Methods for usability evaluations of mobile devices

Fayez Alshehri
University of Wollongong, fa980@uowmail.edu.au

Mark Freeman
University of Wollongong, mfreeman@uow.edu.au

Abstract
The usability of applications designed and developed for mobile devices integral user experience, particularly for current touch-screen mobile devices. This conceptual paper provides a profile of the ways in which mobile device usability evaluation methods are being adapted to better reflect technological. With an increasing capacity for greater user interactivity through a range of tasks and in a variety of contexts, the importance and impact of mobile device context-of-use components is of particular interest to researchers. This requires a need for a shift in evaluation method paradigms with a particular focus on methods involving the user, as part of the overall development process. Having a better understanding of the context-of-use may contribute to improvements in mobile device usability and the overall user experience.

Keywords
methods, usability, evaluations, mobile, devices

Disciplines
Engineering | Science and Technology Studies

Publication Details
Alshehri, F. & Freeman, M. (2012). Methods for usability evaluations of mobile devices. In J. W. Lamp (Eds.), 23rd Australian Conference on Information Systems (pp. 1-10). Geelong: Deakin University.

Methods of usability evaluations of mobile devices

Fayez Alshehri
Mark Freeman
Social Technical Information Systems Research Centre
School of Information Systems and Technology
University of Wollongong
New South Wales, Australia
Email: fa980@uowmail.edu.au
Email: mfreeman@uow.edu.au

Abstract

The usability of applications designed and developed for mobile devices integral user experience, particularly for current touch-screen mobile devices. This conceptual paper provides a profile of the ways in which mobile device usability evaluation methods are being adapted to better reflect technological. With an increasing capacity for greater user interactivity through a range of tasks and in a variety of contexts, the importance and impact of mobile device context-of-use components is of particular interest to researchers. This requires a need for a shift in evaluation method paradigms with a particular focus on methods involving the user, as part of the overall development process. Having a better understanding of the context-of-use may contribute to improvements in mobile device usability and the overall user experience.

Keywords

Human-Computer Interaction, Mobile Devices, usability, development process

INTRODUCTION

Rapid advancements in the design and development of mobile devices have a significant impact on the nature of user interaction, as they can be used at any time in any location. An anticipated outcome of increased diffusion of mobile devices is the increasing interest by researchers and mobile interface device designers on the implications of mobile human-computer interaction (mHCI) with regard to how novel interfaces should be designed as part of the development process. Of current interest is improving the user interfaces of touch-screen mobile devices by focusing on context-of-use factors. These contextual factors include: task, environment; social and technological components; and their impact on user performance and behaviour. This conceptual paper will briefly explore the evolution of HCI studies over recent decades in relation to the increasing awareness of the differences in user interactions stationary computers and mobile devices and what issues need to be taken into consideration when testing interfaces as part of the design stage of the development process. In addition, a discussion will be developed on the paradigm shift taking place in HCI research towards touch-screen mobile devices in relation to both task and contextual factors - mHCI. Additionally, the implications for the development of effective interfaces for touch-screen mobile devices and the users' response to the interfaces will also be explored.

HUMAN-COMPUTER INTERACTION (HCI)

Human-Computer Interaction (HCI) was recently redefined by Te'eni et al. (2007) as *a discipline that attempts to understand and establish how people interact with computers in an organisational context*. The central element of HCI is to identify usability needs and improve the user's interfaces of task-oriented applications in the work context (Chen 2001). Therefore, HCI may also be defined as a study of human factors and machine aspects in the interaction process. Smith (1997) stated that interactive design also requires an appropriate interactive style based on the user's knowledge and tasks. Factors that help determine the interaction style are initiation, flexibility, option complexity, and information load, and interaction styles may differ according to individual users (Smith 1997). Carroll (2010) argues that HCI as a research and practice field began to develop during the early 1980s. According to the author, what started as a specialty area in computer science rapidly expanded over the following three decades into a field of study that includes professionals from a range of disciplines and which incorporates diverse concepts and approaches (Carroll 2010, p. 3). As noted by Jaimes and Sebe (2007), the primary aims of research into HCI is to gain deeper insight into how computer technology can be made more usable by people. Therefore, HCI research can broaden our understanding of how effective HCI can be achieved through high quality interface design that is not only usable and performs well, but which incorporates human elements such as the physical and social experiences into the interaction.

One of the most significant factors to impact the concept and practice of HCI during the early 1980s was the emergence of the Personal Computer (PC). This marked the beginning of an era that could potentially see all people in developed nations being a computer user (Carroll 2010). In turn, as reported by Carroll (2010), this then led to the emergence of HCI studies to address the usability deficiencies of computers for people seeking to use computers as tools. Thus, during the 1980s HCI research was primarily focused on user interactions with office automation programs including word processing, databases and statistical software, before also including graphical user interfaces, during the early 1990s (Lazar et al. 2010). HCI evaluations need to be reframed as an integral part of the ISD process and thus it is important to consider this issue as part of the diverse nature of the ISD landscape (Isomäki and Pekkola 2011).

HCI AND MOBILE DEVICES

As stated by Chi (2009, p. 1), HCI as a field of research and study has progressed far beyond the evaluation setting of "a single user sitting in front of a single desktop computer". Traditional HCI researchers have considered new paradigms that recognize mobile computation and the way that it is impacted by environmental and context-of-use factors (Chi 2009). Thus, the challenge for HCI researchers is to design social and mobile application experiments that are "ecologically valid" (Chi 2009, p. 5). The importance of the shift in HCI paradigms towards human interaction with mobile device interfaces is also emphasized by Jaimes and Sebe (2007) who argued that the increased availability of mobile devices as a result of lower hardware costs and the fact that they are becoming more powerful and complex in relation to their computational capacity. The implication of the transition from stationary computer to mobile device and the ongoing developments in user interfaces is best reflected in the statement by Shackel (cited in Grudin 2009, p. 371) who stated that the core aspect of computer use has shifted ''from system supremacy to personal empowerment". This is evidenced by the evolution of HCI research, and it is factors such as these that have led to the need for more comprehensive research into user interface interactions (Jaimes and Sebe 2007).

Usability evaluation

Evaluation is defined by Preece et al. (2002) as the process of data collection relating to the users' perceptions of a product in relation to a specific use in a particular environment. Evaluation is vital to both the development of a product and its perceived usability as it underpins the direction for product development, the design process, and the perceived outcomes. Evaluation methods used to assess the usability of computer applications have long been an integral aspect of HCI research (Billi et al. 2010; Hornbæk 2010). The aim of the usability test is to gain a third-party assessment of user characteristics and to gauge how effectively and efficiently a user is able to view content or perform a task on a specified device (Blandford 2004; Zhang and Adipat 2005; Blandford et al. 2008; Inzunza et al. 2011). By carrying out such analysis Inzunza et al. (2011) asserted that the usability aspects of the software architecture can be better visualized prior to the implementation phase. In turn, this can then help to determine three core components of the user interface: components required by the task; ideal components for the user; and the specified component to be implemented.

For a usability test to be effective it must be able to elicit feedback from users about the degree of difficulty and the level of user satisfaction in using the application, as well as to evaluate task performance levels achieved by users (Zhang and Adipat 2005). Nielsen and Mollich (1990, cited in de Kock et al. 2009) asserted that UEMs may be categorised into four basic user interface evaluation methods: formal analysis techniques; automatic computerized procedures; empirical performance tests of users; and heuristically, and these classifications are still widely accepted today.

While the terms 'usability' and 'user experience' are sometimes used interchangeably in the literature, researchers such as Heo et al (2009, p. 263) draw a distinction between the two terms. According Heo et al., usability is a much narrower concept than user experience in that it typically just refers to the ease of use of the design features of the device. User experience, on the other hand, is defined by the authors as including the user's thoughts and feelings about their interaction with the device as well (Heo et al. 2009). Furthermore, Petrie and Bevan (2009), asserted that drawing this distinction between the two terms is valid because the users of many new technologies are not necessarily seeking to achieve a task through their interaction with the device, but may simply be engaged with the device to amuse and entertain themselves.

Irrespective of the purpose of the engagement by the user, one of the main goals of evaluating a device is to determine whether the problems with the design are general, specific or task-related (Blandford et al. 2008). Different types of usability evaluations will elicit different problems (Biel et al. 2010), and while some UEMs are suitable to one or more stages of the design lifecycle, to understand their different functions, the literature reveals that a number of researchers have divide them into three different groups: test, inspection and inquiry. UEMs that use testing manipulate representative users to work on archetypal tasks on the device or a prototype and the user interface and their performance is measured or assessed (Heo et al. 2009). UEMs based on inspections usually involve HCI experts (Thimbleby 2007) and can be accomplished at any stage of the design, from prototype to marketplace. UEMs based on inquiry should include the users' preference desires and behaviour and try to constitute the requirements of a design (Thimbleby 2007; Heo et al. 2009). It is established within the literature that two of the most common evaluation methods utilised during the product design stages are empirical user-testing and expert inspection (Blandford et al. 2008; Petrie and Bevan 2009; Moritz and Meinel 2010). As pointed out by de Kock et al. (2009) both of these UEMs vary from formal quantitative experiments including large sample sizes and complex testing procedures to informal qualitative studies that involve only a small number of participants.

A review of the literature reveals that the evaluation processes usually include a number of UEMs that may focus on device characteristics, user knowledge (Moritz and Meinel 2010), or the test procedures themselves. For instance, according to Kjeldskov and Graham (2003) , the six main areas of concern for usability testing are the:

- Skills and/or technological competence of the user
- Costs associated with data collection (participation fees, etc.)
- The presence of researcher changing the phenomenon of interest
- Time and/or personnel required (of the user and tester)
- Inability to completely control experimental variables
- Observations that do not generalise

Zhang and Adipat (2005, pp. 302-303) contributed further to the discussion by suggesting that the usability attributes most often evaluated are learnability, efficiency, memorability, user errors, user satisfaction, effectiveness, simplicity, comprehensibility and learning performance. However, more recently, Petrie and Bevan (2009) have claimed that many practitioners consider flexibility, learnability, memorability and safety to be the main elements of usability, and, as such, are the focus of evaluation procedures.

A range of user-based UEMs are available including scenario-based tasks, interviews, focus groups, and think-aloud sessions (Collins et al. 2010, p. 8). These methods generally require users to perform realistic tasks which the device is designed to support in realistic situations (Petrie and Bevan 2009). The size of the evaluation sample and the tasks that they are required to perform as part of the evaluation process will depend on the whether the purpose of the evaluation is formative or summative in nature (Petrie and Bevan 2009). Importantly, by employing user oriented methods of evaluation it is possible to overcome the limitations of expert simulation evaluations that can never quite replicate the exact nature of user behaviours (Moritz and Meinel 2010). Moreover, Jones and Marsden (2006) and Cooper et al. (2007) argued that it is possible to use in-situ interviews, instant contextual inquiry and observations of the naturalistic behaviours to enrich the design process by identifying the requirements of the user in the initial stages of the design cycle. Therefore, although such methods are more time and resource intense than the expert inspection based alternatives, they are able to deliver data that more accurately reflects the real opinions, problems and proposals presented by the users (Moritz and Meinel 2010).

MOBILE DEVICES

Mobile devices can typically be categorised into four types: unintelligent gadgets, cellular phones (mobile phones), smart phones, and devices with operating systems (Mirza 2008). Users may access distributed services anywhere, anytime using mobile devices that adapt to the users' situations. The aim of pervasive computing is simply to make the applications a part of the environment of the user who interacts with the computer without knowing that they are interested and interacting with the device. As a result, mobile application developers have to take many mobile device characteristics into consideration.

The focus of this paper is on improving usability of touch screen mobile devices through an enhanced understanding of the current situation, and, as a result, particular consideration is given to the distinctive mobile device characteristics. It is well established in the literature that the context-of-use of mobile devices differs from the traditional stationary computer setting in that it is highly dynamic and often involves a range of different software applications (Kjeldskov and Stage 2004; Zhang and Adipat 2005). On the basis of these differences, mobile device usability is increasingly a focus of conventional HCI research and seeks to identify issues such as navigation of complex data on small screens (Björk et al. 2000); tactile feedback (Brewster et al. 2007); mobile usability assessment and texting (MacKenzie and Soukoreff 2002); multiple device handling (Oulasvirta and Sumari 2007); and theoretical frameworks of mobile usability (Ham et al. 2006). Yet, although mobile device context-of-use is not specifically the focus of HCI studies, it nonetheless operates as a backdrop to such studies through its potential impact on the user-device interaction process (Isomäki and Pekkola 2011). It may be noted at this point that to evaluate the usability of a mobile application a functional prototype or the final product is needed (Biel et al. 2010, p. 2077).

Recent research into mobile device usability evaluation is increasingly being focused on the contextual variables impacting the HCI from a holistic perspective (Bernhaupt et al. 2008). Context is part of the user's subjective experience and, as identified by Jumisko-Pyykko and Vainio (2011), central to the exploration of context in mobile device usability evaluations is the question of relevancy (Isomäki and Pekkola 2011). One of the key benefits of real-context user-based UEMs is that interaction between the user and the interface in real-life contexts provides the designer with the most accurate assessment of the mobile device. Context in relation to mHCI is thought to include task, physical, social, temporal and technical components (Isomäki and Pekkola 2011). In broader terms: task context refers to the performance of the actual task and any interruptions that may occur; physical context refers to the sensed circumstances of the situation; social context refers to the presence of other people and their impact; temporal context refers to factors pertaining to past or future experiences; and the technical context refers to the device infrastructure and the connectivity issues (Isomäki and Pekkola 2011).

In relation to these components, many researchers have argued that designers should pay more attention to environmental and contextual evaluations on the usability of mobile devices. Dunlop and Brewster (2002) argued that one of the most challenging aspects facing mobile HCI investigators and designers is mobility, while Pascoe et al. (2000) argued that the development of minimal attention user interfaces needs to be combined with complex contextual demands. Furthermore, Preece et al. (2002) emphasised that when considering the usability of the design it is important to consider when, where and who is going to use the system. In addition, the authors stressed the importance of understanding the activities people do when interacting with the product as this helps determine the type of activities the systems needs to support and optimise the users' interactions with the system.

Elements such as the use of the mobile device in a range of different contexts and situations, the possibility of group interaction, the need for privacy and security, achieving intimacy and the availability of the mobile device have the potential to significantly impact on the user-device interaction flow and so require more specific evaluation methods (Avouris et al. 2008; Billi et al. 2010). Moreover, the specific characteristics of mobile devices such as "small screen, limited input capabilities, limited and costly bandwidth, limited connectivity, limited computational resources, limited power (batteries), and wide heterogeneity inherently" pose various limitations on accessibility and usability (de Sá and Carriço 2008; Billi et al. 2010, p. 339). Moreover, in contrast to the typical context of general computer use, mobile device contexts can include transitions between and within contexts, and this also contributes to the heterogeneous and dynamic context-of-use (Isomäki and Pekkola 2011). Jumisko-Pyykko and Vainio (2011) identify that the dynamic nature of the mobile device context-of-use is based upon the random and highly contingent interrelation between a range of contextual components and is worthy of further study.

However, one of the things apparent from a review of the literature is that evaluation of usability attributes is often used extensively as measures of usability in PC applications. This reveals a tendency towards evaluation methods that primarily focus on the device rather than aspects of the context-of-use (Avouris et al. 2008). This finding is also affirmed by Billi et al. (2010) who also pointed to the issue of traditional HCI UEMs relying on measurements of task performance and task efficiency as the pathway towards evaluating applications. Billi et al. (2010) questioned the validity of measurements of task performance and task efficiency as an evaluation approach given the trend away from structured tasks and predictable HCI settings typically associated with PC activities to the more unstructured and unpredictable user interactions with mobile devices. Furthermore, Zhang and Adipat (2005) also discussed the challenges involved in trying to design UEMs that address all or most of the various contexts-of-use for mobile devices. As the researchers pointed out, the primary difficulties lie in trying to foresee all of the different environmental variables possible while using an application including such things as body movement, visibility, time to accomplish the task, level and type of distraction. Due to these fundamental differences between the device types that a new set of usability paradigms for mobile devices should emerge (de Sá and Carriço 2008).

Yet, as is discussed in the literature, when addressing the different evaluation methods for mobile devices when compared to PC technology, the challenge may not be so much to create a new set of measures, but rather to adapt the existing measures to more accurately reflect the complexity of the context-of-use aspects relating to mobile devices (Avouris et al. 2008; Billi et al. 2010). Billi et al. (2010) explored in detail the many challenges facing mobile devices designers in relation to the device's accessibility and usability, highlighting the points of difference that are sometimes established between accessibility and usability. Notwithstanding these differences, however, the authors choose to look at the concepts of accessibility and usability as a whole on the basis that there is a significant degree of mutual dependence and influence between the two elements.

Furthermore, in their discussion of UEMs specific to mobile devices, Heo et al. (2009) applied the usability test, inquiry and inspection evaluation categorisations and proposed an evaluation framework based on a multilevel, hierarchical model of usability factors that supports both a task-based and an interface-based approach.

Moreover, on the basis of these categorisations, it is evident that both laboratory settings (usability testing and usability inspection) and field settings (usability inquiry) may be utilised as part of the evaluation procedures.

Billi et al. (2010, p. 338) drew attention to the concept of 'universal access'. According to the authors, universal access is a reference to three key elements: the global requirement of catering to diverse characteristics in the target user population; the scope and nature of tasks; and the different contexts-of-use and their effects. Using the phrase coined by Stephanidis and Savidis (2001; cited in Billi et al. 2010, p. 338) 'anyone, anytime, anywhere access', the authors provided a discussion of the important role that early assessment plays in steering the development of the design and implementation process towards the achievement of these accessibility goals. Accessibility and usability evaluation methodologies are vital to the early development process of mobile devices as they help to define the specifics of the product (Billi et al. 2010). Balagtas-Fernandez et al. (2009) focused specifically on the methods used to evaluate mobile device touch screens. The design of the device interface is vital for the success of both the device and the applications. The evaluation method profiled by Balagtas-Fernandez et al. (2009) was laboratory-based user participation, and acknowledged that previous studies conducted on mobile devices in relation to their specific characteristics such as interface use of hand-held devices and effective use of the small size screen typical of such devices. However, the study by these researchers added to the overall understanding of usability attributes of mobile devices by focusing on three usability input components: the overall "interface layout, information input and menu accessibility". Specifically, a comparison of layout for scrollable view versus a tabbed view, direct keyboard input versus touch screen input, and "menu accessibility through the device menu versus the context menu" (Balagtas-Fernandez et al. 2009, pp. 243-244).

User-based evaluation of mobile device usability

Bernhaupt et al. (2008) emphasised that the primary objective of the evaluation process is to produce data upon which to determine usability solutions. To understand users' usage behaviours of mobile technology, usability testing and evaluation methods are undertaken. Such usability evaluation methods seek to ensure that users can operate the mobile technology efficiently, effectively, and satisfactorily.

Bastien (2010) stated that user-based usability evaluations methods invite actual users to perform tasks with a product or to explore it applications while their behaviours are observed and their thoughts are recorded, this information can then be used if future iterations of the development process. Aspects under evaluation are often the time required to complete a task, task-completion rates, and number and types of errors. Subsequent to the identification of the design flaws recommendations are then made to improve the ergonomic quality of the product (Bastien 2010). Biel et al. (2010) noted that there are many mobile usability evaluations constructed as 'user tests' which are designed to identify problems with the interaction between a human and an application's interface, yet more research needs to be done into the mobile context factors and the technology that supports the interaction. The authors identify the relevant mobile context factors as "environment, user, task, and technology" (Biel et al. 2010, p. 2032). Bertini et al.'s (2006) discussion of user-testing UEMs for mobile devices asserted that there are certain requirements that need to be fulfilled for this type of testing method to be effective. Specifically, they asserted that at the very least there is a need for the evaluation method to include real users, real or simulated contexts-of-use and real tasks with real devices (Bertini et al. 2006). The main issue cited within the literature in relation to user-based evaluation methods of mobile devices is the potential for the evaluation procedure to provide reliable and valid data on usability issues applicable to natural user interactions with the device in real contexts-of-use. While the adaption of the laboratory setting can go some way to recreating real contexts-of-use, Bastien (2010) also discussed the need to include additional strategies such as user diaries, log books traces and interviews to monitor and evaluate user behaviours and perceptions.

LABORATORY VERSUS FIELD-BASED SETTINGS

There are a number of challenges that currently face evaluation of the mHCI process for mobile devices; one of these key challenges is whether to conduct the testing as part of the development process in a laboratory or in the field. In addition to the great variability in the types of users and uses of the devices themselves, there is the issue of the great variability in the types of environments in which the devices are being used (Biel et al. 2010). In fact, one of the most prominent challenges to evaluating the usability of mobile devices identified in the literature is testing usability under 'real-life' conditions. Crossan et al. (2007) identified the difficulties are in conducting "realistic" tests without constraining the users in unnatural ways.

The literature identifies that the most common method used to evaluate mHCI has been through the process of experiments and assessments conducted in laboratory settings (Kjeldskov and Stage 2004; Bernhaupt et al. 2008). Kjeldskov and Graham (2003) noted that laboratory settings were used to evaluate mobile devices up to 71% of the time. According to Biel et al. (2010), the prevalence of the laboratory setting is because of the close connection between HCI and the computer sciences. The use of such settings to conduct mHCI has traditionally involved the use of particular techniques and strategies to try to address the challenges posed by the nature of the mobile device in its specific context-of-use (Kjeldskov and Stage 2004). With regard to the advantages of UEMs conducted within a laboratory setting, Zhang and Adipat (2005) highlighted that such evaluations are generally conducted by expert technicians and this environment provides them with greater potential for total control over the experiment. For instance, the expert evaluator is able to set the tasks and procedures to match the aim of the study, as well as to ensure the user follows the necessary instructions (Zhang and Adipat 2005). In addition, the results of laboratory-based evaluation procedures can be more accurately interpreted as all of the irrelevant or unwanted variables can be eliminated from the evaluation process (Kaikkonen et al. 2005). Finally, an audio-visual recording of the evaluation procedure is able to be easily attained providing the expert evaluator with further opportunity to gather data on the user actions and reactions during the evaluation process (Goodman et al. 2004; Kaikkonen et al. 2005). Kjeldskov and Stage (2004) also drew attention to a number of benefits resulting from conducting usability evaluations in a laboratory setting. In particular, the authors pointed to the higher degree of experimental control and collection of high quality data as the primary benefits. This view is affirmed by Avouris et al. (2008) who stated that laboratory settings that can provide highly valuable quantitative data. As such, the authors recommended the use of laboratory settings for the evaluation of such aspects as time to complete the task, success or failure rates, number of attempts, perception metrics such as the user's feeling about the device after use, user satisfaction and perceived costs (Avouris et al. 2008).

However, the limitations apparent to the more traditional or laboratory-based evaluation methods often demand that they be combined with field-based evaluation methods so as to get the best insight into the user-centred design and development process (Biel et al. 2010). For instance, Tamminen et al. (2004) argued that laboratory tests do not take into account environmental pressures which may have an effect on the performance of the users such as noise, interruptions, movement, and multitasking. This view is affirmed by (Coursaris et al. 2007) who stated that controlled laboratory studies do help to promote experimental rigor, but the fact that usability can be significantly affected by context-of-use must not be forgotten.

The use of field-based mobile device usability evaluations is important to gain a better understanding of HCI issues. Avouris et al. (2008) discuss the range of methodologies which are increasingly being linked to the evaluation of mobile devices. According to the authors, the use of scenario-based methods, persona creation and the application of performance techniques are just a few of the methods gaining increasing attention (Avouris et al. 2008). Specific scenario-based evaluation methods are discussed including those that focus on "locations and settings"; "movement and posture"; "workloads, distractions and activities"; "devices and usages"; and "users and personas" (Avouris et al. 2008, p. 39). In turn, the authors suggested that controlled field-based evaluations are far more useful for gathering qualitative data through the measurement of such aspects as user performance, identification of particularly hard points or tasks, perceived workload, and user satisfaction. Moreover, as asserted by Zhang and Adipat (2005, p. 298), one of the main advantages to conducting user evaluations in the field is that they can take into account connectivity issues as a result of "changing and unreliable network conditions and other environmental factors".

What is apparent from a review of the literature is the continuing debate that exists on whether there is a greater need to employ more field-based UEMs on mobile devices (as opposed to laboratory based methods) in order to more accurately reflect their context-of-use (Zhang and Adipat 2005; Billi et al. 2010). For instance, Kjeldskov and Stage (2004) argued that the importance placed on field-based methods is somewhat misguided. In particular, the researchers discuss the complications that arise in trying to recreate "realistic environments", the merits of traditional observation and think-aloud evaluation methods that are sometimes overlooked, and the way in which reliable and accurate data collection processes can be complicated as a result of the unknown variables in a real-life context potentially affecting the evaluation data (Kjeldskov and Stage 2004, p. 600). Zhang and Adipat (2005) asserted that the main disadvantage of field-based UEMs is their incapacity to provide sufficient control over the participants and the conditions in the evaluation. Even though this method of evaluation purports to be done in a 'natural' environment, there is still the difficulty of trying to generalise findings from the evaluation given that particular conditions were present in the evaluation process. Furthermore, traditional data gathering techniques including observation and verbal protocol are difficult to apply given the changing dynamics of the evaluation environment (Zhang and Adipat 2005, p. 300).

Billi et al. (2010, p. 340) recognised that there is a degree of importance to creating "real-world" settings in which to conduct usability testing. However, the authors then went on to state that due to the difficulties associated with gathering data using field-based tests, expert-based traditional heuristic evaluations remain the most effective way to evaluate mobile device usability (Billi et al. 2010). In addition, Fiotakis et al. (2009) recognised the claims made by some researchers against the value of performing field-based evaluations due to the resources required. However, it is worth noting that they went on to argue that the increasing sophistication of field-based evaluation methods means that they are slowly beginning to be valued more highly (Fiotakis et al. 2009). In addition, de Sá & Carriço (2008) claimed that field-based usability evaluations during the early stages of the design process are recognised as important to the process and can yield some effective results, yet, the evaluation methods still rely too heavily on simulations and role playing within controlled scenarios. Finally, in support of field-based UEMs, Zhang and Adipat (2005) argued that the main advantage is that they are able to take into account the dynamic nature of the real mobile context and the unpredictable nature of wireless networks. Moreover, they claimed that such studies more accurately reflect the user experience and so data gained on the subjective aspects of usability are more reliable (Zhang and Adipat 2005). It is interesting to note, however, that Zhang and Adipat (2005) asserted that both controlled laboratory evaluation methods and those conducted in the field are limited in the way that they are able to accurately reflect a mobile environment or in their lack of adequate procedural control.

The choice of whether to use a laboratory or field setting to conduct mobile device user evaluations will depend on the research objectives (Zhang and Adipat 2005; Avouris et al. 2008). Therefore, as one of the research objectives of this study is to evaluate appropriate UEMs for use with mobile touch screen devices in different contexts. To achieve the research objective this study will conduct usability evaluations within a laboratory setting. This is because the evaluation of touch screen usability in this study focuses on a component of mobile applications that is not highly "impacted by mobility, network connectivity, and other contextual factors" which include such aspects of interface layout, data entry methods, information presentation schemes and link structures (Zhang and Adipat 2005, p. 301). Moreover, although Zhang and Adipat (2005, p. 301) have also stated that field studies are more appropriate for the evaluation of user experience, it is the aim of this study to include such evaluations within a laboratory setting.

Results from the Kjeldskov and Stage (2004) study claim that laboratory-based and field-based UEMs yield similar results in relation to user performance; whereas the techniques demonstrated significant differences in relation to perceived effort and overall workload. A similar finding was also presented by Kaikkonen et al. (2005) after their comparison study of laboratory-testing and field-testing of 20 participants using the think-aloud protocol in both environments. The researchers concluded that field-testing does not add significantly to the validity or thoroughness of the user interface evaluation of mobile applications and devices compared to laboratory based testing (Kaikkonen et al. 2005).

Is widely acknowledged within the literature is that expert inspection methods are the preferred UEMs by product manufacturers to test their product prior to market exposure and are generally most useful during the earlier stages of the development cycle (Fiotakis et al. 2009; Hwang and Salvendy 2010); whereas empirical-based user-testing, while useful during any stage of the design process, is certainly an essential component in the later stages of the design process (Blandford et al. 2008; de Kock et al. 2009; Petrie and Bevan 2009; Lee et al. 2011). This is also true for the usability evaluation of applications for mobile devices during their development process.

CONCLUSION

Currently, there is a growing body of literature on the subjective assessment of many of the usability aspects of mobile devices. Many of these issues have been discussed in this paper to provide the reader with a greater understanding of the current situation. However, there are still many usability aspects for which there are no objective measurements. One of the most significant gaps in the literature on the usability of mobile devices is in relation to the methods used to evaluate usability attributes. As a result, there is a lack of reflection and subsequent evaluation of the methodologies being used in the evaluation process (Bernhaupt et al. 2008). Furthermore, of the few studies that have been conducted focusing on mHCI, there is strong evidence to suggest the tendency is towards environment independent and artificial settings for the evaluation process. That is 'real-use' contexts and natural setting evaluations have traditionally been passed over in preference for artificial or laboratory based settings (Bernhaupt et al. 2008). This is significant because Biel et al. (2010) also stressed the importance of considering the mobile usage context when designing and evaluating mobile user-system interaction. To optimize the development of mobile user applications it is necessary to combine an analysis of the software architecture of the mobile devise with the usability evaluation. Carroll (2000) argued that the true story of those performing the task is revealed through scenario testing, and that such an approach is also a useful method for gaining a holistic insight into the design elements of the technology. In particular, the backgrounds of the people and the social and physical settings where the technology will be used can be more readily associated with the process of design.

This paper contributes to the growing understanding of the impact of contextual components of mobile device usability, with advances to the technological capacity of mobile devices taking place and user interactivity with the device becoming more sophisticated across a range of different contexts. As a result, the focus of HCI research has necessarily become more concentrated on mobile device context-of-use components and their impact on touch screen mobile device usability - mHCI. This conceptual paper has explored the paradigm shift in relation to the development of UEMs more suited to the specific nature of mobile device technology. In particular, the focus that is now being assigned to the ways in which evaluation can be used either in a laboratory environment or 'real-world' contexts and the trade-offs of each approach. What has been established is that there is increasing pressure on mobile device UEMs to become more relevant to the context-of-use considerations. This includes a greater focus on the interrelationship between such aspects as task, physical environment, social environment, temporal elements, and the technical components of touch screen mobile devices. Through an enhanced understanding of these issues, both practitioners and researchers can make informed recommendations and decisions to improve the usability or touch screen mobile devices.

REFERENCES

Avouris, N., Fiotakis, G. and Raptis, D. 2008. "On measuring usability of mobile applications," In . E. L-C. Law, N. Bevan, G. Christou, M. Springett & M. Lárusdóttir (eds), *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM),* Reykjavik, Iceland, pp. 38-44.

Balagtas-Fernandez, F., Forrai, J. and Hussmann, H. 2009. "Evaluation of user interface design and input methods for applications on mobile touch screen devices," *in Proceedings of the 12th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT), pp. 243–246.*

Bastien, J. 2010. "Usability testing: a review of some methodological and technical aspects of the method," *International Journal of Medical Informatics* (79:4), pp e18-e23.

Bernhaupt, R., Mihalic, K. and Obrist, M. 2008. "Usability Evaluation Methods for Mobile Applications," *Handbook of research on user interface design and evaluation for mobile technology* (2), pp 745-758.

Bertini, E., Gabrielli, S., Kimani, S., Catarci, T. and Santucci, G. 2006. "Appropriating and assessing heuristics for mobile computing," *in Proceedings of the International Conference in Advanced Visual Interfaces*, pp 119-126.

Biel, B., Grill, T. and Gruhn, V. 2010. "Exploring the benefits of the combination of a software architecture analysis and a usability evaluation of a mobile application," *Journal of Systems and Software* (83:11), pp 2031-2044.

Billi, M., Burzagli, L., Catarci, T., Santucci, G., Bertini, E., Gabbanini, F. and Palchetti, E. 2010. "A unified methodology for the evaluation of accessibility and usability of mobile applications," *Universal Access in the Information Society* (9:4), pp 337-356.

Björk, S., Redström, J., Ljungstrand, P. and Holmquist, L. 2000. "POWERVIEW Using information links and information views to navigate and visualize information on small displays," London: Springer.

Blandford, A. 2004. "Understanding Users' Experiences: Evaluation of Digital Libraries," *in Proceedings of the Delos workshop on evaluation of Digital libraries,* Padova, Italy, pp 31-34.

Blandford, A. E., Hyde, J. K., Green, T. R. G. and Connell, I. 2008. "Scoping analytical usability evaluation methods: a case study," *Human–Computer Interaction* (23:3), pp 278-327.

Brewster, S., Chohan, F. and Brown, L. 2007. "Tactile feedback for mobile interactions," In Proceedings of the SIGCHI conference on Human factors in computing systems, San Jose, CA, pp 159-162.

Carroll, J. M. 2010. "Conceptualizing a possible discipline of human-computer interaction," *Interacting with Computers* (22:1), pp 3-12.

Chen, Q. 2001. *Human computer interaction: issues and challenges,* Hershey: IGI Global.

Chi, E. 2009. "A position paper on'living laboratories': Rethinking ecological designs and experimentation in human-computer interaction," *Human-Computer Interaction. New Trends*, pp 597-605.

Collins, K., Kanev, K. and Kapralos, B. 2010. "Using games as a method of evaluation of usability and user experience in human-computer interaction design." *Proceedings of the 13th International Conference on Humans and Computers,* Aizu-Wakamatsu, Japan.

Cooper, A., Reimann, R. and Cronin, D. 2007. *About Face 3: The Essentials of User Interface Design* Indianapolis: Wiley.

Coursaris, C. K., Hassanein, K., Head, M. and Bontis, N. 2007. "The Impact of Distractions on the Usability and the Adoption of Mobile Devices for Wireless Data Services." *In Proceedings European Conference on Information Systems, Paper 28.*

Crossan, A., Murray-Smith, R., Brewster, S. and Musizza, B. 2007. "Instrumented usability analysis for mobile devices," *Handbook of Mobile HCI,* Hershey: IGI Global.

de Kock, E., van Biljon, J. and Pretorius, M. 2009. "Usability evaluation methods: Mind the gaps." *in Proceedings of Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists Vanderbijlpark,* Emfuleni, South Africa*, pp. 122-131*.

de Sá, M. and Carriço, L. 2008. "Lessons from early stages design of mobile applications," MobileHCI, ACM Press, New York. pp. 127-136.

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

Fiotakis, G., Raptis, D. and Avouris, N. 2009. "Considering cost in usability evaluation of mobile applications: Who, where and when," *in Proceedings of Interact, International Conference on Human-Computer Interaction,* Uppsala, Sweden.

Goodman, J., Brewster, S. and Gray, P. 2004. "Using Field Experiments to Evaluate Mobile Guides." *Mobile HCI 2004*, Glasgow, UK: Springer.

Grudin, J. 2009. "Brian Shackel's contribution to the written history of Human–Computer Interaction," *Interacting with Computers* (21:5), pp 370-374.

Ham, D., Heo, J., Fossick, P., Wong, W., Park, S., Song, C. and Bradley, M. 2006. "Conceptual framework and models for identifying and organizing usability impact factors of mobile phones," ACM.

Heo, J., Ham, D., Park, S., Song, C. and Yoon, W. 2009. "A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors," *Interacting with Computers* (21:4), pp 263-275.

Hornbæk, K. 2010. "Dogmas in the assessment of usability evaluation methods," *Behaviour & Information Technology* (29:1), pp 97-111.

Hwang, W. and Salvendy, G. 2010. "Number of people required for usability evaluation: the 10±2 rule," *Communications of the ACM* (53:5), pp 130-133.

Inzunza, S., Juarez-Ramirez, R., Mejia, A. and Gomez-Ruelas, M. 2011. "Implementing user-oriented interfaces: From user analysis to framework's components," International Conference on Uncertainty Reasoning and Knowledge Engineering: IEEE.

Isomäki, H. and Pekkola, S. 2011. Introduction: Reframing Humans and Information Systems. *Reframing Humans in Information Systems Development*. H. Isomäki and S. Pekkola. (Ed.) London, Springer-Verlag**,** 1-14.

Jaimes, A. and Sebe, N. 2007. "Multimodal human-computer interaction: A survey," *Computer Vision and Image Understanding* (108:1-2), pp 116-134.

Jones, M. and Marsden, G. 2006. *Mobile interaction design*. New York: John Wiley & Sons.

Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T. and Kankainen, A. 2005. "Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing," *Journal of Usability Studies* (1:1), pp 4-16.

Kjeldskov, J. and Graham, C. 2003. "A review of mobile HCI research methods," *Human-computer interaction with mobile devices and services*, pp 317-335.

Kjeldskov, J. and Stage, J. 2004. "New Techniques for Usability Evaluation of Mobile Systems," *International Journal of Human-Computer Studies* (60, pp 599-620.

Lazar, J., Feng, J. H. and Hochheiser, H. 2010. *Research methods in human-computer interaction*: United Kingdom: Wiley.

Lee, K. I., Jin, B. S. and Ji, Y. G. 2011. "The scenario-based usability checklist development for home appliance design: A case study," *Human Factors and Ergonomics in Manufacturing & Service Industries* (21:1), pp 67-81.

MacKenzie, I. and Soukoreff, R. 2002. "Text entry for mobile computing: Models and methods, theory and practice," *Human–Computer Interaction* (17:2), pp 147-198.

Mirza, A. 2008. "Potential of mobile devices in New Zealand healthcare: a thesis submitted to the faculty of Massey University at Albany in partial fulfilment of the requirements for the degree of Masters of Engineering in Software."

Moritz, F. and Meinel, C. 2010. “Mobile Web Usability Evaluation-Combining the Modified Think Aloud Method with the Testing of Emotional, Cognitive and Conative Aspects of the Usage of a Web Application.” *9th IEEE/ACIS International Conference on Computer and Information Science, pp. 367-372.*

Oulasvirta, A. and Sumari, L. 2007. “Mobile kits and laptop trays: managing multiple devices in mobile information work,” *in Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, CA, 1127-1136

Pascoe, J., Ryan, N. and Morse, D. 2000. “Using while moving: HCI issues in fieldwork environments,” *ACM Transactions on Computer-Human Interaction (TOCHI)* (7:3), pp 417-437.

Petrie, H. and Bevan, N. 2009. *The evaluation of accessibility, usability and user experience*: In C. Stephanidis (Ed.) The Universal Access Handbook. London: Taylor & Francis.

Preece, J., Rogers, Y. and Sharp, H. 2002. *Interaction design: beyond human-computer interaction* New York: Wiley.

Smith, A. 1997. *Human computer factors: A Study of Users and Information Systems* New York: McGraw-Hill.

Tamminen, S., Oulasvirta, A. and Toiskallio, K., & Kankainen, A. 2004. “Understanding mobile contexts,” *Journal of Personal and Ubiquitous Computing.* (8: Special Issue), pp 135-143.

Te'eni, D., Carey, J. and Zhang, P. 2007. *Human computer interaction: developing effective organizational information systems* New York: Wiley.

Thimbleby, H. 2007. “Interaction walkthrough: evaluation of safety critical interactive systems,” *Interactive Systems. Design, Specification, and Verification*, pp 52-66.

Vermeeren, A. P. O. S., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J. and Väänänen-Vainio-Mattila, K. 2010. “User experience evaluation methods: current state and development needs,” Proceedings: NordiCHI, Reykjavik, Iceland, pp. 521-531.

Zhang, D. and Adipat, B. 2005. “Challenges, methodologies, and issues in the usability testing of mobile applications,” *International Journal of Human-Computer Interaction* (18:3), pp 293-308.

COPYRIGHT