difference in culture may impact the number of identified problems and redesign proposals. Familiarity between evaluator and user may also impact results.
TA tests concern also non-task issues	Easterners are more likely to have a socio-emotional orientation. Thus they may perceive the relationship with the evaluator as being broader than solving tasks or thinking out loud.

2.3.6 Reviewing Cross-Cultural Usability from the Mobile Industry Perspective

The research in the field of cross-cultural HCI drafts a framework that is relatively challenging to apply in the design of mobile applications. The sufficient coverage of markets for each usability and user research phase, the *feasible scope of culture* from the viewpoint of product development, is rarely discussed. For instance, one could easily hypothesize that merely translated user interface for a music player is equally usable for Norway, Denmark, Sweden and Iceland. Moreover, the level of cultural variation in perception, preferences, goals and overall user requirements related to *different application domains* is for the most part not discussed within cross-cultural HCI. For instance, the level of cultural variation in end user needs and usability may differ between communication applications, such as messaging and social networking, and media applications, such as imaging and music. Furthermore, it is largely undefined whether cultural differences are equally significant, from industry perspective, in user groups representing *different dimensions of experience*. The question of whether cultural differences are equally significant for instance within a cross-cultural user group who have been using mobile phones for years and use most functions on mobile phones often compared to cross-cultural user group who do not own a mobile phone or share one with their family.

Cross-cultural research has identified problems in the current practices for cultural accommodation of user interfaces. However, from mobile industry perspective, evidence that prove such profound cultural differences related to perception, user interface preferences and variation in the significance associated with different usability attributes, consequently indicating that universal UI software platforms and globalized and interfaces should be abandoned, is still limited. Furthermore, the conclusions related to effectiveness of cross-cultural usability research in iterative design of mobile user interfaces are somewhat indefinite.

Several areas and specific questions within the cross-cultural HCI, from the perspective of mobile application design, require further exploration and evidence. This research situates in the product development environment within mobile industry, specifically, in the evaluation phase of user interface design cycle of smartphone multimedia. The following chapter will describe the approach for this study and point out which of the

aforementioned gaps in the existing body of knowledge in the practice of cross-cultural usability will be addressed by this research.

3 Study Design

This chapter describes the empirical research conducted to study cross-cultural usability of a multimedia application, as a part of user interface design lifecycle within Nokia. The research problem is defined with the aim of informing future multimedia application design and development projects by evaluating the current cross-cultural usability testing practice. Finally, the methods and procedure employed in the research are presented and discussed.

3.1 Problem Definition

This master's thesis discusses cross-cultural usability testing and, more specifically, addresses the question of whether cross-cultural usability testing has practical significance in the user interface design lifecycle of mobile multimedia applications. Cross-cultural usability testing of a multimedia application is conducted according to industry practices; in selected target markets and with relatively small samples. Furthermore, findings are analyzed and reported from a product development perspective; focusing on evaluating the design against usability targets and identifying usability problems that can be addressed by re-design (usability problems that are identified in all cultures), further globalization, or including more UI elements to the localizable layer of the interface. This research involves usability testing within a specific target group of the multimedia application; expert mobile phone users with extensive mobile communication experience and with existing knowledge about multimedia applications or services. The specific research questions are:

1. In the case of mobile multimedia applications, are there cultural differences in the fulfillment of usability targets obtained by a think-aloud usability test conducted according to industry practices?
2. In the case of mobile multimedia applications, are there cultural differences in the type or frequency of usability problems identified in a think-aloud usability test conducted according to industry practices?
3. If differences can be found, are they significant enough to justify continuous cross-cultural usability and user testing in a product-development environment?

3.2 Methods

3.2.1 Apparatus

The user interface employed in this research was a multimedia application quick launcher providing access to multimedia content (music, photos and videos), games, contacts, Internet bookmarks and landmarks, called the Multimedia Menu (MM). Music player, media gallery and phonebook functionality integrated into the quick launcher were also included in the scope of the usability test. The application was running on Nokia N81 8GB Symbian S60 smartphone depicted in Figure 7. An early, globalized version of the product software was used. Furthermore, the hardware employed was also a product prototype, although closely resembling the hardware of the sales version. The usability of the hardware was not the primary focus of this research, and the study was not designed to yield a comprehensive evaluation on it. However, as software and the hardware are inseparable parts of the user interface, also hardware-related usability problems were included in the analysis where detected. Moreover, when interpreting the usability findings, it needs to be taken into account that the hardware issues may have contributed to the results.

Figure 7. Nokia N81 8GB (Nokia 2010).

There are some demerits related to cross-cultural usability testing of non-localized user interfaces, and localized variants are often employed in the studies within the cross-cultural HCI literature (e.g. Sturm et al. 2005, Vatrapu & Pérez-Quiñones 2006).

However, in order to get the highest possible impact, usability testing of product software must be conducted as early as possible, and localized variants are not available until right before launch. Moreover, the user testing of localization is conducted as a separate research phase in product development. The alternative approach, building a localized prototype, poses some risks to the validity of usability testing. For example, response times of the user interface and their impact on the user experience cannot be evaluated with a prototype. Moreover, even dynamic Flash prototypes do not include all the functions and navigation paths available on the actual interface (Sarjanoja 2008), limiting the possibility to freely navigate in the UI according to any user-preferred strategy, consequently impacting the validity of usability evaluation against targets. Finally, creating localized prototypes for usability testing is highly time-consuming, which impacts the feasibility of the prototyping approach in projects which yield product software, mature enough for research, relatively fast. For instance, Ketola (2002) reported that in their case study of cross-cultural testing of a mobile application, creating and localizing a paper prototype took four to five weeks. Building a dynamic flash prototype would likely take significantly longer.

3.2.2 Sampling

The participants for the study were chosen based on criteria derived from the product development project and from cross-cultural usability literature. The aim was to select a sample that would cover the cultural variation within the product target markets representatively, based on theoretical frameworks and empirical evidence in cross-cultural HCI literature, and also to include other possible sources of variation within the product target user group. The screening criteria included:

1. The participants represent the target experience dimensions of both the smartphone and the application: extensive mobile phone experience (have been using mobile phones for at least five years) and medium experience in multimedia usage (take digital photos and listen to digital music at least monthly, browse online at least weekly). In addition to the aim of aligning the sample to the product target user group, experience with mobile phones and multimedia use was controlled due to the possible bias that different experience profiles may inflict the usability test findings between cultures.

2. The sample consists of an even split of Nokia and non-Nokia mobile phone users in each of the testing locations. In addition to the aim of aligning the sample with the product target user group (current customers and potential customers), the phone model in use was controlled due to the possible bias that familiarity with different UI styles may inflict the usability test findings between cultures.
3. Sample includes an even gender split.
4. The participants speak English fluently. English fluency was required due to the use of non-localized UI.
5. The participants represent key target markets of both the smartphone and the application.
6. Both Easterners and Westerners are included in the sample. Differences in perception between the two groups have been studied and show within the cross-cultural usability literature (e.g. Clemmensen et al. 2009, Kim & Lee 2007) and, thus, both groups should be represented in the sample.
7. The selected countries include a representative spectrum of Hofstede's dimensions. Hofstede's cultural model is widely used to study and explain differences in cross-cultural usability (e.g. Smith et al. 2004, Marcus & Gould 2000) and, thus, an extensive representation of different value orientations should be included in the sample.

Screening criteria 1, 2, 3 and 4 were used to recruit the individual participants within the selected countries. The countries were selected based on criteria 4, 5, 6 and 7 and included Finland, Singapore (participants with Chinese ethnic background), Canada and India. Hofstede's index scores for the selected countries are presented in Figure 8. The maximum and minimum scores for each dimension revealed in Hofstede's research of 53 countries are included in the figure to anchor the scores.

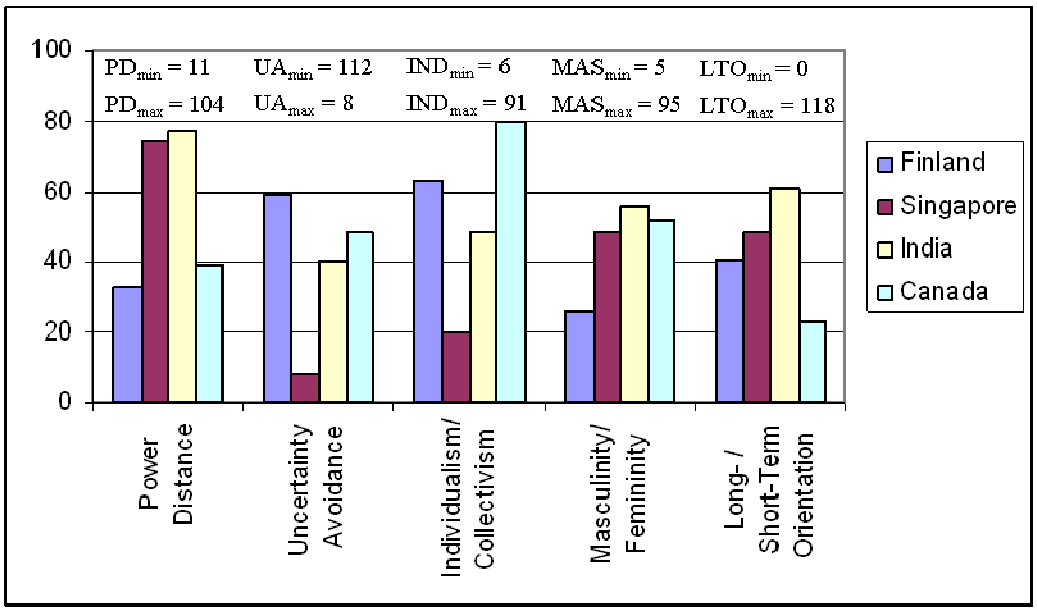


Figure 8. Index scores for Hofstede's dimensions in countries included in the sample (adapted from Hofstede 2001).

The countries selected for the research represent the different levels of power distance, individualism vs. collectivism and long-term vs. short-term orientation relatively well. With respect to uncertainty avoidance, the sample is somewhat biased towards low uncertainty avoidance, and with respect to masculinity vs. femininity, the sample is somewhat biased towards femininity. Full list of index scores and country ranks is presented in Appendix 1, and a description of the dimension profile of the sample is presented in Appendix 2.

Eight users were recruited in each of the countries. Sample size can be considered relatively small from reliability perspective, however, typical for industry usability testing (Nielsen 1993).

3.2.3 Usability Test

Formative think-aloud usability testing was selected to study the research problem. The method involves individual participants performing predefined tasks with the system that is being studied and thinking aloud while interacting with the system (Nielsen 1993, Clemmensen et al. 2009). A description of different elements of think-aloud usability test is presented in Table 11.

Table 11. Elements of think-aloud usability test (adapted from Nielsen (1993) and Clemmensen et al. (2009)).

Welcoming and purpose of the test	Participant is explained the purpose and course of the test, methods used to record user actions and confidentiality of the results. Participant is made comfortable by explaining that they are not being tested and they are given the chance to ask questions before the test begins. The moderator tries to establish a productive relationship with the participant and ensure that participant feels free to make both positive and negative comments.
Instructions and tasks	Participant interacts with the system based on a set of instructions including tasks (selected so that they are representative of the tasks that users will perform with the system after it is launched) and explanation on how to think out loud.
Verbalization	While solving the tasks, the user verbalizes her thoughts. If the user falls silent for longer periods of time, moderator uses prompts to resume verbalization.
Reading the user	One or more evaluators observe the user's behavior and verbalizations. Based on the observations, evaluator(s) extract, describe and report usability problems.

Recommendations for conducting cross-cultural think-aloud usability testing, presented in Section 2.3.5, were used as a basis to design the test. A local moderator sharing a cultural background with the participants was used in each of the countries. The testing was conducted in English in Singapore, Canada and India. For the Canadian users, English is the native language. Singaporean and Indian samples consisted of users who used spoken and written English daily. Both the user interface and the tasks were presented in English also for the Finnish sample, but as the Finnish participants did not use spoken English daily, they were given the option to think aloud in Finnish if preferred.

The moderators were given instructions to explain the background and the context for the testing according to procedure suitable for the culture to reduce the risk of introducing cultural bias. The same set of tasks, presented in Appendix 3, was used in all the countries. The tasks were constructed to include a scenario to support the Easterners paying more attention to the larger context of the task. The guiding of verbalization was carried out according to local practices, allowing users to take pauses if needed.

The test sessions were conducted in usability laboratories. Even though usability testing in real context of use is often recommended (e.g. Oulasvirta et al. 2005), laboratory testing was chosen due to advantages. Observation by the author in a separate room was easier to arrange in laboratory settings eliminating the interference with the relationship between moderator and the participant. Moreover, recording of both the user interface (video camera attached to the device) and the participant was easier to arrange in laboratory settings to ensure the availability of material for detailed analysis. Laboratory testing was considered to yield valid usability findings based evidence which indicate that testing in a laboratory does not significantly impact the usability problems found, even though it might impact the frequency of encountering problems to some extent (Kaikkonen et al. 2005). Finally, from the view point of industry, cost efficiency and ability to produce results fast is important, favoring the selection of laboratory testing. The aim of this research is to evaluate existing cross-cultural usability practices, and the laboratory testing method selected may be considered a valid reproduction of industrial usability testing.

3.2.4 Data Preparation and Analysis

All the usability tests were recorded and the recordings were transcribed after the study by the author (single evaluator). Task completion rates and subjective satisfaction metric were employed to evaluate the usability of the multimedia application against the usability targets set in the product development team. The scheme for analyzing and categorizing task completion is presented in Table 12.

Table 12. Analysis scheme for task completion.

Main category	Sub-category
Completed	Fully completed (with 1st or 2nd try)
	Completed, but with observed or expressed difficulty (3 or more tries or expressed difficulty)
Failed	Participant needed a hint from the moderator
	Participant gave up
	Participant had the misconception that the task was complete even though it was not

Binary task completion is a widely used usability metric (Hornbæk 2006). Unlike traditional performance measures, which typically do not yield reliable results when combined to formative think-aloud usability testing (Nielsen 1993), task completion can be combined to the method (Nielsen 2001). The analysis scheme used in this study was constructed based on Nielsen's (2001) *success rate* and usability test coding scheme suggested by Tullis & Albert (2010). The success rate measure proposed by Nielsen (2001) includes a category for partially completed tasks, but the category was not included in this research due to the relatively small number of tasks in the study that consisted of clearly distinguishable sub-tasks. Task completion can be studied either from the view point of individual tasks (percentage of users completing a specific task) or from the viewpoint of the overall success rate with the system studied (percentage of tasks that users complete successfully), as suggested by Nielsen (2001). From the industry perspective, both rates are useful. When both primary (frequent and basic) and secondary (rare and more advanced) tasks are included in the task set, emphasis is on the individual task completion rates, because the goal for primary tasks is higher than for secondary tasks. In this study, the task set included four secondary tasks in the set of 14 tasks. Thus, even though overall success rates were analyzed for cultural comparison purpose, the evaluation against usability targets was conducted on a task level. The target task completion rates for primary and secondary tasks are presented in Table 13.

Table 13. Target task completion rates.

Task category	Target
Primary task	75% of users fully complete a primary task
Secondary task	75% of users complete a secondary task (either fully complete or complete but with observed or expressed difficulty)

The employed task completion rates allow the identification of tasks that do not yet meet the usability goals and the comparative evaluation of the overall usability of the system (from the task completion perspective). However, due to the relatively small sample in a single country, the findings related to fulfillment of goals in each of the countries are only indicative.

To measure the participants' subjective satisfaction with the system, a SUS score was used. *System Usability Scale* (SUS) developed by Brooke (1996) is a popular tool used for subjective assessment of overall usability of a system (Lewis & Sauro 2009). The

questionnaire was developed from the viewpoint of industry to yield comparative data of system usability with a simple post-test questionnaire. The SUS includes ten statements:

1. I think I would use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I would imagine that most people would learn to use this system very quickly
8. I found the system very cumbersome to use
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going with this system

The SUS is a bipolar, five-point, partly labeled Likert scale. A respondent indicates the degree of agreement or disagreement on a five-point scale from 1 (anchored with "strongly disagree") to 5 (anchored with "strongly agree"). Points from 2-4 are not labeled. As a result, the SUS yields a single score, from 0 to 100 representing the (subjective) overall usability. (Brooke 1996) Even though Brooke does not refer to the SUS as a satisfaction score, several HCI researchers have used it to measure the satisfaction aspect of usability (e.g. Everett et al. 2006, Bangor et al. 2008) The reliability and validity of the SUS have been investigated and confirmed in several studies (Lewis & Sauro 2009).

In case of the multimedia application studied, the goal for the (mean) SUS score was set to 75% in the product development team. The fact that four secondary tasks were included in the task set may decrease the SUS scores to some extent as SUS is intended for satisfaction measurement related to basic tasks. Moreover, as with task completion rates, the SUS scores only yield indicative findings due to the small sample.

To support the usability measurements, the number of comments indicative of different aspects of usability was analyzed from the verbal protocol. Following categorization, adapted from Tullis & Albert (2010), was used for the analysis:

- Stated confusion

- Stated frustration
- Other negative comment
- Stated variation from expectation
- Suggested improvement
- Question
- Positive comment

The attributes of usability – effectiveness, efficiency and satisfaction – were all addressed by the employed measurements. Task completion is indicative of effectiveness of the user interface (Hornbæk 2006). To some extent, the efficiency aspect was also built into the task completion measurement as the task set included some repetition of same functions indicative of learnability. Furthermore, efficiency-related data was collected from the verbal protocol as stated frustration is indicative of perceived (subjective) lack of efficiency. However, besides the verbal protocol, which is very sensitive to cultural bias as presented in Section 3.2.5, efficiency was not directly measured resulting in some incompleteness in the degree to which the findings represent usability broadly. Satisfaction was measured with SUS scores and with the number of positive comments on the interface.

In addition to the aforementioned usability metrics, all encountered usability problems were extracted during the analysis. Even though qualitatively identifying and interpreting usability problems is inherently relatively inexact and prone to cultural bias and evaluator effect, it is considered one of the most important and effective ways to improve usability (Nielsen 1993). Guidelines used to recognize usability problems in the data include:

- Verbal protocol: user states confusion, frustration, variation from expectation, or gives other negative comments; user describes the system behavior or representations in an incorrect way.
- Behavior: user navigates in the interface and tries functions without advancing or accomplishing the task at hand, user completes a task in an incorrect way (and states that the task is completed), user does not complete a task.

Even though recognized as a valuable tool to support systematic usability problem detection and correction in software projects, usability problem categorizations presented in Section 2.1.5 were not employed because, in this case study, the user

interface design team and product management, not the software development or quality assurance teams, were the primary audiences for the usability test findings. The design team was considered to benefit the most from detailed usability problem descriptions pointing out the de facto problems encountered. For the product management, evaluation of the system usability against the targets was considered the most informative outcome of the study.

4 Results

4.1 Overview on Observations

The results related to verbal protocol, usability metrics and detected usability problems are presented in the following sections. In this section, some general observations related to the fieldwork are presented to provide an overview on the characteristics of usability testing sessions in the four different countries included in the sample.

The Singaporeans had a clearly distinct approach to performing the tasks compared to the other three countries in the sample. Even though they were given a lengthy introduction to the background and purpose of the testing, as well as instructed on how to think aloud, they tended to verbalize less than participants from the other three countries. When given a task, Singaporeans typically browsed around the user interface, appearing to be exploring the different options provided by the system before getting started with the tasks at hand. Compared to the other groups, the Singaporean participants were significantly faster in their navigation of the user interface, clicking around and trying out different functionality seemingly unrelated to the task. Thus, even though navigating through clearly more views than the other groups, Singaporeans tended to perform with the tasks relatively quickly and without any apparent indication of frustration due to the navigation. What is more, the Singaporean group also seemed relatively quick to come up with alternative solutions when their primary strategy failed, without expressing confusion or frustration. Compared to Finnish and Canadian participants, who often seemed to be interested in determining the underlying logic of functions and the purpose of different views, Singaporeans often used the "quick exit", provided by the End key taking them back to home screen, when they encountered something that they did not find helpful or perhaps had difficulty interpreting.

The Canadian participants evidently took the most informal and conversational approach to the testing. Sometimes, rather than focusing on performing the tasks, Canadians seemed to be focused on evaluating the path to complete a task. Several participants in the Canadian sample appeared to be aiming at maximizing the amount of constructive feedback. Thinking aloud while performing the tasks seemed relatively natural for them. They formed a strategy for each task which they were able to articulate,

and difficulties encountered or the change of primary strategy was easy to determine from their verbalizations.

The Finnish participants approached the testing relatively similarly to the Canadian participants even though they seemed to be slightly less confident in blaming the system when they had difficulties with the tasks. They preferred statements describing their own confusion, such as "*I don't know which way to rotate this*." (female, Finland) over statements that would indicate a fault in the system, such as "*This is not good, I'm trying to go back and it makes me choose*." (male, Canada). Similarly to the Canadians, the Finns were apparently trying to establish a model of the logic with which the system is working. Even after completing a task, several of the Finnish participants continued experimenting with some functions, appearing to be making sure that they understood the logic behind the function.

Excluding the fact that the Indians made only little attempts to explore the underlying logic of the system, the Indian participants approached the usability testing relatively similarly to the Canadian and the Finnish participants. They were straightforward with their feedback on the system and expressed their emotions relatively freely.

4.2 Verbal Protocol

Results from the verbal protocol characterize the differences between usability sessions in different countries well. The number of verbalizations in each of the categories for all the countries is presented in Figure 9. Even though some individual differences in think-aloud activeness were detected in all the countries, the total number of comments reflects the thinking aloud patterns in each of the countries.

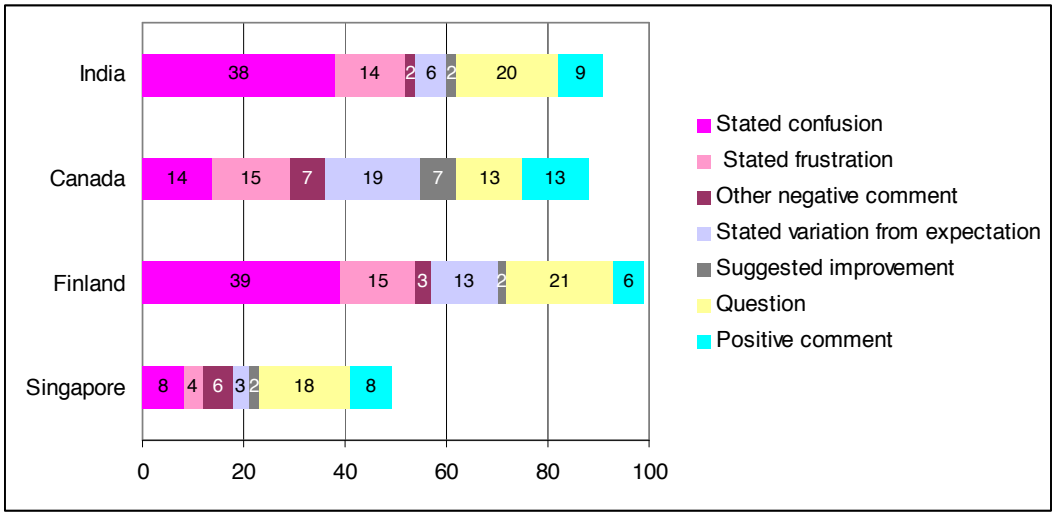


Figure 9. Number of verbalizations in different categories.

The total number of usability-related comments in the categories analyzed was close to equal in Finland, India and Canada, the number of verbalizations totaling 99, 91 and 88 respectively. Verbalizations in Singaporean sample were notably fewer, a total of only 49 comments were made by the Singaporeans. The results support the qualitative observation that Singaporeans tended to be relatively silent when completing the tasks and gave few evaluative comments after they had completed a tasks.

Most of the comments made by the Singaporeans during the tasks were questions seeking assurance from the moderator, such as "*is this correct?*" (female, Singapore), or non-specific expressions of confusion, such as "*I don't know*" (female, Singapore) or "*I'm not sure*" (male, Singapore). Compared to the other countries, the Singaporeans also made (percentually) many general evaluative comments not directly related to the problem solving task at hand, such as "*these keys are a bit sensitive*" (female, Singapore) or "*I like this multimedia menu place, it is so accessible*" (male, Singapore). Questions and general comments (both positive and negative) represent approximately 65% of all the comments made by Singaporeans and are emphasized compared to other countries where the same categories represent significantly less than half the comments. In the other countries, comments directly related to the task or the user interface element at hand, comments expressing emotions (confusion or frustration), and comments evaluating the user interface against expectations tended to dominate the verbalization.

For India and Finland, the thinking aloud patterns were strikingly similar. Stated confusion was clearly the most prominent category of verbalizations followed by

questions and expressions of frustration. The only category in terms of which the countries differed is expressing variation from expectation which was somewhat more frequent among the Finnish participants. Moreover, based on the observations during the sessions, the Finnish and Indian participants tended to verbalize somewhat similarly even thought they were moderated in a slightly different manner. The Finnish moderator was supportive when encouraging thinking aloud, using phrases like "*You are doing well with thinking aloud*." and "*What is on your mind right now?*" while the Indian moderator was more commanding, using phrases like "*Keep talking*.". In both countries, participants made a lot of comments related to specific functions, terms and behavior, such as "*I don't understand what this* arrange tiles *is all about*." (male, India), "*I just saw ten songs somewhere but I don't know how to get there*." (male, Finland) and "*This is a fun slideshow, it looks like a movie*." (female, India). Like the Singaporeans, participants in both Finland and India also made evaluative comments like "*This keypad is a little iffy*." (male, India). Compared to the other groups, the Indian participants tended to use more extreme expressions in their comments, for instance phrases such as "*I hate typing*." (female, India), "*Stop, now!*" (female, India) and "*What is the point of having this option?*" (male, India).

The Canadian verbalization pattern is clearly distinct from the other two patterns (Finnish-Indian and Singaporean) including an almost equal number of variations from expectation (19), stated frustrations (15), stated confusions (14), questions (13) and positive comments (13). The Canadians expressed confusion less often and asked fewer questions than the Finnish and Indian participants. However, they made more comparisons to their expectations and made more suggestions for improvement. They also made more positive comments than the Finnish and Indian participants.

4.3 Task Completion

In this section, the task completion rates are presented at a country level. Some individual differences were detected in the task completion rates. However, there were no significantly poorly performing participants in any of the country samples to whom the country level differences could be solely attributed to. In Finland, Singapore and Canada, the number of failed tasks varied from none to three per participant, and in

Canada the variation was between one and two failed tasks per participant. None of the participants in the sample fully completed all the tasks.

The task completion rates for Singapore are presented in Figure 10. For the Singaporean sample, the overall task completion rates were 92% completed and 70% fully completed. In the evaluation of task completion against the targets for primary and secondary tasks, altogether one secondary task (Check gallery details) and five primary tasks (Call Jake, Create playlist, Add contact to MM, Pause music and Play paused music) failed to meet the usability targets presented in Table 13. In Figures 10-13, primary tasks are indicated with (P) and secondary tasks with (S). The reasons for the altogether nine failed task performances included: moderator assistance needed (2) and, for the majority of failed tasks, the participant had the (stated) misconception that the task was correctly completed even though it was not (7).

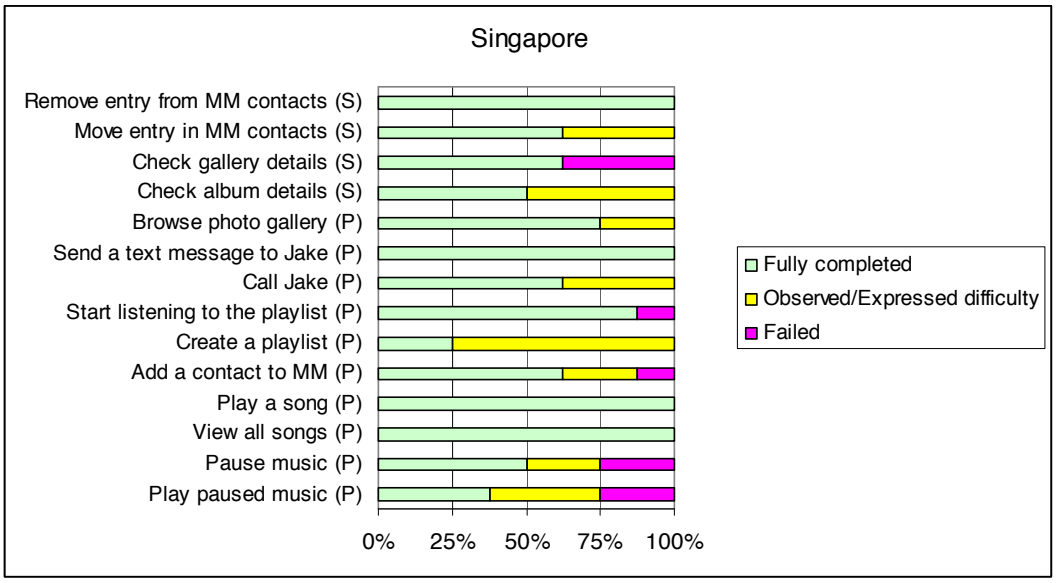


Figure 10. Task completion rates for Singapore.

The task completion rates for Finland are presented in Figure 11. The overall task completion rates for Finland were 87% completed and 71% fully completed. Altogether one secondary task (Check gallery details) and four primary tasks (Start listening to playlist, Create a playlist, Add contact to MM and Play paused music) failed to meet the usability targets. Out of the altogether fifteen failed task performances, six were completed with moderator assistance, and on nine occasions the participant incorrectly thought that the task had been completed.

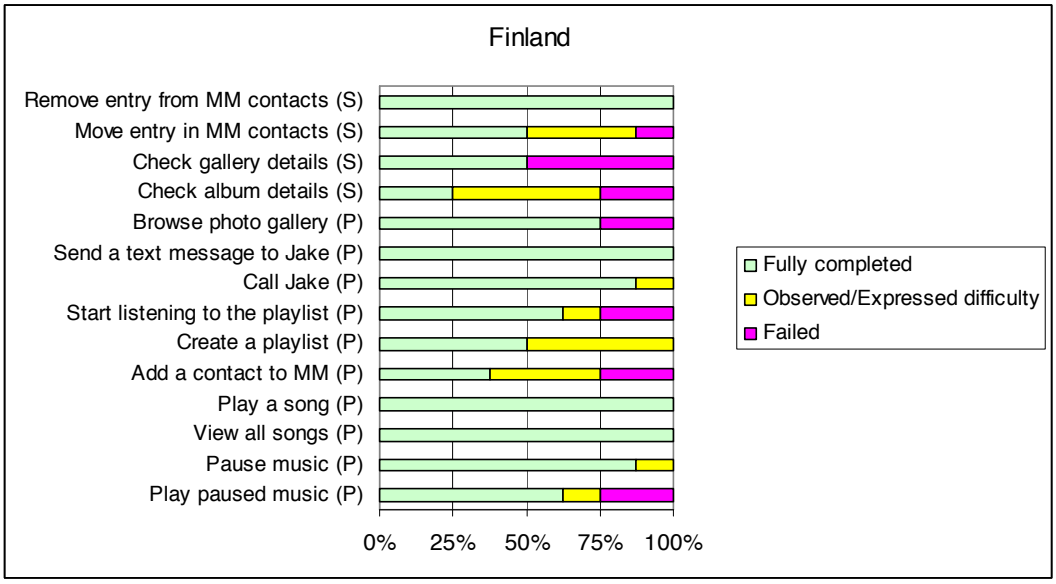


Figure 11. Task completion rates for Finland.

The task completion rates for Canada are presented in Figure 12. The overall task completion rates for Canada were 92 % completed and 75% fully completed. Altogether two secondary tasks (Check gallery details, Check album details) and one primary task (Play paused music) failed to meet the usability targets. In the Canadian sample, altogether nine task performances failed, out of which the participant was assisted on two occasions, gave up on two occasions and misinterpreted the task as correctly completed on five occasions.

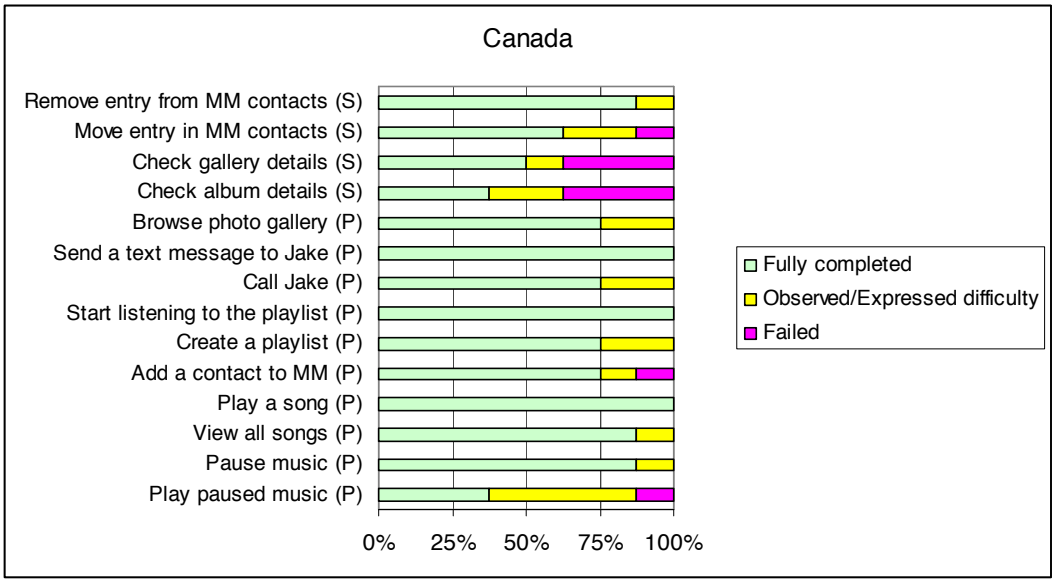


Figure 12. Task completion rates for Canada.

The task completion rates for India are presented in Figure 13. The overall task completion rates for India were 89% completed and 70% fully completed. Altogether one secondary task (Check gallery details) and three primary tasks (Browse photo gallery, Add contact to MM and Play paused music) failed to meet the usability targets. Out of the altogether twelve failed task performances, six occasions were assisted by the moderator, two failed due to the participant giving up, and on four occasions the participant had the misconception (stated) that the task was completed even though it was not.

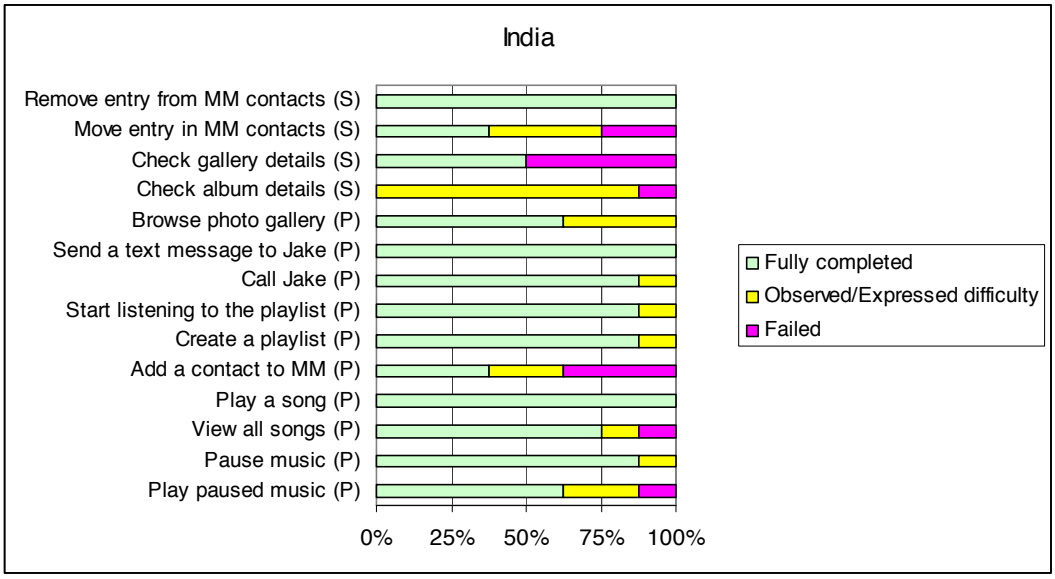


Figure 13. Task completion rates for India.

The overall task completion scores were relatively close to each other in all the countries varying from 87 % to 92% for completed tasks and from 70% to 75% for fully completed tasks. Singapore and Canada scored slightly better for both completed and fully completed tasks than India and Finland. However, when specific tasks were evaluated against usability targets, the number of tasks that failed to meet the target varied from three to six, and out of the fourteen tasks included in the test, altogether nine tasks scored less than required for the target in at least one of the countries. Accumulation of tasks that were identified as problematic was detected to some extent. Two of the tasks (one primary and one secondary) were identified as not reaching the target in all four countries. In addition, one task failed in three of the countries, and one failed in two of the countries.

4.4 SUS Scores

Both the mean and median SUS scores, presented in Figure 14, reached the set target of 75% in all the countries included in the study. Some degree of variation was detected in the individual scores, and in each of the countries there were one (India) to three (Singapore) participants who gave a score below the target. The mean deviations were the smallest in India (4.9) and Canada (5.4), somewhat greater in Finland (7.5) and the greatest in Singapore (9.7). The SUS scores for each participant are presented in Appendix 4. The difference in SUS scores between countries is minor; both mean and median SUS scores in all countries are within a range from 75.6 to 80.

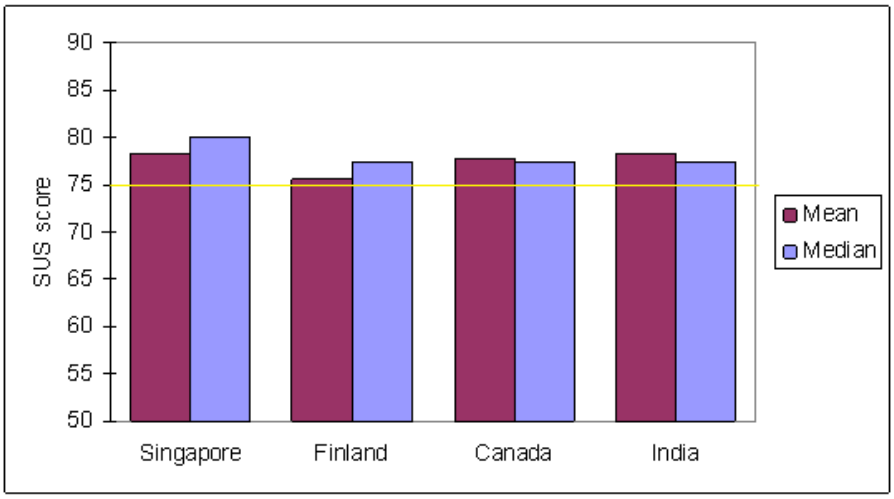


Figure 14. Mean and median SUS scores.

4.5 Usability Problems

A full list of usability problems identified in the study is presented in Appendix 5. The research uncovered altogether 48 usability problems in the system studied, out of which six were encountered by only one participant. The Finnish sample identified the most usability problems, altogether 38, equaling 79% of all the detected problems. Least usability problems were detected in Singapore, a total of 28 (58%). Canadians found 34 (71%) and Indian participants 33 (69%) usability problems.

A total of 19 (40%) usability problems were detected in all the countries. If the usability problems that were encountered by one participant are excluded, 45% of usability problems were identified in all countries included in the study. However, out of the total

48 usability problems, only 10 (21%) were unique (detected in only one country) and out of these problems six were identified only by one participant. Out of the unique usability problems, only one was detected by more than two participants, namely; four Indian participants interestingly misread Arrange *tiles* as Arrange *titles*.

Examination of the most prominent (encountered by three or more participants) usability problems, typically emphasized when prioritizing correction requirements, reveals that even though the Singaporean sample encountered fewer problems in total, the number of prominent problems was nearly equal for the Singaporean, Finnish and Indian samples, 13, 13, and 14 respectively. The Canadian sample identified the least prominent usability problems, 9 in total.

The cross-cultural study revealed several usability problems which would not have been recognized or which would have been interpreted differently if one or more countries been had been excluded from the sample. For instance, the hardware problems and few other usability problems, presented in Table 14, which likely originated from the speed of input and navigation, were prominently represented in the Singaporean sample. These problems may not have been considered severe in any of the other locations.

Table 14. Usability problems rooting in input and navigation speed.

Usability problem	Singapore	Finland	Canada	India
User scrolls past the tile she is looking for (correct tile is hard to recognize and distinguish from others)	4	2	1	2
User accidentally clicks the media keys (too close to other keys or not elevated enough from the surface)	7	3	2	2
User perceives buttons hard to press or feel (too much force needed)	5	1	2	1
User repeatedly clicks player controls due to slow response	4	2	1	1
User accidentally adds the same song to a playlist several times (song added indication is not prominent enough)	6	4	2	-

Another detected usability problem type, presented in Table 15, unequally represented in different countries, was likely rooted in the fact that the Finnish and the Canadian participants appeared to be examining the logic behind the system behavior and meaning of the different views more thoroughly. The majority of participants in Canada

and Finland were confused or annoyed by the fact that they could not predict or immediately reason the behavior of the Back button. The majority of Singaporean and Indian participants paid little attention to the fact that Back did not actually take them back to the last view they had visited. The usability problem can be considered particularly relevant from the viewpoint of the multimedia application that was being evaluated. However, the implications of the severity would have been different if the usability testing had been conducted in only some of the countries. Likely due to the same reason, some participants (four in Canada, three in Finland and two in India) were notably confused or annoyed by the fact that they had to focus their attention to choose between two different options when accidentally opening the SMS editor (the choice between deleting a message or saving a draft) rather than being able to cancel the accidental action. Most of the Singaporean and Indian, and some Canadian and Finnish participants, overcame the dialog easily just by clicking the End button or seemingly randomly selecting one of the options.

Table 15. Usability problems rooting in analytical examination of the UI.

Usability problem	Singapore	Finland	Canada	India
Mental model of quick launcher is faulty: the user does not realize that launching an item in MM takes her to a specific view inside and app and that selecting back is actually back in the app rather than back to the launcher	1	6	7	3
User gets looped in the SMS editor after accidentally launching it (Cancel functionality missing)	-	3	4	2

A prominent problem, presented in Table 16 and detected in all countries except Canada, was that the participants were looking for an option to copy or move a contact to the shortcut menu (find the contact you want in Phonebook and move it to the shortcut menu) rather than looking for an option to add a contact inside the shortcut menu (select add in the desired destination and then pick a contact to be added). Based on the verbal protocol and observations, no apparent reason for the difference in the choice of strategy in this task could be found.

Table 16. Usability problem rooting in incorrect mental model on moving or copying contacts.

Usability problem	Singapore	Finland	Canada	India
User is looking for Add to MM option in phonebook	5	5	-	3

In addition to the usability problems with varying frequencies depending on the country, several prominent universal usability problems, presented in Table 17, were detected.

Table 17. Universal usability problems.

Usability problem	Singapore	Finland	Canada	India
User has difficulty interpreting how the 5-way navigation key is mapped to Music player controls on the graphical user interface	5	3	5	2
User assumes that Albums includes all photos on the device	5	2	6	4
User selects Arrange tiles to rearrange items on a specific tile	7	7	6	3
User accidentally opens the SMS editor with select (not intuitive function for select)	5	4	2	7
User fails to notice the video file indication, confusing a video with a photo	6	3	4	5

5 Discussion

The significance of this Master's thesis emerges primarily from its practical implications for the mobile industry, resonating both from the in-depth examination and analysis of existing literature in the field of cross-cultural usability, and the empirical part of this research. Based on the literature review, the author argues, usability testing methodology lies at the heart of determining the scope and practices for cultural adaptation. If usability cannot be reliably and validly evaluated across cultures, also with hands-on industry methodology, it is challenging to design for and incorporate into product development process. Several researchers have studied usability testing across cultures (e.g. Clemmensen et al. 2009, Vatrapu & Pérez-Quiñones 2006, Evers 2002), and this thesis has presented guidelines, based on existing research, that can be adopted within the industry. However, due to the unique characteristics of mobile user interfaces, the product development process and the extraordinarily broad user base, more focused research from the perspective of the mobile industry is needed. The existing knowledge on cross-cultural usability testing was extended in the empirical part of this research, from the viewpoint of mobile multimedia, and the results obtained are discussed in the remainder of this chapter.

Regarding the fulfillment of usability targets, the results of this study are twofold. Satisfaction metrics, SUS scores (Brooke 1996) and the number of positive comments in verbal protocol, and the overall task completion rates did not indicate cultural differences that would have practical significance. However, in terms of the task completion rates related to individual tasks, minor, however not compelling, cultural differences were detected. The number of tasks that failed to fulfill the usability target set, varied between the countries (6/14 in Singapore, 5/14 in Finland, 4/14 in India and 3/14 in Canada), even though some overlap in problematic tasks was detected. For the SUS scores, the validity and reliability have been proven in several studies (Lewis & Sauro 2009) and, in general, employing standardized questionnaires for measuring satisfaction is recommended (Hornbæk 2006). However, Finstad (2006) presents evidence that the validity and reliability of SUS scores might be somewhat decreased for non-native English speakers due to understandability problems related to specific statements in the questionnaire. In this study, this was noted by giving the participants the possibility to ask questions related to the scale items. Fixed explanations for each of the scale items were used to further explain the meaning of each question if asked by

the participants. For the task completion rates, the evaluator effect, particularly notable due to the possibility of cultural bias in the evaluation, impacts the reliability of task completion rates to some extent. However, these reliability issues were addressed in the analyzing phase by following the guidelines proposed by Clemmensen et al. (2009).

For the evaluation of cross-cultural usability, the task completion rate proved to be a problematic, perhaps somewhat arbitrary, measure despite its established status and wide use in the practice of usability. The observations of the sessions revealed that participants from different cultures adopted varying approaches to completing the tasks complicating valid interpretations on the actual task completion. Sturm et al. (2005) predicted that interaction styles vary between cultures, which is supported by the findings of this study. The Singaporeans explored the various views and navigated significantly more than participants from other cultures. This behavior is perhaps rooted in the holistic cognitive style of Easterners; the preference towards flexibility, spontaneity and parallelism (Kim & Lee 2007). It may be also hypothesized, that low uncertainty avoidance scores for Singapore determined the explorative behavior; Marcus & Gould (2000) propose that low uncertainty avoidance may translate as a preference for wandering and risk taking when interacting with a computer system. If typical task completion indicators, such as the number of tries required to complete a task, were used without cultural consideration, the rates would likely indicate lack of usability not necessarily perceived by the participants. The observation can be considered as an improvement requirement for the use of usability metrics within the industry.

Another, likely culturally rooted, behavioral patter regarding the task completion was that the Canadian participants seemed to focus on giving constructive feedback rather focusing primarily on completing the tasks. The behavior could have been primed by the testing situation, perhaps the introduction and instructions given before the test, as suggested by Clemmensen (2009), with implications to the validity of task completion as an indicator of usability. Furthermore, both the Finnish and Canadian participants were frequently examining the logic of UI behavior and purpose of the different views, perhaps even more in the usability testing situation than they would have in actual use of the system. This behavior may be attributed to the analytic cognitive style of the Westerners, proposed to be manifesting as preference for rule-based structures, as well as planned and organized interaction style (Kim & Lee 2007). Notably, distinct patterns

characterizing the approach to task completion of the Indian participants were not identified. Similarly to the Finnish and Canadian participants, the Indians approached the tasks rather linearly and sequentially, judging from their verbalization and navigation, but seemed to be less analytical in their dissection of the UI. The findings related to task completion rates of the Indian participants indicate that even though both cultural cognition approach suggested by Clemmensen et al. (2009) and Hofstede's dimensions were found to be useful theoretical frameworks for interpreting the findings, the theoretical foundation related to cross-cultural usability perhaps still fails to predict and explain some cultural differences in usability between cultures.

Regarding the type and frequency of usability problems detected, cultural differences were relatively clear and practically significant. Several prominent usability problems were detected, for which the interpretation of severity would have differed had they been viewed based on the results of only one country, and few prominent problems would also have remained unidentified. Only approximately half of the usability problems were detected in all of the countries included in the sample. On the other hand, unique problems, identified only in one country, were scarce. Based on the rareness of unique usability problems, the fact that several universal usability problems were detected and the usability problem types, it may be argued that conflicting usability improvement requirements based on the findings were exiguous. In contrast, specific problems or problem types seemed to be emphasized in specific countries, largely due to the same reasons that were described to have impacted the task completion rates, namely the analytical approach of the Finnish and Canadian participants, and the explorative behavior of the Singaporeans. A possible interpretation for the lack of conflicting usability findings, and possible reason for the relatively small number cultural differences in usability problems regarding the UI style (for example, Select, Back and End key functionality, Options menu functionality, Phonebook and Media gallery concepts), is the experience dimension profile represented by the participants and the application domain of the studied system. It may be argued that the participants' experience with mobile phone use in general, and with mobile multimedia, had resulted into adaptation to platform and domain conventions, which dominated the perceived usability instead of cultural preferences, value orientation or cognitive style, as anticipated by Smith et al. (2004). However, the conclusion is rather hypothetical as different experience dimension profiles were not studied.

If used as a evaluative usability measurement, for instance, to evaluate the time or cost to implement the changes required before product launch, the number of usability problems identified in each country indicated some level of cultural differences: the Finnish sample encountered slightly more usability problems (38) than the Canadian (34) and Indian samples (33), and clearly more than the Singaporean sample (28). However, the findings regarding prominent problems (encountered by 3 or more participants), indicated nearly identical number of usability improvement requirements based on the Singaporean, Finnish and Indian samples (13, 13 and 14 respectively), and somewhat less need for improvement based on the Canadian sample (9). Similarly to the overall task completion rates, the comparison of the number of identified usability problems appears to have little cultural sensitivity. It may be inferred that if the overall task completion rates and the number of usability problems were used as a comparative metric between different releases, the results would be unpredictable, as the metrics do not seem to capture the cultural differences in how *quality of use* is constructed or how culturally rooted differences in behavior, likely not indicative of perceived usability, might bias the results. A specific set of usability improvements between the releases, for instance, regarding the response times of the user interface and the ergonomics of the keypad, may significantly improve the usability results yielded by the metrics in one culture but have practically no impact on the results of another.

Based on the task completion rates related to individual tasks and the number of usability problems detected, this research provides some evidence indicating cultural differences in the overall level of usability, namely that the Canadian sample identified fewer prominent usability problems and fewer tasks that did not fulfill the usability targets compared to the other countries. However, evidence related to differences between the other countries is inconclusive. Other researchers have also found cultural differences in overall usability difficult to point out with quantifiable measurements, namely performance measurements (Ford & Gelderblom 2003), and effectiveness, efficiency and satisfaction measurements (Vöhringer-Kuhnt 2002). This may be due to the inherent cultural bias built into the usability testing situation suggested by Clemmensen (2009), or perhaps the insensitivity of the used metrics to cultural differences that are observed in other ways, namely, by comparing the type and prominence of usability problems, qualitative observations and the verbal protocol. The findings of this study support both interpretations. Consequently, the practical implication is that critical evaluation and further development of the quantitative

usability metrics and the usability target setting is needed within the industry to ensure that differences in how usability is constructed in different cultures is taken into account in product development, and the cultural sensitivity of the UCD process is increased.

The findings indicating only minor cultural differences in the overall level of usability and the lack of conflicting usability improvement requirements provide some opposing arguments to the demands that, essentially, in order to be usable, the product needs to be comprehensively redesigned from the perspective of the target culture, proposed for instance by Sturm et al. (2005) and Shen et al. (2006). A possible explanation to the conflict of findings lies in the external validity of this research: the experience dimensions of the participants and the application domain employed in this research likely do not justify making comprehensive claims about ICT products in general, but rather indicate that the cultural variation in usability depends on the target user group and the application domain. Another explanation for the contradicting findings is that usability testing as a method is not sensitive towards, or even designed to, detecting variations in the perceived usefulness of the system or users' tasks and goals, but the testing focuses on effectiveness, efficiency and satisfaction. Thus, the results of this study do not provide comprehensive understanding regarding the cross-cultural product acceptance, which can be considered as a weakness of this research, while the models proposed by Sturm et al. (2005) and Shen et al. (2006) adopt a broader scope for cultural accommodation of user interfaces. However, the findings may also be interpreted as supporting and extending Cockton's (2009) notion of glocalized interfaces. Users may be able to adapt to and cherish foreign influences on a user interface, indicating that categorical demands for local interfaces should be subjected to critical evaluation.

The verbal protocol supports the finding that usability is constructed differently between cultures, even though the overall level of usability, quantitatively, may be close to equal. The verbalizations are a vital tool for usability problem detection, in general and in this research, and they also provide a means for further interpreting the task completion rates and identified usability problems. The verbal protocol, however, is relatively prone to cultural effects, impacted by cultural differences in cognition of the participants and dependent on the evaluator's perhaps culturally biased interpretation, which impacts the cross-cultural reliability and validity of usability problem detection. Moreover,

individual differences in the think-aloud activeness impact the validity of conclusions rooting in the verbal protocol.

The number of verbalizations in the Singaporean sample was significantly fewer than in the other samples, likely contributing to the fewer number of identified usability problems. The theoretical framework proposed by Clemmensen et al. (2009) predicted the mismatches between the think-aloud method and the cognitive style of Easterners. However, even though recommendations suggested by Clemmensen et al. (2009) were taken into account in the study design, the differences compared to the other countries were apparent. In addition to the thinking-aloud activeness, the pattern of usability-related verbalizations was distinct in the Singaporean sample, emphasizing general evaluative commentary and questions rather than focusing on specific UI behaviors or views, expressing emotions, or stating variation from expectations. The pattern may indicate that Singaporeans are more adaptable to the features and behavior of the system. The theoretical background of this claim lies in the cognitive style of Easterners, who are less prone to experiencing surprise and have the tendency to find compromises in seemingly contradictory situations (Clemmensen et al. 2009). The interpretation of adaptability of the Singaporeans is supported by the findings presented by Evers & Day (1997) which imply that Chinese users deemphasize the ease-of-use when they consider the system useful. However, it needs to be noted that equating Chinese and Singaporean with Chinese ethnic background is debatable. The accentuation of seemingly objective feedback (commenting on the keys, commending visual appearance of the interface or innovativeness features) may also be due to the high power distance decreasing the tendency to criticize the system if the moderator is considered as an authority, or as Clemmensen et al. (2009) suggest, due to the socio-emotional orientation of the Easterners.

Compared to the Singaporean sample, negative emotional expressions, namely stated confusion and frustration, were more common in the other three countries. A possible interpretation is to attribute the finding to the higher individualism scores in Finland, Canada and India, which may have provoked commenting the user interface from an individual perspective rather than adopting an objective perspective. Moreover, the uncertainty avoidance scores for Finland, Canada and India are higher than for Singapore, which may be interpreted so that the uncertainty and ambiguity faced during the test was perceived more uncomfortable and, thus, evoked more negative emotional

statements in the Finnish, Canadian and Indian samples. For the Finnish and Canadian samples, the number of statements describing variation from expectation was greater than detected in the Indian and Singaporean samples. The results may be due to the typical Western tendency to polarize and experience surprise as suggested by Clemmensen et al. (2009). After adopting a strategy to complete a task and failing to produce a successful outcome evoked the expressions of surprise in the Finnish and Canadian participants.

Regarding the justification of continuous cross-cultural usability testing, the findings indicate that there are practically significant cultural differences in the findings of the formative think-aloud usability test, and that testing in a single location does not provide a basis for comprehensive understanding of the usability of a user interface. However, the findings also indicate that there are limitations regarding the industrial cross-cultural usability testing practice that need to be considered in order for the testing to be effective and in order to ensure cross-cultural validity of the results. Cross-cultural usability testing in different phases of the product development cycle may prove to be challenging from the viewpoint of validity of the results. Iterative usability testing often involves prototype research conducted before product-level software is available. The findings in this study suggest, however, that testing with a prototype with limited functionality may pose threats to the validity of the usability testing results in Eastern countries; the Singaporeans relied on exploration rather than following a sequential strategy when completing the tasks, which is not supported by prototype testing. Moreover, prototype research often involves participants giving feedback on specific views and discussing the different steps required to complete a task. However, both the Indian and Singaporean participants seemed to pay relatively little attention to the details of the UI and did not dissect the purpose of each view and the underlying logic of UI behavior. Discussing the flows and views step-by-step in prototype research may not yield valid results and may be unnatural for the participants. For the evaluative testing situating in a late phase of product development, and sometimes after launch, the findings of this research indicate that the commonly employed metrics may not provide a comprehensive or even correct evaluation on the usability across cultures. Firstly, this study indicated that the Singaporean sample was more adaptable to the behavior of the user interface, and thus, measures that emphasize effectiveness, efficiency and encountered errors may not provide a comprehensive measure of usability for the Easterners. Secondly, as the styles of interaction were found to vary between the

cultures – from analytical and sequential to explorative and risk-taking – the results indicate that careful consideration is required when task completion rates are used as usability indicators across cultures.

This research provides a broad examination of the impact of culture on the usability test findings, which can be considered as a strength of the study, in addition to the contribution to the industrial usability testing practice. The external validity of the study was addressed by conducting the study in several countries, including Eastern and Western countries and extensive coverage of Hofstede's dimensions. The fact that the sample included only English-speaking participants, however, may have impacted the results to some extent. Clemmensen (2009) suggests that participants may adopt different cultural models of usability depending on the usability testing situation and the studied user interface. It is possible that particularly for the non-Western samples, the results may fail to uncover some cultural differences and may indicate behaviors that the users would not adopt outside the usability testing situation. The sampling also involved internal validity considerations as the experience dimensions and current phone model were controlled decreasing the possibility that differences in usability could be attributed to other than cultural factors. Due to the controlled experience dimension profile and the fact that only one user interface was tested, however, the possibilities to generalize the findings are limited. The sample size was relatively small within each country, increasing the possibility that individual differences impacted the results, and thus decreasing the external and internal validity to some extent. The construct validity was addressed by the inclusion of several different ways of measuring and observing usability and by using standardized or established methodology.

On one hand, the field of cross-cultural usability is still under construction and perhaps the recent revolutions in the mobile industry – convergence of devices and services, growth of mobile applications market and open source development (Oulu Innovation 2010) – will expand the scope of cross-cultural usability in mobile domain even further. On the other hand, the developments will provide a chance to study cross-cultural usability from a novel perspective, as applications will be designed and developed by 3rd parties, likely from more local viewpoint, all over the world.

6 Conclusions

Cross-cultural usability has become an important consideration within the mobile industry, as products are targeted at an increasingly international user group; already, more than a third of the world's population has a mobile phone. Within the industry, there is a need for more in-depth understanding on how culture impacts the usability of the products and, especially, how cross-cultural usability should be incorporated into the existing product development processes. This Master's thesis studied cross-cultural usability from the perspective of usability engineering in the mobile multimedia application development. An extensive review on the cross-cultural usability literature was conducted to outline the theoretical foundation for the interrelationship between culture and usability, to explore frameworks for approaching cultural accommodation in the industry, and to identify the gaps in the body of knowledge from the viewpoint of practitioners in mobile domain. In the empirical part of this research, the current cross-cultural usability testing practice was evaluated by comparing the fulfillment of usability targets and the identification of usability problems in a think-aloud usability test conducted within the target user group, experienced mobile phone users and multimedia users, in Singapore, Finland, Canada and India. The purpose of this study was to evaluate whether differences in usability testing findings have practical significance to justify expensive and time-consuming cross-cultural usability testing. Furthermore, the think-aloud usability testing methodology and metrics employed in the product development were evaluated from a cross-cultural viewpoint.

The verbal protocol and qualitative observations of the sessions proved to be a valuable tool for interpreting the usability test findings. Compelling evidence of cultural differences was found regarding verbalization activeness and patterns. The Singaporeans were clearly less active in their verbalizations compared to the Finns, Canadians and Indians, complicating the usability problem detection. Thus, it may be concluded that the number of encountered usability problems in a think-aloud usability test should not be used as a comparative metric of usability without careful cultural consideration if Easterners are included in the study sample. The verbalization pattern of the Singaporeans emphasized general evaluative feedback seemingly unrelated to specific view on the UI or specific UI behavior. The emotional expressions of frustration and confusion were clearly fewer than in the other countries. The findings suggest that Easterners may be more adaptable to the characteristics or a user interface

and, consequently, careful consideration should be used if comprehensive claims about usability are made based on usability metrics that emphasize effectiveness, efficiency or number of errors.

The observations of the sessions, alongside with the verbal protocol, revealed that participants from different cultures adopted varying approaches to completing the tasks. The Finnish and the Canadian participants adopted a sequential and analytical approach, examining the purpose of the views and logic of the UI behavior, the Indian participants also adopted sequential approach and were able to verbalize their task completion strategy, however dissecting the UI less analytically than the Finns and the Canadians, and the Singaporeans adopted a more exploratory approach, navigating on the UI clearly more than the participants from the other countries. Based on the findings, it may be concluded that the task completion rates are prone to cultural misinterpretations, namely, that careful consideration is required if task completion rates are used as an indicator of usability.

In the fulfillment of usability targets for task completion, minor, however not compelling, cultural differences were detected: namely, the number of tasks that failed to meet the targets set, varied from 3/14 (Canada) to 6/14 (Singapore) and the altogether 9/14 tasks failed to fulfill the target in at least one of the countries. The satisfaction metric, the SUS score, did not reveal cultural differences in the fulfillment of the target set. Moreover, the overall task completion scores (regarding the entire task set) were close to identical in all the countries studies. Related to the type and frequency of usability problems, however, cultural differences were relatively clear, and practically significant. Several prominent usability problems were detected, for which the interpretation of severity would have differed had they been viewed based on the results of only one country, and few prominent problems would also have remained unidentified. Supported by the qualitative observations and verbalizations, several usability problems were attributed to the different approaches that the users adopted when completing the tasks, namely the input and navigation speed and the explorative interaction style of the Singaporeans and the careful examination of the UI by the Finns and the Canadians. Importantly, practically no conflicting usability improvement requirements were detected. Instead, the cross-cultural usability study yielded a richer, and more complete, set of usability improvements compared to testing in just one national culture. Furthermore, several universal usability problems were identified in

the study. Thus, it may be concluded that in the case of mobile multimedia targeted at experienced mobile phone and multimedia users, cultural differences in usability exist, but rather than in the overall level of usability – the quantity of usability – the differences exist in how the *quality of use* is constructed in different cultures. Furthermore, the study suggests users' experience with mobile phones and the application domain might have an impact on the magnitude of cultural differences in usability, but the claim needs to be subjected to further study.

The practical implications to the usability testing practice in the mobile industry suggested by this research are manifold. This research provides evidence to support the importance of cross-cultural formative usability testing: testing in a single country can not be considered a comprehensive evaluation on the usability of the user interface. However, several considerations that need to be addressed to ensure the cross-cultural applicability of usability testing methodology were indentified in the study, alongside with limitations posed by cultural differences related to established usability measures and the extent to which they measure usability reliably and broadly. Supported by the review of existing body of knowledge within the cross-cultural HCI, the findings show that, like user interfaces, usability testing and metrics need to be globalized and localized. Recommendations to improve cross-cultural usability practice within mobile industry include:

- When measuring usability, attention should to be paid on the cross-cultural applicability of the measures. Evaluation of task completion rates and identification of usability problems is prone to cultural bias due to cultural differences in how users complete tasks and approach usability testing. Furthermore efficiency, effectiveness and error focused measures may not measure usability broadly across cultures.
- Usability target setting, effective and valid across cultures, should be further developed. Task completion rates, number of usability problems encountered and satisfaction measures may be insensitive to cultural differences detected by other means, namely, the verbal protocol and analyzing the detected usability problems.
- Novel usability testing methodology should be employed to support usability problem detection in the Eastern markets, where participants may be limited by the verbal protocol.

- Novel prototype usability research methodology should be employed in order to ensure validity outside the Western markets. The limited exploration possibilities (emphasis on linear task completion strategy) and the analytical review of specific screens and flows may be unnatural for non-western participants.

The recommendations introduce a variety of topics for future research. Novel methods for cross-cultural usability testing of both product software and prototypes should be explored and validated. Furthermore, the comparison of different usability metrics across cultures is needed, for instance studying the weightings of different attributes of usability in different cultures. The research related to cross-cultural usability of mobile interfaces should be expanded to user groups with different experience dimension profiles and to different application domains to extend and validate the findings presented in this research and to support the practitioners within mobile domain. Importantly, more research is needed to support different levels of cultural considerations, especially the contextual level, determining the required functionality and necessary elements of the UI.

7 References

Ambler S (2010) Introduction to Agile Usability User Experience Activities on Agile Development Projects. [online]. Date of publish unknown [retrieved 11.5.2010]. Available: http://www.agilemodeling.com/essays/agileUsability.htm.

Bangor A, Kortum P & Miller J (2008) An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction, 24, (6), pp. 574-594. ISSN: 1044-7318.

Barber W & Badre A (1998) Culturability: The Merging of Culture and Usability. [online] In: Proceedings of the 4th Conference on Human Factors and the Web. Published June 1998. [retrieved 30.11. 2006]. Available: http://research.microsoft.com/en-us/um/people/marycz/hfweb98/barber/.

Bevan (2008) UX, Usability and ISO Standards. Now Let's Do It In Practice: User Experience Evaluation Methods In Product Development. [online]. In: Proceedings of the workshop on CHI 2008. Published April 2008 [retrieved 20.3.2010]. Available: www.cs.tut.fi/ihte/CHI08_workshop/papers/Bevan_UXEM_CHI08_06April08.pdf.

Bevan N, Kirakowski J & Maissel J (1991) What Is Usability? In: Bullinger H J (Ed) Proceedings of the 4th International Conference on Human Computer Interaction, Stuttgart, Elsevier.

Beyer H & Holtzblatt K (1998) Contextual Design: Defining Customer-Centered Systems. San Francisco, Morgan Kaufman.. 472 p. ISBN: 1-55860-411-1.

Blom J, Chipchase J & Lehikoinen J (2005) Contextual and Cultural Challenges for User Mobility Research. Communications of the ACM, 48, (7), pp. 37 – 41. ISSN: 0001-0782.

Bourges-Waldegg P & Scrivener S A R (2000) Applying and testing an approach to design for culturally diverse user groups. Interacting with Computers, 13, (2), pp. 111-126. ISSN: 0953-5438.

Bourges-Waldegg P & Scrivener S A R (1998) Meaning, the central issue in cross-cultural HCI design. Interacting with Computers, 9, (3), pp. 287-309. ISSN: 0953-5438.

Brooke J (1996) SUS: A "Quick and Dirty" Usability Scale. In: Jordan P W, Thomas B, Weerdmeester B A & McClelland A L (Eds) Usability Evaluation in Industry. pp. 189-194. London, Taylor & Francis, London, UK. pp. 189-194. ISBN: 978-0748404605.

Carroll J (2002) Human-Computer Interaction in the New Millennium. New York, Addison-Wesley Professional. 703 p. ISBN: 0-201-70447-1.

Carroll J (1997) Human-Computer Interaction: Psychology as a Science of Design. Annual Review of Psychology, 48, pp.61-83.

Clemmensen T (2009) Towards a Theory of Cultural Usability: A Comparison of ADA and CMU Theory. [online] Presented in HCI2009, San Diego. Published 2009 [retrieved 15.2.2010]. Available: http://culturalusability.cbs.dk/downloads/HCII%202009/hcii2009%20clemmensen%20proofread%20v4.pdf.

Clemmensen T, Hertzum M, Hornb K, Shi Q & Yammiyavar P (2009) Cultural Cognition in Usability Evaluation. Interacting with Computers, 21, (3), pp. 212–220. ISSN: 0953-5438.

Cockton G (2009) Understanding Cultural Differences in HCI: The Diamond Model of Culture. [online]. Presentation given at Cross-Cultural UX Design Seminar 2009 in Tampere University of Technology. Published 18.9.2009 [retrieved 12.4.2010]. Available: http://www.cs.tut.fi/ihte/projects/suxes/pdf/Cross-Cultural%20UX%20Design%20Seminar_180909_all%20presentations.pdf.

Cooper A, Reimann R & Cronin D (2007) About Face 3: The Essentials of Interaction Design. Indianapolis, Wiley Publishing. 610 p. ISBN: 978-0-470-08411-3.

Davis F D (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information technology. MIS Quarterly, 13, (3), pp. 319-340.

del Galdo E (1990) Internationalization and Translation. In: Nielsen J (Ed) Designing User Interfaces for International Use. Amsterdam, Elsevier Science Publishers, pp. 1-10. Advances in Human Factors / Ergonomics, 13. ISBN 0-444-88429-9.

Downey S, Cordova-Wentling R M, Wentling T & Wadsworth A (2005) The Relationship Between National Culture and the Usability of an E-Learning System. Human Resource Development International, 8, (1), 2005, pp.47-64.

Ebling M R & John B E (2000) On the Contributions of Different Empirical Data in Usability Testing. In: Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques, New York, pp. 289 – 296. ISBN: 1-58113-219-0.

Encyclopædia Britannica (2006) Culture. [online]. Encyclopædia Britannica Academic Edition. Date of publish unknown [retrieved 5.12. 2006]. Available: http://search.eb.com/eb/article-9109857Encyclopædia.

Everett S, Byrne M & Greene K (2006) Measuring the Usability of Paper Ballots: Efficiency, Effectiveness and Satisfaction. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, Santa Monica, pp. 2547-2551.

Evers V (2002) Cross-Cultural Applicability of User Evaluation Methods: A Case Study amongst Japanese, North-American, English and Dutch Users. In: CHI '02 extended abstracts on Human factors in computing systems, Minneapolis, pp. 740 – 741. ISBN: 1-58113-454-1.

Evers V (2001) Cultural Aspects of User Interface Understanding An Empirical Evaluation of an E-Learning Website by International User Groups. Doctoral Thesis. University of Amsterdam. 377 p.

Evers V & Day D (1997) The Role of Culture in Interface Acceptance. In: Howard S, Hammond J & Lindegaard G (Eds) Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction, Sydney. London, Chapman & Hall Ltd., pp.260-267. ISBN: 0-412-80950-8.

Fallman D (2003) Design-oriented Human-Computer Interaction. In: Proceedings of the SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, pp. 225-232. ISBN: 1-58113-630-7.

Fernandes T (1995) Global Interface Design: A Guide to Designing International User Interfaces. Boston, AP Professional, Inc. 191 p. ISBN: 0-12-253790-4.

Finstad K (2006) The System Usability Scale and Non-Native English Speakers. Journal of Usability Studies, 1, (4), pp. 185-188. ISSN: 1931-3357.

Ford G & Gelderblom H (2003) The Effects of Culture on Performance Achieved through the Use of Human Computer Interaction. In: Proceedings of the 2003 annual research conference of the South African Institute of Computer Scientists and Information Technologists on Enablement through Technology, Gauteng, pp. 218-230. ISBN: 1-58113-774-5.

Gould E W, Zakaria N & Yusof S A M (2000) Applying Culture to Website Design: A Comparison of Malaysian and US Websites. In: Proceedings of IEEE professional communication society international professional communication conference and Proceedings of the 18th annual ACM international conference on Computer documentation: technology & teamwork, Cambridge. Piscataway, IEEE Educational Activities Department, pp. 161-171. ISBN: 0-7803-6431-7.

Hall E T (1989) Beyond Culture. 2nd ed. New York, Doubleday. 298 p. ISBN: 0-385-12474-0.

Hagen P, Robertson T, Kan M & Sadler K (2005) Emerging Research Methods for Understanding Mobile Technology Use. In: Proceedings of OZCHI 2005, Canberra, pp. 1-10. ISBN: 1-59593-222-4.

Hassenzahl M (2008) User Experience (UX): Towards an Experiential Perspective on Product Quality. In: Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, Metz. New York, ACM, pp. 11-15. ISBN: 978-1-60558-285-6.

Herman L (1996) Towards Effective Usability Evaluation in Asia: Cross-Cultural Differences. In: Proceedings of the 6th Australian Conference on Computer-Human Interaction (OZCHI '96), Hamilton. Washington, IEEE, p.135. ISBN: 0-8186-7525-X.

Hewett T, Baecker R, Card S, Carey T, Gasen J, Mantei M, Perlman G, Strong G & Verplank W (1996) ACM SIGCHI Curricula for Human-Computer Interaction. [online]. New York, Association for Computing Machinery, Inc. 90 p. Published 1996, last updated 7.3.2007 [referred 11.3.2010]. Available: http://old.sigchi.org/cdg/cdg2.html.

Hofstede G (2001) Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. 2nd ed. Thousand Oaks, Sage Publications. 596 p. ISBN: 0-8039-7323-3.

Hofstede G (1991) Cultures and Organizations: Software of the Mind. New York, McGraw-Hill. 279 p. ISBN 0-07-029307-4.

Hoft N (1996) Developing a Cultural Model. In: del Galdo E M & Nielsen J (Eds) International User Interfaces. New York, John Wiley & Sons, Inc. pp. 41-72. ISBN: 0-471-14965-9.

Holden N J (2002) Cross-Cultural Management: A Knowledge Management Perspective. Harlow, Pearson Education Limited. 328 p. ISBN: 0-273-64680-X.

Honold P (2000) Culture and Context: An Empirical Study for the Development of a Framework for the Elicitation of Cultural Influence in Product Usage. International Journal of Human-Computer Interaction, 12, (3/4), pp. 327-345. ISSN: 1044-7318.

Hornbæk K (2006) Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human-Computer Studies, 64, (2), pp.79-102. ISSN: 1071-5819.

Horton W (2005) Graphics: The Not Quite Universal Language. In: Aykin N (Ed) Usability and Internationalization of Information Technology. New Jersey, Lawrence Erlbaum Associates, Inc., Publishers. pp. 157-187. ISBN: 0-8058-4479-1.

Human Factors and Ergonomics Society (2010) HFES History. [online]. Date of publish unknown [retrieved 1.4. 2010]. Available: www.hfes.org/web/AboutHFES/history.html.

Hvannberg E B & Law L-C (2003) Classification of Usability Problems (CUP) Scheme. In: Rauterberg, Menozzi M & Wesson J (Eds) Proceedings of Human-Computer Interaction – Interact'03, Zurich, pp. 655-662.

Häkkilä J (2006) Usability with Context-Aware Mobile Applications Case Studies and Design Guidelines. Dissertation. University of Oulu, Acta Universitatis Ouluensis C263. 77 p. ISBN 951-42-8323-6.

International Ergonomics Association (2010) What is Ergonomics. [online]. Published 31.3.2010 [retrieved 10.5. 2010]. Available: http://www.iea.cc/browse.php?contID=what_is_ergonomics.

International Telecommunication Union (2010) Measuring the Information Society. Geneva, International Telecommunication Union. 124 p. ISBN 92-61-13111-5.

ISO/IEC 9126-1 (2001) Quality Model. Geneve, International Organization for Standardization. 25 p.

ISO 9241-11 (1998) Guidance on Usability. Geneve, International Organization for Standardization. 28 p.

ISO 13407 (1999) Human-centred design processes for interactive systems. Geneve, International Organization for Standardization. 31p.

Ito & Nakakoyi (1996) Impact of Culture on User Interface Design. In: del Galdo E M & Nielsen J (Eds) International User Interfaces. New York, John Wiley & Sons. pp. 105-126. ISBN: 0-471-14965-9.

Jacobsen N E, Hertzum M, John B E (1998) The Evaluator Effect in Usability Studies: Problem Detection and Severity Judgements. In: Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, pp. 1336-1340.

Jokela T, Iivari N, Matero J & Karukka M (2003) The Standard of User-Centered Design and the Standard Definition of Usability: Analyzing ISO 13407 against ISO 9241-11. In: CLIHC Proceedings of the Latin American Conference on Human-Computer interaction, volume 46, Rio de Janeiro, pp. 53-60.

Jones M & Marsden G (2006) Mobile Interaction Design. Chichester, John Wiley & Sons. 377 p. ISBN : 0-407-09089-8.

Joshi D & Avasthi V (2007) Mobile Internet UX for Developing Countries. Position Paper. In: Proceedings of Mobile HCI 2007, Singapore, pp. 16-20.

Kaikkonen A, Kekäläinen A, Cankar M , Kallio T & Kankainen A (2005) Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. Journal of Usability Studies, 1, (1), pp. 4-17. ISSN: 1931-3357.

Kamppuri M, Bednarik R & Tukiainen M (2006) The Expanding Focus of HCI: Case Culture. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction, Oslo. New York, ACM Press, pp. 405-408. ISBN: 1-59593-325-5.

Keenan S L, Hartson H R, Kafura D G & Schulman R S (1999) The Usability Problem Taxonomy: A Framework for Classification and Analysis. In: Empirical Software Engineering 4. Boston, Kluwer Academic Publishers, pp. 71–104.

Ketola P (2002) Integrating Usability with Concurrent Engineering in Mobile Phone Development. Dissertation. University of Tampere. 144 p. ISBN: 951-44-5359-X.

Ketola P & Röykkee M (2001) The Three Facets of Usability In Mobile Handsets. In: Proceedings of CHI 2001 Workshop Mobile Communications: Understanding Users, Adoption & Design, Seattle.

Khaslavsky J (1998) Integrating culture into interface design. In: CHI 98 Conference Summary on Human Factors in Computing Systems, Los Angeles. New York, ACM Press, pp. 365-366. ISBN: 1-58113-028-7.

Kiljander H (2004) Evolution and Usability of Mobile Phone Interaction Styles. Dissertation. Helsinki University of Technology. Espoo, Otamedia Oy, Publications in Telecommunications Software and Multimedia TML-A8. 234 p. ISBN: 951-22-7319-5.

Kim H, Kim J, Lee Y, Chae M & Choi Y (2002) An Empirical Study of the Use Contexts and Usability Problems in Mobile Internet. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS'02), Volume 5, p.132. ISBN: 0-7695-1435-9.

Kim J H & Lee K P (2007) Culturally Adapted Mobile Phone Interface Design: Correlation between Categorization Style and Menu Structure. In: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services, Singapore, pp. 379-382. ISBN: 978-1-59593-862-6.

Kuniavsky M (2003) Observing the User Experience: A Practitioner's Guide to User Research. San Francisco, Morgan Kaufman Publishers. 560 p. ISBN: 1-55860-923-7.

Kuutti K (1995) Activity Theory as a Potential Framework for Human-Computer Interaction Research. In: Nardi B (Ed) Context and Consciousness: Activity Theory and Human Computer Interaction. Cambridge, MIT Press, pp. 17-44.

Lewis J M (2006) Sample Sizes for Usability Tests: Mostly Math, Not Magic. Interactions, 13, (6), pp. 29-33. ISSN: 1072-5520.

Lewis J R & Sauro J (2009) The Factor Structure of the System Usability Scale. In: Proceedings of the 1st International Conference on Human Centered Design, San Diego, pp. 94-103. ISBN: 978-3-642-02805-2.

Localization Industry Standards Association (2010) OSCAR, LISA's Standards Committee. [online]. Date of publish unknown [retrieved 8.4.2010]. Available: http://www.lisa.org/Standards.79.0.html.

Marcus A (2002) Culture Class vs. Culture Clash. Interactions, 9, (3), pp. 25-28. ISSN: 1072-5520.

Marcus A (2001) Cross-Cultural User-Interface Design. In: Smith M J & Salvendy G (Eds) Proceedings of Human-Computer Interface International (HCII) Conference, volume 2, New Orleans. Mahwah, Lawrence Erlbaum Associates, pp. 502-505.

Marcus A & Gould E W (2000) Crosscurrents: Cultural Dimensions and Global Web User-interface Design. Interactions, 7, (4), pp. 32-46. ISSN: 1072-5520.

Massey A P, Hung Y-T C, Montoya-Weiss M & Ramesh V (2001) When Culture and Style Aren't About Clothes: Perceptions of Task-Technology "Fit" in Global Virtual Teams. In: Ellis C & Zigurs I (Eds) Proceedings of the 2001 international ACM SIGGROUP Conference on Supporting Group Work, Boulder, Colorado. New York, ACM Press, pp. 207-213. ISBN: 1-58113-294-8.

Mayhew D J & Bias R G (2005) Cost-Justifying Usability Engineering for Cross-Cultural User Interface Design. In: Aykin N (Ed) Usability and Internationalization of Information Technology. New Jersey, Lawrence Erlbaum Associates, Inc., Publishers, pp. 213-249. ISBN: 0-8058-4479-1.

McNamara N & Kirakowski J (2005) Defining Usability: Quality of Use or Quality of Experience? In: Proceedings of International Professional Communication Conference 2005 (IPCC 2005), pp. 200-204. ISBN: 0-7803-9027-X.

Millen D R (2000) Rapid Ethnography: Time Deepening Strategies for HCI Field Research. In: Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques. New York City, pp. 280 – 286. ISBN: 1-58113-219-0.

Mrazek D & Baldacchini C (1997) Avoiding Cultural False Positives. Interactions, 4, (4), pp. 19-24. ISSN: 1072-5520.

Myers M D & Tan F D (2003) Beyond Models of National Culture in Information Systems Research. In: Tan F D (Ed) Advanced Topics in Global Information Management. Hershey, PA, Idea Group Publishing. pp. 14-29. ISBN: 1-59140-064-3.

Nielsen J (2001) Success Rate: The Simplest Usability Metric. [online]. Published 18.2.2001 [retrieved 1.5. 2010]. Available: www.useit.com/alertbox/20010218.html.

Nielsen J (1993) Usability Engineering. San Francisco, CA, Morgan Kaufmann Publishers Inc. ISBN: 0-12-518406-9.

Nielsen J (1990a) Preface. In: Nielsen J (Ed) Designing User Interfaces for International Use. Amsterdam, Elsevier Science Publishers. pp. v-vii. Advances in Human Factors / Ergonomics, 13. ISBN 0-444-88429-9.

Nielsen J (1990b) Usability Testing of International User Interfaces. In: Nielsen J (Ed) Designing User Interfaces for International Use. Amsterdam, Elsevier Science Publishers. pp. Advances in Human Factors / Ergonomics, 13. ISBN: 0-444-88429-9.

Nielsen J & Mack R L (1994) Usability Inspection Methods. New York, John Wiley & Sons, Inc. 413 p. ISBN: 0-471-01877-5.

Nokia (2010) Media Resources. [online]. Date of publish unknown [retrieved 8.4.2010]. Available: http://www.nokia.com/press/media_resources.

Onibere E A, Morgan S, Busang E M & Mpoeleng D (2001) Human-Computer Interface Design Issues for a Multi-Cultural and Multi-Lingual English Speaking Country – Botswana. Interacting with Computers, 13, (4), pp. 497-512(16). ISSN: 0953-5438.

Oshlyansky L, Thimbleby H & Cairns P (2004) Breaking Affordance: Culture as Context. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction (NordiCHI'04), Tampere. New York, ACM Press, pp. 81-84. ISBN: 1-58113-857-1.

Oulasvirta A, Tamminen S, Roto V & Kuorelahti J (2005) Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing systems, Portland, pp. 919 – 928. ISBN: 1-58113-998-5.

Oulu Innovation (2010) Mobile Linux - Towards New Mobile Ecosystems and Business Models. [online]. Innovator, 1. Published 2010 [retrieved 10.5.2010]. Available: http://issuu.com/ouluinnovation/docs/innovator_mobilelinux.

Preece J, Rogers Y & Sharp H (2002) Interaction Design: Beyond Human-Computer Interaction. New York, John Wiley & Sons, Inc. 519 p. ISBN: 0-471-49278-7.

Preece J, Rogers Y, Sharp H, Benyon D, Holland S & Carey T (1994) Human-Computer Interaction. Harlow, Addison-Wesley Longman Limited, Inc. 775 p. ISBN: 0-201-62769-8.

Rosenzweig E (2006) World Usability Day: A Challenge for Everyone. Journal of Usability Studies, 1, (4), pp151-155. ISSN: 1931-3357.

Russo P & Boor S (1993) How Fluent is Your Interface? Designing for International Users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Amsterdam. New York, ACM press, pp. 342-347. ISBN: 0-89791-575-5.

Sacher H, Tng T-H, & Loudon D (2001) Beyond Translation: Approaches to Interactive Products for Chinese Consumers. International Journal of Human-Computer Interaction, 13, (1), pp. 41-51. ISSN: 1044-7318.

Sarjanoja A-H (2008) Prototyping Tool for Designing Animated Mobile User Interfaces. Master's Thesis. University of Oulu. 63p.

Shen S-T, Woolley M & Prior S (2006) Towards Culture-Centred Design. Interacting with Computers, 18, (4), pp. 820-852. ISSN: 0953-5438.

Smith A, Dunckley L, French T, Minocha S & Chang Y (2004) A Process Model for Developing Usable Cross-Cultural Websites. Interacting with Computers, 16, (1), pp. 63-91. ISSN: 0953-5438.

Smith D & Yetim F (2004) Editorial: Global Human–Computer Systems: Cultural Determinants of Usability. Interacting with Computers, 16, (1), pp. 1-5. ISSN: 0953-5438.

Soares A M, Farhangmehr M & Shoham A (2007) Hofstede's Dimensions of Culture in International Marketing Studies. Journal of Business Research, 60, (3), pp. 277-284.

Sturm C, Moreno-Rocha M A & Quaet-Faslem P (2005) Questioning the Implicitness: Issues in Developing and Testing Mobile Devices for International Markets. Paper presented at the 11th International Conference on Human-Computer Interaction, Las Vegas.

Sukaviriya P & Moran L (1990) User Interfaces for Asia. In: Nielsen J (Ed) Designing User Interfaces for International Use. Amsterdam, Elsevier Science Publishers. pp. 189-218. Advances in Human Factors / Ergonomics, 13. ISBN 0-444-88429-9.

Sun H (2006) The Triumph of Users: Achieving Cultural Usability Goals With User Localization. Technical Communication Quarterly, 15, (4), pp. 457–481.

Sun H (2004) Expanding the Scope of Localization: A Cultural Usability Perspective on Mobile Text Messaging Use in American and Chinese Contexts. Doctoral Thesis. Troy, New York, Rensselaer Polytechnic Institute. 287 p.

Sun H (2002a) Exploring Cultural Usability. In: Proceedings of Professional Communication Conference 2002 (IPCC 2002), Portland. Piscataway, IEEE International, pp.319-330. ISBN: 0-7803-7591-2.

Sun H (2002b) Why Cultural Contexts Are Missing: A Rhetorical Critique of Localization Practices. [online]. 2002 Annual Conference – Society for Technical Communication. Published 2002 [retrieved 3.2. 2010]. Available: http://www.stc.org/confproceed/2002/PDFs/STC49-00090.pdf.

Taylor D (1990) Creating International Applications A Hands-On Approach using the Hewlett-Packard NLS Package. In: Nielsen J (Ed) Designing User Interfaces for International Use. Amsterdam, Elsevier Science Publishers. pp. 123-158. Advances in Human Factors / Ergonomics, 13. ISBN 0-444-88429-9.

Tractinsky N (1997) Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, pp. 115 – 122. ISBN: 0-89791-802-9.

Trompenaars F (1993) Riding the Waves of Culture. London, Nicholas Brealey Publishing. 192 p. ISBN: 1-85788-0331.

Tullis T & Albert B (2010) Measuring the User Experience [online]. Date of publish unknown [retrieved 15.4.2010]. Available: http://measuringux.com/.

Vatrapu R & Pérez-Quiñones M A (2006) Culture and Usability Evaluation: The Effects of Culture in Structured Interviews. Journal of Usability Studies, 1, (4), pp. 156-170. ISSN: 1931-3357.

Venturi G & Troost J (2004) Survey on the UCD Integration in the Industry. In: Proceedings of the Third Nordic Conference on Human-Computer Interaction, Tampere, pp. 449 – 452. ISBN: 1-58113-857-1.

Vöhringer-Kuhnt T (2002) The Influence of Culture on Usability Paper Draft. [online]. Master's Thesis. Freie Universität Berlin. 20 p. Published 2002 [retrieved 7.4.2010]. Available: http://wise.vub.ac.be/Members/mushtaha/PhD/phd-activity/results.pdf.

Wania C, Atwood M & McCain K (2006) How do Design and Evaluation Interrelate in HCI Research? In: Proceedings of the 6th Conference on Designing Interactive systems, University Park. New York, ACM, pp. 90-98. ISBN: 1-59593-367-0.

Winschiers H & Paterson B (2004) Sustainable Software Development. In: Proceedings of SAICSIT 2004, Stellenbosch. New York, ACM press, pp.274-278. ISBN: 1-58113-982-9.

Wobbrock J (2006) The Future of Mobile Device Research in HCI. In: CHI '06 Workshop Proceedings: What is the Next Generation of Human-Computer Interaction, pp. 131-134.

Appendix 1

Hofstede's dimensions: index scores and ranks for countries and regions (adapted from Hofstede 2001, 500).

	PD		UA		IND/COL		MAS/FEM		LTO/STO	
Country	**Index**	**Rank**	**Index**	**Rank**	**Index**	**Rank**	**Index**	**Rank**	**Index**	**Rank**
Argentina	49	35-36	86	10-15	46	22-23	56	20-21		
Australia	36	41	51	37	90	2	61	16	31	22-24
Austria	11	53	70	24-25	55	18	79	2	31	22-24
Belgium	65	20	94	5-6	75	8	54	22	38	18
Brazil	69	14	76	21-22	38	26-27	49	27	65	6
Canada	39	39	48	41-42	80	4-5	52	24	23	30
Chile	63	24-25	86	10-15	23	38	28	46		
Colombia	67	17	80	20	13	49	64	11-12		
Costa Rica	35	42-44	86	10-15	15	46	21	48-49		
Denmark	18	51	23	51	74	9	16	50	46	10
Ecuador	78	8-9	67	28	8	52	63	13-14		
Finland	33	46	59	31-32	63	17	26	47	41	14
France	68	15-16	86	10-15	71	10-11	43	35-36	39	17
Germany	35	42-44	65	29	67	15	66	9-10	31	22-24
Great Britain	35	42-44	35	47-48	89	3	66	9-10	25	28-29
Greece	60	27-28	112	1	35	30	57	18-19		
Guatemala	95	2-3	101	3	6	53	37	43		
Hong Kong	68	15-16	29	49-50	25	37	57	18-19	96	2
Indonesia	78	8-9	48	41-42	14	47-48	46	30-31		
India	77	10-11	40	45	48	21	56	20-21	61	7
Iran	58	29-30	59	31-32	41	24	43	35-36		
Ireland	28	49	35	47-48	70	12	68	7-8	43	13
Israel	13	52	81	19	54	19	47	29		
Italy	50	34	75	23	76	7	70	4-5	34	19
Jamaica	45	37	13	52	39	25	68	7-8		
Japan	54	33	92	7	46	22-23	95	1	80	4
Korea (S)	60	27-28	85	16-17	18	43	39	41	75	5
Malaysia	104	1	36	46	26	36	50	25-26		
Mexico	81	5-6	82	18	30	32	69	6		
Netherlands	38	40	53	35	80	4-5	14	51	44	11-12
Norway	31	47-48	50	38	69	13	8	52	44	11-12
New Zealand	22	50	49	39-40	79	6	58	17	30	25-26
Pakistan	55	32	70	24-25	14	47-48	50	25-26	0	34
Panama	95	2-3	86	10-15	11	51	44	34		
Peru	64	21-23	87	9	16	45	42	37-38		
Philippines	94	4	44	44	32	31	64	11-12	19	31-32
Portugal	63	24-25	104	2	27	33-35	31	45	30	25-26
South Africa	49	35-36	49	39-40	65	16	63	13-14		
Salvador	66	18-19	94	5-6	19	42	40	40		
Singapore	74	13	8	53	20	39-41	48	28	48	9
Spain	57	31	86	10-15	51	20	42	37-38	19	31-32
Sweden	31	47-48	29	49-50	71	10-11	5	53	33	20
Switzerland	34	45	58	33	68	14	70	4-5	40	15-16
Taiwan	58	29-30	69	26	17	44	45	32-33	87	3
Thailand	64	21-23	64	30	20	39-41	34	44	56	8
Turkey	66	18-19	85	16-17	37	28	45	32-33		
Uruguay	61	26	100	4	36	29	38	42		
United States	40	38	46	43	91	1	62	15	29	27
Venezuela	81	5-6	76	21-22	12	50	73	3		
Yugoslavia	76	12	88	8	27	33-35	21	48-49		
Regions:										
Arab countries	80	7	68	27	38	26-27	53	23		
East Africa	64	21-23	52	36	27	33-35	41	39	25	28-29
West Africa	77	10-11	54	34	20	39-41	46	30-31	16	33

Appendix 2

Description of the Hofstede's dimension profile of the sample.

For power distance, Finland and India represent the lowest and highest score within the sample, Finland scoring 33 and India 77. The highest score in Hofstede's original sample of 53 countries is 104 (Malaysia) and the lowest 11 (Austria). The scores of sample countries in this research represent the different ends of the power distance spectrum relatively well, 7 countries in Hofstede's original sample scored lower than Finland and 9 countries scored higher than India. Canada, alongside with Finland, represents the lower end of power distance. Singapore, on the other hand, scores close to India representing high power distance.

Canada represents the individualistic countries ranking 4th-5th (score 80). For Finland, there is somewhat more collectivistic tendency (score 63, rank 17th). Singapore represents collectivistic countries, ranking 39th-41st, and India scores in the middle of the dimension (48, ranking 21st). The extremes of the dimension are individualistic United States (scoring 91) and collectivistic Guatemala (scoring 6). Thus, individualism-collectivism dimension is well represented in the sample selected for this research.

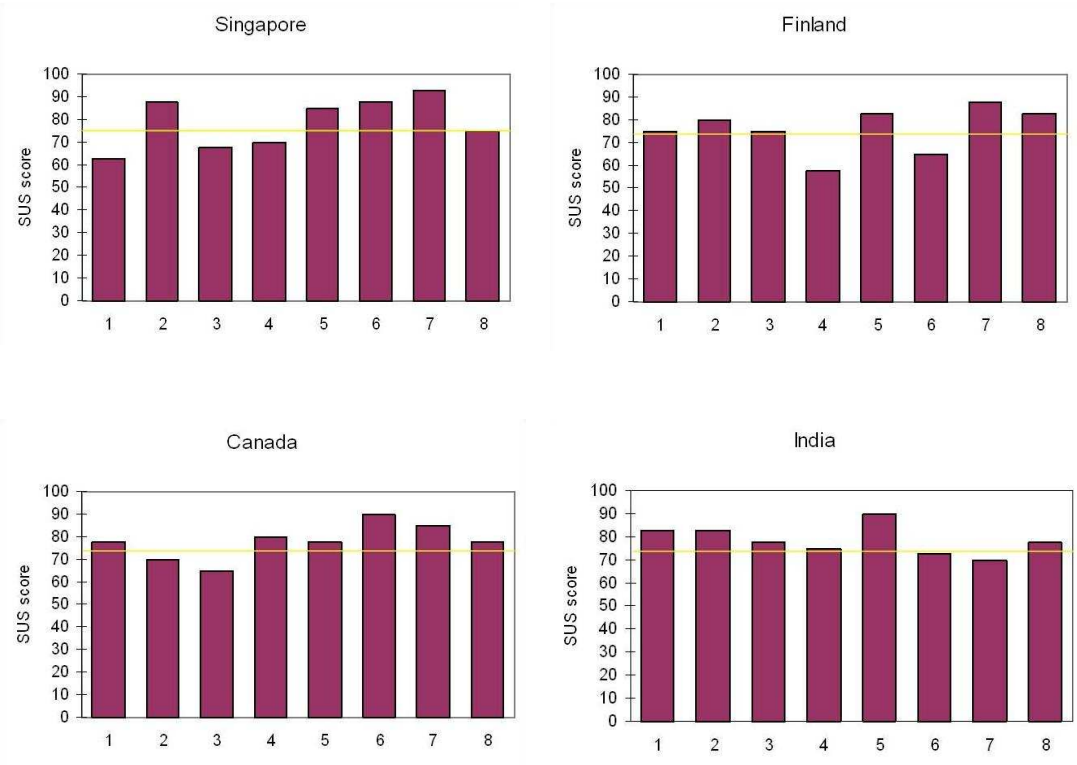
For long-term-short-term orientation the spectrum is relatively well represented. India is the most long-term oriented country in the sample, ranking 7th in the sample of 34 countries (for long-term-short-term orientation the Hofstede's research provides fewer index scores than for the other dimensions), closely followed by Singapore ranking 9th. Canada represents the short-term oriented countries, ranking 30th and Finland falls in the middle of the spectrum ranking 14th.

In terms of uncertainty avoidance, the sample is somewhat biased towards the low uncertainty avoidance countries, Singapore representing the lowest score in Hofstede's sample (rank 53rd) and Finland the highest uncertainty avoidance country in the sample of this research, with the rank 31st-32nd. Canada and India are relatively close to Finland with respect to uncertainty avoidance scores, ranking 41st-42nd and 45th respectively. With respect to masculinity-femininity dimension, the sample selected for the study is somewhat biased towards femininity. India has the highest masculinity score in the sample (56, ranking 20th-21st) and both Singapore and Canada score very close to India (score for Singapore is 48 and rank 28th, score for Canada is 52 and rank 24th). Finland is the most femininity oriented country in the sample, scoring 26 (ranking 47th), which is relatively close to the most femininity oriented country in Hofstede's sample (Sweden scoring 5). The highest masculinity score in Hofstede's research is Japan (scoring 95).

Task list.

1. Let us assume that you have been listening to music with your phone. You paused the music for a while when you ran into a friend of yours. Start listening to the music from where you left off.
2. Pause the music.
3. You cannot remember which Radiohead album you are listening to. Check the details related to the album.
4. You finished the album and want to start listening to something else. Check what other songs you have in the phone.
5. Play one of the songs.
6. You are on the phone with your best friend Susan all the time. Add Susan to your shortcut menu for contacts
7. You cannot remember if you have transferred the photos from your phone to PC. Check how many pictures you have in total on this device.
8. You want to put together a good set of music for your long travel tomorrow. Create a playlist called “Favorites” that includes the songs “Jump”, “Morning bell” and “They don’t want music”
9. Start listening to the playlist.
10. Edit the shortcut contact list so that “Susan” becomes the first shortcut.
11. You have noticed that you do not keep in touch with your friend Chris that much with your phone. Make room for another friend of yours, Jake, in the shortcut menu for contacts and delete Chris form the menu.
12. You want to talk to Jake about a trip you have been planning for tomorrow. Call Jake.
13. Jake does not seem to be answering the phone. Send him a text message saying: “Hi, where are you?”
14. Let us say that you did the trip with Jake yesterday and you took some nice photos. View the three of the latest photos on the phone.

SUS scores for individual participants.



Full list of identified usability problems.

Usability problem	**Singapore**	**Finland**	**Canada**	**India**
User does not notice that MM tiles are vertically scrollable (a list that has collapsed 1-row items and expanded 2-row items in MM loses the vertical scrollability affordance)	-	1	1	-
User assumes that you have to scroll to the vertically first item in MM in order to scroll horizontally in MM	-	-	-	1
User does not notice that MM can be scrolled horizontally	1	2	2	2
User does not know how to scroll MM horizontally	-	1	1	-
User scrolls past the tile she is looking for (correct tile is hard to recognize and distinguish from others)	4	2	1	2
Mental model of quick launcher is faulty: the user does not realize that launching an item in MM takes her to a specific view inside and app and that selecting back is actually back in the app rather than back to the launcher	1	6	7	3
User accidentally clicks the media keys (too close to other keys or not elevated enough from the surface)	7	3	2	2
User accidentally selects the wrong item (select is hard to operate accurately)	3	4	2	5
User perceives buttons hard to press or feel (too much force needed)	5	1	2	1
User has discoverability problems related to HW media keys	-	-	1	-
User repeatedly clicks player controls due to slow response	4	2	1	1
User has difficulty interpreting how 5-way navigation key is mapped to music player controls	5	3	5	2
User does not realize that media keys control the music in now playing regardless what you have highlighted	2	7	2	2
User thinks you have to go to Now playing to control playing	-	2	1	2
User expects contextual options for MM list items	-	2	3	4
User does not understand Now playing item in MM is actionable	2	1	2	3
User has difficulty identifying the music details provided for the music in Now playing	1	2	2	3
User fails to identify the name of the album on the title pane (focuses attention to the content pane)	1	1	4	3
User associates Playlists with all the music on the device	2	1	1	3
User thinks that Go to now playing will start playing the highlighted song	-	1	-	1
User is looking for a function to play album/playlist in albums/playlists view instead of collection opened view	-	3	-	3
User is looking for an option to sort content in Gallery based on different criteria	1	2	-	1
User does not understand that Gallery means all photos	-	-	-	1
User in uncertain of which photos are included in slideshow	-	2	-	-
User assumes that Albums includes all photos on the device	5	2	6	4
User does not realize how to add in phonebook selection list (user does not perceive Select control natural for Add function)	-	-	1	-

Usability problem	Singapore	Finland	Canada	India
User does not understand what is happening in click to add -> select flow (difference between contact lists is not clear)	-	-	1	-
User does not perceive Click to add option to be selectable	1	2	-	2
User thinks Copy to MMC in option means copy contact to multimedia menu contacts	-	2	-	2
User is looking for Add to MM option in phonebook	5	5	-	3
User tries LSK to erase text and interrupts playlist creation (platform familiarity)	1	-	-	2
User does not realize placeholder name for a playlist has to be erased in order to name the list)	2	-	-	-
User accidentally adds the same song to a playlist several times (song added indication is not prominent enough)	6	4	2	-
User thinks that selecting a song can be used to determine whether a song has been added to the playlist	-	1	1	-
User misreads Arrange tiles as arrange titles	-	-	-	4
User selects Arrange tiles to rearrange items on a specific tile	7	7	6	3
User has difficulty exiting Arrange tiles (Cancel functionality missing)	1	3	1	2
User accidentally opens SMS editor with select (not intuitive function for select)	5	4	2	7
User gets looped in the SMS editor after accidentally launching it (Cancel missing)	-	3	4	2
User does not understand grab-drop functionality	-	1	1	-
User fails to click Done after moving contacts (moving as a mode is not intuitive)	1	1	4	-
User doesn't understand term remove in MM contacts context	-	-	2	-
User fails to notice video file indication confusing a video with a photo	6	3	4	5
User looks for call function in options or with selecting (call key is not immediately intuitive in MM even though recognized in PB)	4	1	2	1
User thinks that you can type on the To field in message editor instead of adding a recipient	1	-	-	-
User does not interpret MM as a place for standard communication (call and text)	2	1	-	1
Options on photo full view disappear too quickly	-	2	1	-
User thinks that latest photos can be found by scrolling right instead of left from the last captured photo	-	2	1	-