Appropriating Heuristic Evaluation for Mobile Computing

E. Bertini (1), T. Catarci (2), A. Dix (3), S. Gabrielli (4), S. Kimani (5), G. Santucci (2)

(1) DIUF - University of Fribourg
Bd de Pérolles, 90
CH-1700 Fribourg, Switzerland
Phone: +41 26 300 9299
Fax: +41 26 300 9726
enrico.bertini@unifr.ch

(2) DIS - University of Rome "La Sapienza"
Via Ariosto 25 – 00185 Rome, Italy
Phone/Fax: +39-0677274007
{catarci, santucci}@dis.uniroma1.it

(3) Computing Department - Lancaster University
InfoLab21, South Drive, Lancaster, UK
LA1 4WA
Phone: +44 7887 743 446
Fax: +44 1524 510 492
dixa@comp.lancs.ac.uk

(4) HCI Lab - Department of Math and Computer Science
University of Udine
Via delle Scienze, 206 - 33100 Udine, Italy
Phone: +39 0432 558457
gabrielli@dimi.uniud.it

(5) Institute of Computer Science and Information Technology
JKUAT
PO Box 62000 - 00200 Nairobi, Kenya
Phone: +254-67-52244 or 52245
stevekk@iname.com

Appropriating Heuristic Evaluation for Mobile Computing

Abstract
Heuristic evaluation has proven popular for desktop and web interfaces, both in practical design and as a research topic. Compared to full user studies, heuristic evaluation can be highly cost-effective, allowing a large proportion of usability flaws to be detected ahead of full development with limited resource investment. Mobile computing shares many usability issues with more conventional interfaces. However, it also poses particular problems for usability evaluation related to aspects such as limited screen real estate, intermittent user attention, and contextual factors. This paper describes a modified collection of usability heuristics that are designed to be appropriate for evaluation in mobile computing. They have been systematically derived from extensive literature and empirically validated. They therefore offer a sound basis for heuristic-based evaluation in mobile computing. Besides introducing the reader to the practical use of heuristic evaluation, the paper also closes with a description of potential future research in the area.

1. Introduction

Expert-based evaluation techniques, such as heuristic inspection (Nielsen et al., 1990) and cognitive walkthrough (Wharton et al., 1994) typically benefit from providing evaluators with guidance (for instance, a set of heuristics or a checklist) for identifying a prioritized list of usability flaws (Kjeldskov et al., 2005). Expert-based evaluation techniques are also well-known methods that can realize a relatively quick and easy evaluation.

According to Po (2003), mobile computing devices are typically 'smart products' or 'information appliances', and are generally consumer products. Their users are thus a 'heterogeneous group' (Sade et al., 2002) and so it may be more difficult to find suitable surrogate users for user-based testing in mobile computing (Po, 2003). Po further observes that even if appropriate surrogate users were found, the realistic recreation of the user context in laboratories would be challenging because of user mobility, which makes observation and video recording difficult (Vetere et al., 2003). “Given the problems associated with user-based evaluations of mobile devices, expert-based usability techniques are considered to be more appropriate” (Po, 2003). However, it is worth noting that expert-based techniques have in the past been criticized for finding proportionately fewer problems in total, and disproportionately more cosmetic problems (Karat et al., 1992). In mobile computing the capacity of expert-based techniques to adequately capture the multiple contextual factors that affect user–system interactions in real settings has been questioned (for instance: Kjeldskov et al., 2003; Johnson, 1998).

We believe that heuristic evaluation can be enriched and adapted toward capturing contextual factors. This paper describes how standard heuristic evaluation can be made more appropriate/relevant for mobile computing. In particular, the paper describes a modified collection of usability heuristics that are designed to be appropriate for this area. The heuristics have been systematically derived from extensive literature and empirically validated, and so offer a sound basis for heuristic-based evaluation of mobile computing. As well as introducing the reader to the practical use of heuristic evaluation, the paper also describes potential future research in the area.

This work has been carried out in the context of MAIS[1], a project whose research goal is to provide a flexible environment to adapt the interaction, and the information and services provided, according to ever changing requirements, execution contexts, and user needs.

The rest of the paper is organized as follows: Section 2 highlights some of the challenges posed by mobile devices, applications and context; Section 3 discusses the standard heuristic evaluation method; Section 4 describes the methodology we adopted to appropriate heuristic evaluation for mobile computing and the results that we obtained; Section 5 contains reflections regarding this research activity; and Section 6 concludes the paper and highlights some future work.

Our focus in this paper is on usability problems in mobile devices and a discussion of their sources. However, this should be set against a broader view of the fantastic world of new opportunities, advantages, and benefits that mobile devices and contexts bring. While we will not touch explicitly on this again, the paper should be read in the light that the problems and limitations are ones worth tackling because of the opportunities offered by the technology.

2. Mobile devices, applications, and their context

In order to better understand how usability in mobile computing can be evaluated and improved, it is useful to outline specific limitations inherent in mobile devices, applications and their context. These fall into two broad categories: limitations due to the nature of the devices themselves, and limitations due to context of use and style of interaction.

2.1 Limits posed by the device

Small-screen - In order to be portable, mobiles must necessarily be small, and tend to have small screens, therefore problems due to the screen real estate are intrinsic and can be addressed only by figuring out new techniques to organize information visually.

Limited input – Because of device format, input mechanisms are inherently limited. Currently the most common means of interaction are: numeric keypads, which are used in almost all cell phones; and styluses, which are the primary input means in PDAs and smart phones.

Limited bandwidth and cost – Mobile Internet connections are still slow. This is in fact still one of the main factors limiting mobile Internet access. To this we must also add the problem of the cost model. Most companies offer their Internet access in a pay per KByte policy that obviously limits the size of pages and the number of requests.

[1] http://www.mais-project.it

Limited connectivity – Perhaps more than bandwidth, the latency of the connection affects its usability. The limited coverage of different networks and the consequent intermittent connection makes the latency extremely variable, as well as giving rise to problems of how to portray these hidden network properties to the user. There is also the problem of seamlessly switching between different types of network, e.g. WiFi to GPRS.

Limited computational resources - This means the capabilities of applications are limited. However, this should be overcome in the near future as new processors and memories specifically designed for mobile devices increase their quality and speed.

Limited power (batteries) - This is often an underestimated issue, but the batteries are still a big problem for every kind of mobile system (laptops included). This has a big impact on end users: limited autonomy means limited availability, which in turn means limited reliability.

Wide heterogeneity (of OSs and physical properties) - Users of mobile systems must always adapt to new forms of interaction as they switch to different mobiles. Changing the physical device and operating system usually translates into the need to re-learn functions, operations, messages, etc., with an enormous waste of resources.

2.2 Limits posed by context and interaction

Variable context – Since mobile devices, by definition, are mobile, the context in which they are used is continually changing. This poses challenging new issues because, though context has always been considered a fundamental aspect to analyse in usability studies, only now must we address such frequent and complex context variations within the same device, application, or single user.

Kind of interaction – The nature of interaction also changes in mobile settings. In general, users tend to interact in small and focused chunks of activities, more so than in fixed settings. A high proportion of tasks in a mobile environment consist of a few fast steps that the user should be able to execute without cognitive effort. In addition, mobile tasks may happen in conditions where users' attention is necessarily reduced, or may be part of more complex activities with which they should not interfere.

Interruptions – Mobile devices/applications are always "with us". If this, on one hand, means that computation and data are always available, it is also true that notifications and requests for attention can happen at inappropriate moments and that some tasks may be interrupted. This raises two kinds of problem: appropriateness of notifications and recovery from interruptions.

Privacy and security – Privacy issues become more prominent. While staying mobile, users find themselves in a variety of spaces (private and public), in a variety of situations (formal and informal), and in a variety of infrastructures (wireless and cable connection). Moving through these settings means having different needs for privacy and security.

Intimacy and availability - Because mobile devices are mobile, they are personally available in a way that fixed devices are not. Moreover, they seem to engender a sense of being “personal” in a deeper sense than desktop PCs (e.g., not just my PC but my PDA and definitely my phone).

3. Heuristic Evaluation

3.1 Introduction to Heuristic Evaluation

Heuristic evaluation (Nielsen et al., 1990; Nielsen, 1994b) is an inspection usability evaluation method. In heuristic evaluation, experts scrutinize the interface and its elements against established design rules. The experts should have some background knowledge or experience in HCI design and usability evaluation. Three to five experts are considered to be sufficient to detect most of the usability problems. The enlisted experts individually evaluate the system/prototype under consideration. They assess the user interface as a whole and also the individual user interface elements. The assessment is performed with reference to some usability heuristics. When all the experts are through with the assessment, they come together and compare and appropriately aggregate their findings. In Molich et al. (1990) and Nielsen et al. (1990) Rolf Molich and Jakob Nielsen initially proposed a set of usability heuristics for the design of user interfaces. Aiming to maximize the explanatory power of the heuristics, Nielsen later refined them (Nielsen, 1994b), thereby deriving the following set:

1. Visibility of system status: The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. Match between system and the real world: The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3. User control and freedom: Users often choose system functions by mistake and will need a clearly marked “emergency exit” to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4. Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.
6. Recognition rather than recall: Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.
7. Flexibility and efficiency of use: Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
8. Aesthetic and minimalist design: Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

9. Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
10. Help and documentation: Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

3.2 Strengths of Heuristic Evaluation

Here are some strengths of heuristic evaluation:

- Its ease of implementation and high efficiency (Law et al., 2002; Nielsen, 1994b). It is considered to have a good success rate in that typically only 3–5 usability experts are needed to detect most (75–80%) of the usability flaws a system presents (Nielsen, 1994b).
- Its early applicability in the development lifecycle and low cost: it requires neither a working prototype nor the real users (Nielsen, 1994b).
- It is becoming part of the standard HCI curriculum and therefore known to many HCI practitioners (Greenberg et al., 1999.). The heuristics are well documented and therefore easy to learn and put to use, so it may be argued that heuristic evaluation can also be effectively conducted by non-usability experts (Nielsen, 1994b).

On the whole, heuristic evaluation is considered to be a cost-effective evaluation method. Its main strengths lie in providing discovery and analysis resources (Cockton et al., 2003), such as domain and system knowledge, where it generally outperforms other popular inspection techniques like guideline-based methods or cognitive walkthrough (Wharton et al., 1994).

3.3 Limitations of Heuristic Evaluation

Here are some specific limitations of heuristic evaluation:

- Heuristic evaluation is highly dependent on the skills and experience of the specific usability expert(s) involved. At a high level of generality, the heuristics are “motherhood statements that serve only to guide the inspection rather than prescribe it” (Greenberg et al., 1999).
- Participants are not the real users. Regardless of the experts' skills and experience, they are still “surrogate users” (i.e. experts who emulate real users) (Kantner et al., 1997), therefore the resulting data are not really representative of the real users.
- Heuristic evaluation does not fully capture or take into account the context of use of the system under evaluation but rather evaluates it “as a relatively self-contained object” (Muller et al., 1995).
- It has been said that the majority of usability flaws detected by heuristic evaluation are ‘minor’ usability problems (for instance, by Nielsen (1994a)), or false positives, problems that do not negatively impact user performance or users' perception of system quality (Simeral and Russell, 1997).

When compared to other expert techniques, such as guideline-based methods and cognitive walkthrough, heuristic evaluation is strong in terms of thoroughness (percentage of problems found), but weak in terms of efficiency (number of true positives vs. false positives) and, like other inspection methods, is vulnerable to expert biases (Cockton et al., 2003).

4. Appropriating Usability Heuristics

Analyses of HE have shown that it is more likely for this method to miss relevant usability problems when the system to be evaluated is highly domain-dependent, and when evaluators have little domain expertise[2]. To overcome these limitations of the method when applied to mobile systems and settings, we have conducted an in-depth investigation of usability issues affecting mobile applications. The work leading to the set of specialized heuristics for mobile computing presented in Section 4.2 is based on this empirical evidence. The goal of the mobile heuristics described in Section 4.2 is to better support and contribute to the domain expertise of evaluators applying HE to mobile computing.

4.1 Methodology for Realizing Mobile Issues

To develop usability heuristics for mobile computing, three authors of this paper worked as usability researchers at the following activities:

1. Each of the three was assigned a unique set of papers to analyze independently. The papers originated from the list used in Kjeldskov et al. (2003); a recent meta-analysis of HCI research methods in mobile HCI[3]. We updated the list with papers published in the period 2004–2005 and selected only those with elements of evaluation. The analysis entailed documenting, for each of the papers, appropriate values for the following dimensions:
    - *Evaluation goal* – is the evaluation mainly intended to demonstrate whether one technique is better than another, that is, a comparative study; or is it mainly exploratory, that is, understanding what kind of usability problems may rise with a given design.
    - *Evaluation method* – is the evaluation method expert-based (made by experts through inspection), user-based (observing users performing tasks), or model-based (computing usability metrics through formal models).
    - *Evaluation setting* – is the evaluation conducted in a laboratory (or any other controlled setting) or in the field.
    - *Real device/emulator* – is the application under inspection tested with a real device or in a emulated environment.
    - *Location matters* – does the application take location into account or not.

   Moreover, each of the usability researchers individually documented mobile usability issues that were indicated by (or evident from) each of the papers. At the end of this process, three different lists of usability issues were produced, containing the analysis of all the papers collected in the first phase.

---

[2] See also http://www.useit.com/papers/heuristic/usability_problems.html

[3] The papers were selected from top-level conferences and journals like CHI, AVI, UIST, TOCHI, etc. For details see J. Kjeldskov and C. Graham. A Review of Mobile HCI Research Methods. In L. Chittaro, editor, *International Symposium on Human Computer Interaction with Mobile Devices and Services - Mobile HCI'03*, pages 317–335. Springer-Verlag, 2003.

2. In the next step, the usability researchers came together and consolidated their individual realizations. Individual findings had to be cross-checked and merged into a single consolidated list. This was done in the form of a spreadsheet.

3. Each researcher was then given the same realized list of mobile usability issues and asked to independently categorize (group or cluster) the issues. The idea was to find a way to summarize all the encountered issues and present them at a higher level of abstraction. On reflection, this was useful to check whether traditional heuristics covered each class of usability problems, or not. The researchers then presented and shared their individual categorization results with each other. Each researcher was requested to individually work further on his/her categorization with reference to the other categorizations, by eliminating redundant usability issues, clarifying the mobile usability issues, and grouping the obtained issues to an abstraction level that would be appropriate for developing/generating heuristics. Finally, they came together again to brainstorm and consolidate their work, and to harmonize the terminology used to describe the issues. Figure 1 shows an image of the set of issues produced. Here we present the top-level classes of problems with examples of subclasses to make their meaning clear.

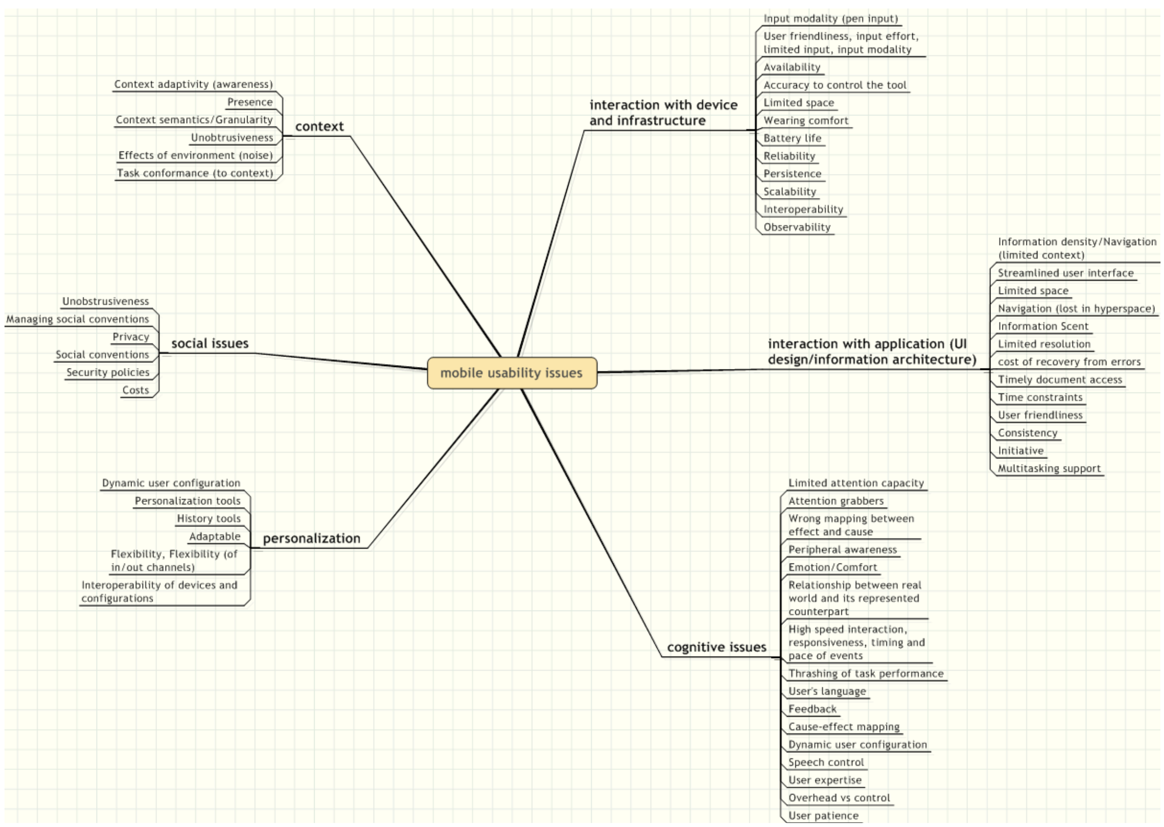

Figure 1 - Mobile Usability Issues

4.1.1 Mobile Usability Issues

Below we describe the usability issues collected and grouped in high-level classes as described in Section 4.1.

- *Interaction with device and infrastructure* – many of the problems we have found in our research have a strong connection with the limits of the device and/or the infrastructure it is connected to.
- *Interaction with application* – this collects classes of problems connected to traditional screen design and information architecture.
- *Cognitive issues* – here we characterize usability problems stemming from an overload of cognitive resources or a mismatch between a cognitive model and reality. While this aspect has always been taken into account in traditional studies, in mobile settings it becomes more evident and presents new challenges.
- *Personalization* – standard heuristics tend to overlook problems connected to personalization or adaptation. While in standard settings this issue can be considered minor with respect to others, with mobile devices this aspect can really be critical.
- *Social issues* – mobile devices and applications are used in a wide spectrum of environments and social conditions: private or public, alone or in groups, etc. This means that the social impact of adopted design solutions cannot be underestimated. Issues like privacy, security, user image, and social conventions thus become of great importance.
- *Context* – similarly to social issues, it is necessary to take into account how the environment can affect interaction. Not only do social conventions and relationships with people matter, but also how potential physical environment features affect the design of an interface.

4.2 Methodology for Realizing Mobile Heuristics

This section discusses our research toward developing a set of mobile usability heuristics and also our efforts toward assessing the proposed mobile usability heuristics.

4.2.1 Toward a set of heuristics

The brainstorming activity described in Section 4.1 was continued and further articulated in a series of new individual or collaborative tasks aimed at developing a set of heuristics for mobile computing evaluations. By capitalizing on the outcome of our previous analysis of mobile usability issues, we decided to rely on the following developmental process to come up with a new set of heuristics, better suited to be applied to mobile evaluation settings.

Phase 1:
Each of the 3 usability researchers was provided with a table reporting Nielsen's traditional heuristics (Nielsen, 1994b) together with their corresponding definitions. Each researcher worked individually at assessing: which of Nielsen's heuristics were considered irrelevant for mobile settings; which of Nielsen's heuristics were relevant, but needed some revision or modification; and which additional heuristics needed to be included in the original set to cover relevant aspects of mobile applications. To better steer our individual relevance judgment of the heuristics, we thought it useful to define a guiding principle to be adopted and shared during the assessment

work: this was a concise answer to the question: "What are the primary goals of mobile applications?", which we expressed as follows: "To enable a user-friendly navigation of relevant information or features in mobile conditions of use". The assessment and brainstorming activity performed in this phase was also informed by the consolidated version of the mobile usability issues that had been previously realized (Section 4.1).

Phase 2:
Each of the usability researchers compared her/his own table of proposed heuristics with that of another researcher, to produce a new consolidated table. This activity was meant to be carried out individually, but based on comparing the work done by two researchers. The aim was to speed up the improvement of the set of heuristics proposed, in terms of their clarity and relevance to the mobile application field.

Table 1: Mobile usability heuristics

Mobile Heuristic	Description
Heuristic 1	Visibility of system status and losability/findability of the mobile device
Heuristic 2	Match between system and the real world
Heuristic 3	Consistency and mapping
Heuristic 4	Good ergonomics and minimalist design
Heuristic 5	Ease of input, screen readability and glancability
Heuristic 6	Flexibility, efficiency of use and personalization
Heuristic 7	Aesthetic, privacy and social conventions
Heuristic 8	Realistic error management

Phase 3:
A new refinement process was started on the set of heuristics included in the three consolidated tables produced in phase 2. It involved: first, a discussion meeting among the usability researchers to arrive at a shared table consolidated from the three developed in phase 2; then, submitting this set of heuristics (with their definitions) to a number of targeted HCI researchers and professionals in the mobile computing and usability community, to elicit feedback on the adequacy of the heuristics proposed. We contacted 19 experts, in person, by email or by phone, and we received feedback from 8 of them. The 3 researchers then met to discuss and compare the experts' comments with the researchers' consolidated table, and arrived at the final set of mobile usability heuristics summarized in Table 1 and described below:

- Heuristic 1 - Visibility of system status and losability/findability of the mobile device: Through the mobile device, the system should always keep users informed about what is going on. Moreover, the system should prioritize messages regarding critical and contextual information such as battery status, network status, environmental conditions, etc. Since mobile devices often get lost, adequate measures such as encryption of data should be taken to minimize loss. If the device is misplaced, the device, system or application should make it easy to recover it.
- Heuristic 2 - Match between system and the real world: Enable the mobile user to interpret the information provided correctly, by making it appear in a natural and logical order; whenever possible, the system should have the capability to sense its environment and adapt the presentation of information accordingly.

- Heuristic 3 - Consistency and mapping: The user's conceptual model of the possible function/interaction with the mobile device or system should be consistent with the context. It is especially crucial that there be a consistent mapping between user actions/interactions (on the device buttons and controls) and the corresponding real tasks (e.g. navigation in the real world).
- Heuristic 4 - Good ergonomics and minimalist design: Mobile devices should be easy and comfortable to hold/carry along as well as robust to damage (from environmental agents). Also, since screen real estate is a scarce resource, use it with parsimony. Dialogues should not contain information that is irrelevant or rarely needed.
- Heuristic 5 - Ease of input, screen readability and glancability: Mobile systems should provide easy ways to input data, possibly reducing or avoiding the need for the user to use both hands. Screen content should be easy to read and navigate through notwithstanding different light conditions. Ideally, the mobile user should be able to quickly get the crucial information from the system by glancing at it.
- Heuristic 6 - Flexibility, efficiency of use and personalization: Allow mobile users to tailor/personalize frequent actions, as well as to dynamically configure the system according to contextual needs. Whenever possible, the system should support and suggest system-based customization if such would be crucial or beneficial.
- Heuristic 7 - Aesthetic, privacy and social conventions: Take aesthetic and emotional aspects of the mobile device and system use into account. Make sure that users' data is kept private and safe. Mobile interaction with the system should be comfortable and respectful of social conventions.
- Heuristic 8 - Realistic error management: Shield mobile users from errors. When an error occurs, help users to recognize, to diagnose, if possible to recover from the error. Mobile computing error messages should be plain and precise. Constructively suggest a solution (which could also include hints, appropriate FAQs, etc). If there is no solution to the error or if the error would have negligible effect, enable the user to gracefully cope with the error.

4.2.2 Assessing heuristics performance

To investigate the potential benefits of applying our set of heuristics for the evaluation of mobile applications, we devised and conducted an experimental study aimed at comparing the support provided by our new set of mobile heuristics vs. standard usability heuristics (here Nielsen's heuristics) to experts performing heuristic evaluation of mobile applications.

Experimental Design

Here we describe various parameters pertaining to the set-up or design of the experimental study.

Participants and Materials:
The study enlisted 8 usability experts[4], as participants, to perform a heuristic evaluation (HE) of two mobile applications for which we had already identified a number of usability flaws. The two criteria we used to select the applications to test were: being a typical application whose problems

[4] All the experts were new to the novel set of heuristics and none of the experts involved in the generation of heuristics discussed above were involved in the experimental study.

are known and evident; application whose tasks are simple and/or self-evident. After searching for applications fulfilling the foregoing conditions, we chose the following two applications: Appl.1) a mobile device application in which location matters or that primarily relies on mobility: we considered a PDA-based supermarket application; Appl.2) a mobile device application in which interface navigation is key: we considered a web-based freeware email application for PDAs[5]. We also prepared the following materials for the evaluators: consent form, demographics questionnaire, post-evaluation form for participant's comments (to be filled out by the study moderator), a set of Nielsen's 10 usability heuristics, our proposed set of mobile usability heuristics (Table 1), and Nielsen's five-point Severity Ranking Scale (SRS) (Nielsen, 1994b) (which is described in Table 2).

Table 2: Severity Ranking Scale (SRS)

Rating	*Description*
0	I don't agree that this is a usability problem at all
1	Cosmetic problem only. Need not be fixed unless extra time is available on project
2	Minor usability problem. Fixing this should be given low priority
3	Major usability problem. Important to fix, so should be given high Priority
4	Usability catastrophes. Imperative to fix this before product can be released

Experimental Conditions:
The experiment had the following two experimental conditions:

- Condition 1: N. 4 experts individually performed the HE by applying Nielsen's standard set of heuristics and Nielsen's SRS to both applications.
- Condition 2: N. 4 experts individually performed the HE by applying our set of mobile heuristics and Nielsen's SRS to both applications.

Procedure:
The 8 usability experts were randomly split into two groups, each assigned to one of the foregoing two experimental conditions (that is a between-subjects design). They all had previous expertise in the HCI evaluation field and were familiar with both the application of traditional HE methods and the use of mobile applications. Nevertheless, they were all given some brief instruction on the technique before starting the evaluation. The following protocol was used for both experimental conditions:

- Pre-evaluation session: This entailed first welcoming and greeting each evaluator. After that the goals of the study, the testing procedures, and the confidentiality issues were explained in detail. Scripts were prepared in advance and used for each usability evaluator to ensure consistency across experts and conditions. In a demographics questionnaire experts were asked about their level of education/academic status, relevant experience in both HCI and mobile computing, experience in using both a PDA and Nielsen's heuristic

[5] We used hp iPAQ Pocket PC series h5500 PDAs with integrated wireless LAN (802.11b), 48 MB ROM, 128 MB RAM, and Intel processor 400 MHz. The PDAs were running Windows CE.

evaluation method; the collected demographic data can be seen in Table 3. Most of the participants have a high level of education and an average knowledge of HCI and mobile devices. Six participants consider themselves almost knowledgeable about heuristic evaluation, while two give themselves an average rating. A training session was conducted with each evaluator to ensure that they fully understood the usability heuristics, and especially the mobile heuristics, which the participants were not familiar with; this involved the facilitator stepping through each usability heuristic and inviting the evaluators to ask questions in order to clarify the meaning of each heuristic and their understanding of the overall process.

- Evaluation session: The usability evaluators performed the usability evaluation on the mobile device by identifying usability problems and prioritizing them according to Nielsen's SRS (Table 2). Presentation of the two applications to be evaluated was counterbalanced to avoid any order effect. While evaluating the mobile device, each usability evaluator was asked to 'think aloud' to explain what s/he was trying to do and to describe why s/he was taking the action. Their comments were recorded by one of the evaluation moderators.

- Debriefing session: This focused on the evaluators' experiences of the process, and providing an opportunity to probe where behavior was implicit or puzzling to the researchers.

Table 3: Participants' demographics. Each value is ranked on a scale between 1 (min) and 4 (max)

Part.	Edu	HCI	PDAs	HE
p1	1	2	3	3
p2	4	2	2	3
p3	4	3	3	3
p4	1	2	2	3
p5	3	3	3	3
p6	2	2	1	2
p7	3	3	2	3
p8	1	1	1	2

Data Analysis

The data collected were analyzed both qualitatively and quantitatively. Comparison of HE effectiveness in the 2 experimental conditions was assessed.

Number of Flaws and Variation Among Experts:
From Table 4, it appears that the use of the mobile heuristics has increased the number of flaws identified in the analysis of both applications, and has reduced variation among experts' analyses. In comparing the type of flaws detected by using the two different sets of heuristics, we did not find evidence of problems identified only by using Nielsen's heuristics. The additional flaws

found by applying mobile heuristics were usually different from the ones identified by using Nielsen's heuristics; also, the problems identified by each expert in the mobile heuristics condition were a small number from a larger set of usability difficulties presented by the two applications, although we could find some overlaps (problems pointed out by more than one expert), which supports the idea of inter-expert consistency when applying mobile heuristics.

Table 4: Number of usability problems identified

	Appl. 1	**Appl. 2**	**Total**	**Mean (SD)**
NHE	22	28	**50**	**12.5 (10.40)**
MHE	26	38	**64**	**16 (3.74)**

Severity of Flaws and Distribution:
As depicted by Table 5 and Figure 2, Nielsen's heuristics have produced a more equally distributed severity ranking of problems detected for both applications. On the other hand, the mobile heuristics have produced a more positive evaluation of Appl.1 (61% of problems are considered minor or cosmetic) while for Appl.2 the ranking seems to be equally distributed among the four severity levels. Considering the mean values in Figure 2, it does appear that Nielsen's heuristics identify fewer Minor and Major flaws compared to the mobile heuristics. It also seems that Nielsen's heuristics have a relatively even distribution of severity ranking for the problems identified. Nielsen's heuristics could therefore do a moderate job of identifying flaws at any design level. The mobile heuristics do seem to be especially good at identifying Minor and Major flaws rather than those at the extremes.

Table 5: Actual Number of Flaws and Severity

	NHE			**MHE**		
	Appl. 1	**Appl. 2**	**Mean % for both appl.s**	**Appl. 1**	**Appl. 2**	**Mean % for both appl.s**
Cosmetic	5	7	6	5	5.75	5.375
Minor	5.5	6.5	6	11	13.25	12.125
Major	6.5	6.5	6.5	9	10.25	9.625
Catastrophe	5	8	6.5	1	8.75	4.875

Total of flaws	22	28	26	38

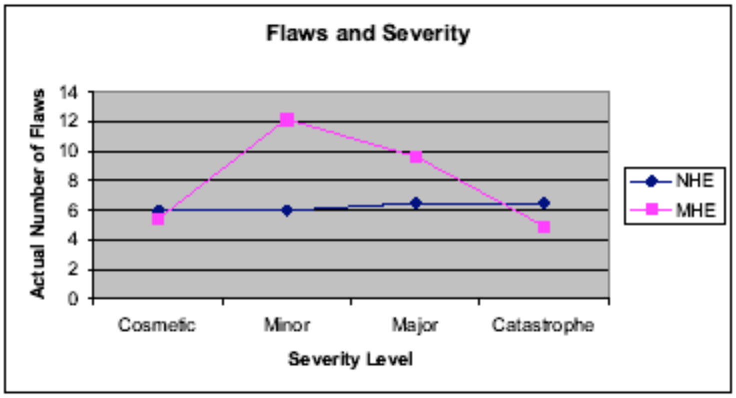

Figure 2: Actual number of flaws, severity and distribution

Figure 3 can be used for further analysis of how the specific heuristics from both sets fare with regard to average severity and average number of flaws. Figure 3 indicates that mobile heuristics are more effective in supporting the detection of flaws, while Nielsen's heuristics seem better suited to cover the case in which high severity flaws are present; also, mobile heuristics seem to support a more detailed evaluation of the mobile application (without considering the flaws classified as catastrophic). It is worth noting that some of the foregoing observations from Figure 3 are similar to those from Figure 2.

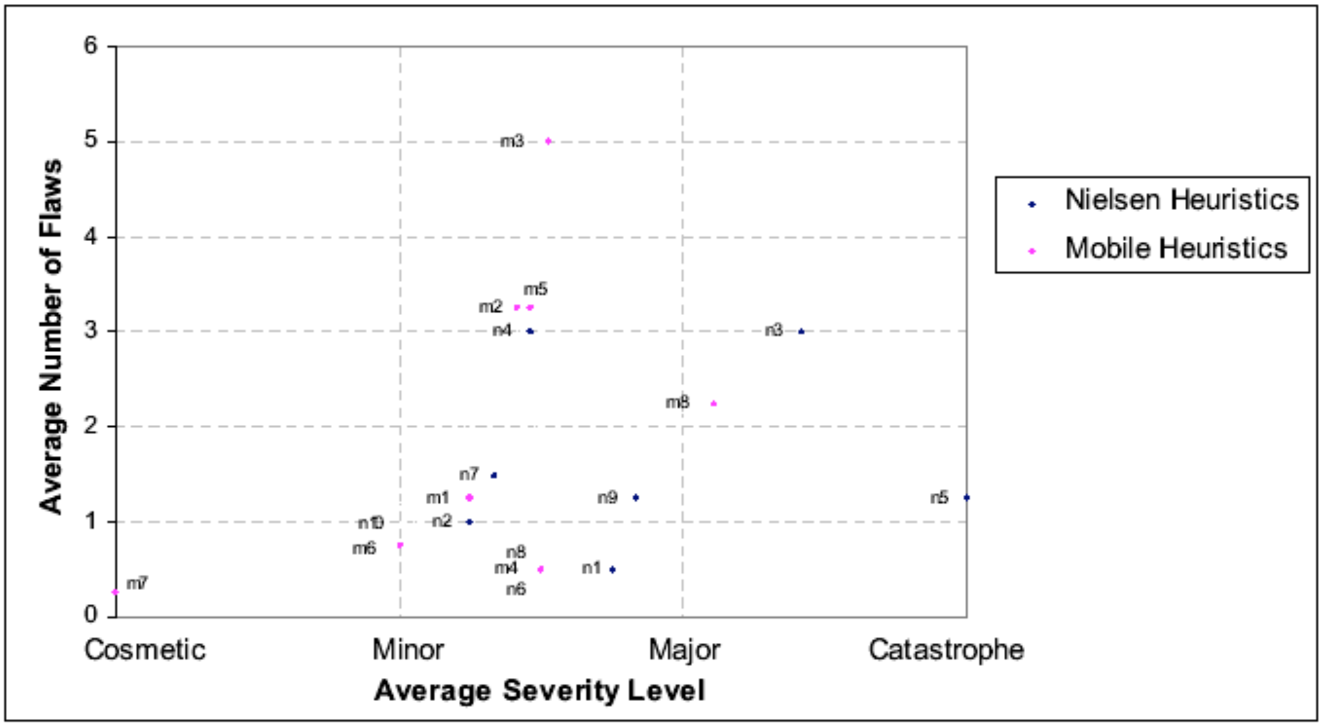

Figure 3: Comparison of the sets of heuristics: flaws and severity

So far it might be observed that the mobile heuristics produce a more accurate evaluation in terms of number of problems detected (more flaws are identified), reduced variation among experts' analyses, and problems' severity ranking (this is actually also supported by the qualitative data collected during the evaluation, where most experts said that Appl.1 was much better designed for a mobile use when compared to Appl.2). Thus the mobile heuristics tend to focus the evaluation on the mobile issues instead of directing experts' attention at a more general level (although the kind of setting we used in this study promoted an evaluation of applications that was more functionalities-based than contextual). Moreover, the mobile heuristics could be applied when/where the extreme flaws have been addressed or are not an issue in the design. If such flaws have to be identified before proceeding, mobile heuristics could be applied after Nielsen's heuristics. It is worth recalling that there are some problems that Nielsen's heuristics failed to identify (based on Table 4). Some might now be identified by mobile heuristics and might lie between Minor and Major severity levels (Table 5 and Figure 2).

Usability Flaws and Heuristics:
As seen in Table 6 and Table 7, the most frequently used/highlighted heuristics in the mobile applications are as follows[6]:
Nielsen's heuristics: Nielsen's heuristic 4 (12 times), Nielsen's heuristic 3 (12 times). The foregoing are [each] less than any of the following mobile heuristics.

---

[6] It should be noted that some of the participants indicated that some of the flaws were individually related to more than one type of heuristic (and thus the number of counts for the heuristics shown in Table 6 (and also Table 7) is greater than the number of flaws as shown in Table 4).

Mobile heuristics: mobile heuristic 3 (20 times), mobile heuristic 5 (13 times), mobile heuristic 2 (13 times).

Table 6: Nielsen's heuristics and corresponding usability problems

Nielsen's Heuristics 1-10	Number of Usability Problems	Description of Heuristic
4	12	Consistency and standards
3	12	User control and freedom
7	6	Flexibility and efficiency of use
9	5	Help users recognize, diagnose, and recover from errors
5	5	Error prevention
2	4	Match between system and the real world
10	3	Help and documentation
1	2	Visibility of system status
6	2	Recognition rather than recall
8	2	Aesthetic and minimalist design

Table 7: Mobile heuristics and corresponding usability problems

Mobile Heuristics 1-8	Number of Usability Problems	Description of Heuristic
3	20	Consistency and mapping
5	13	Ease of input, screen readability and glancability
2	13	Match between system and the real world
8	9	Realistic error management
1	5	Visibility of system status and losability/findability of the mobile device
6	3	Flexibility, efficiency of use and personalization
4	2	Good ergonomics and minimalist design
7	1	Aesthetic, privacy and social conventions

It is interesting to observe that these highlighted Nielsen's heuristics (4 [Consistency and standards], 3 [User control and freedom]) are related to the highlighted mobile heuristics (3 [Consistency and mapping], 5 [Ease of input, screen readability and glancability], 2 [Match between system and the real world]). We could consider the highlighted Nielsen's and mobile heuristics as being those most violated or most noticed, although their recurrence could be due to the particular type of evaluation/application(s) that were provided to experts.

The mobile heuristics probably scored such high figures (i.e. were able to identify more flaws under these related heuristics) because of the way the mobile heuristics have been revised and/or extended to capture mobile computing aspects.

It is interesting to note that the mobile heuristics "bring to the top" heuristics that are related to context. For instance: Nielsen's heuristic 2 has a score of 4; the related revised heuristic for mobile computing (mobile heuristic 2) scores 13. It may therefore be observed that the mobile heuristics make issues and flaws that have to do with context more apparent during the evaluation. Also, from our qualitative analysis of experts' reports it was found that when the evaluator identified a flaw that could not be straightforwardly mapped to a specific mobile heuristic, s/he chose to assign it to mobile heuristic 2 or 3. Moreover, an evaluator stressed the need to make more explicit the word ‘context’ in the description of mobile heuristic 3. The description of the heuristic is found in Section 4.2.

As seen in Table 6, the participants reported some usability problems regarding ‘Help and documentation’ (Nielsen's heuristic 10). This observation may be an indication that people using mobile applications still expect such applications to provide help. Though they might prefer that the help be ‘interactive’, non-distractive, not be a separate task, etc., the designer could consider the use of audio or some ‘light-weight’ approach (e.g. FAQs).

Time Taken to Evaluate:
It seems that the application of the mobile heuristics was more time demanding during the whole evaluation, as seen in Table 8. This may be due to the experts’ relative unfamiliarity with these heuristics compared to Nielsen’s heuristics. We tried to reduce this familiarity issue (it cannot be eliminated simply in an evaluation session) by giving the experts who were using the mobile heuristics some extra time at the beginning of the evaluation to study the mobile heuristics, in order to familiarize themselves with them and to ask questions. Although the application of the mobile heuristics was more time demanding, we should, however, also observe that variation among experts was relatively high, confirming that heuristic evaluation is an evaluation technique strongly dependent on experts’ previous knowledge and expertise with the heuristics, the application domain, etc.

Table 8: Time taken in minutes

	Appl. 1	Appl. 2	Total	Mean (SD)
NHE	106	92	**198**	**49.5 (27.196)**
MHE	155	136	**291**	**72.75 (44.776)**

As a general observation, it is worth mentioning that because our study adopted a between-subjects design, with the inherent risk that individual differences between participants can bias results, the fact that the application of mobile heuristics results in reduced variation among the participants' analyses is therefore notable.

5. Reflections toward deeper principles

Many of the heuristics in ‘traditional’ Heuristic Evaluation appear to be phrased in a way that is general over all systems, and this is true also of many of the other forms of design guidelines or principles used in interaction design. This raises a number of questions:

(i) Why do we need specific heuristics for mobile devices - why not use standard ones?

If (as this paper assumes) we do need them, then this raises further questions

(ii) How do we know the heuristics we have presented are correct?
(iii) How do we know they are sufficient or complete?
(iv) How do we know they work in real design?
(v) Can we assess their scope – do our heuristics simply reflect current mobile technology and applications?

The answer to (i) is that we do need specific heuristics because the traditional heuristics implicitly embody assumptions about static desktop location and use. The differences between standard heuristics and the mobile usability heuristics presented in Section 4.2.1 are precisely due to the differences between mobile and fixed use.

Our confidence on the correctness (ii) and sufficiency (iii) is based on the rigorous methodology used to derive the heuristics, distilling the knowledge and expertise in published work (Section 4.1), more analytic refinement of established heuristics (Section 4.2.1, phase 1) and review by experts (Section 4.2.1, phase 3). The feedback from experts gives some confidence in utility (iv), and this was confirmed by the empirical study (Section 4.2.2), which also bolstered confidence in the correctness and sufficiency.

No set of heuristics or guidelines will be complete, but it can be sufficient to cover the more common or serious pitfalls. However, while the process of distillation from expert opinion and empirical testing suggests that the heuristics are sufficient for current mobile applications, on their own they do not tell us about applicability in the future (v). Mobile technology is changing rapidly and new applications are emerging. While it would be foolish to believe that we can foresee all the ramifications of these, we can try to ensure that we understand the scope of the new mobile heuristics. In particular if we attempt to make explicit the different assumptions that underlie the new heuristics, we will be able to tell when these change and be in a better position to add to or modify this set in the future.

We will examine the differing assumptions that underlie desktop and mobile use under four headings: the nature of mobile devices, the environment of mobile infrastructure, the context of mobile use and the purpose of mobile tasks. For each we will examine the extent to which they are reflected in the heuristics and, where appropriate, how they may develop in the medium term.

5.1 The nature of mobile devices

One of the most obvious differences between mobile devices and fixed ones is size.

It has been pointed out that desktop screen design is often lazy design – putting everything on screen and letting the user worry about what is important (Dix, 1999). In contrast, for the small screen of a mobile device it is crucial that just the right information and input options are available at the right time – mobile device designers have to think far more carefully about the user's task than desktop designers. This is emphasized in Heuristic 4 "since screen real estate is a scarce resource, use it with parsimony. Dialogues should not contain information that is irrelevant or rarely needed."; on a desktop application we would (lazily) just show everything. This is also reflected in Heuristic 2: "Enable the mobile user to interpret the information provided correctly, by making it appear in a natural and logical order", and in Heuristic 6: "Allow mobile users to tailor/personalize frequent actions…". While these are both good advice for any interface it is particularly important on a small screen to help deliver the right information at the right time.

Several heuristics pick out issues of system adaptation. The system should (Heuristic 1) "prioritize messages", (Heuristic 2) "… sense its environment and adapt the presentation of information accordingly", (Heuristic 6) "suggest system-based customization" and Heuristic 8 "Constructively suggest a solution" for errors. In desktop systems 'intelligent' system features can often get in the way and it is often better to have simple consistent interfaces. Of course this consistency is itself also more important when descriptions of actions are by their nature more parsimonious and Heuristic 3 focuses on this "consistent mapping between user actions/interactions". The balance between consistency and intelligence changes as the input/output bandwidth diminishes and the potential annoyance of wrong adaptations may be less problematic than the cost of doing everything by hand. Note too that Heuristic 3 is as much about *external* consistency with the environment as *internal* consistency over time.

The overall small physical size is also central to Heuristic 1's focus on losability/findability. A small device can easily get lost both in public places and in the home. However, its size means that it is often kept close at hand, both allowing it to be used as a proxy for the user in location services and also meaning that it becomes a very personal device, often used for private purposes. The importance of privacy is picked up in Heuristic 1: "Since mobile devices often get lost, adequate measures such as encryption of data should be taken to minimize loss" and Heuristic 7: "Make sure that users' data is kept private and safe". The personal nature is also picked up in Heuristic 7: "Take aesthetic and emotional aspects of the mobile device and system use into account". One of the unexpected lessons that mobile phone manufacturers had to learn quickly was that mobile phones are fashion items as well as functional devices. In addition, the content of mobile communications is often very rich and personal.

5.2 The environment of mobile infrastructure

A key difference between driving across Africa and driving across Europe is the different transport infrastructure. The road system, signage, garages for repairs, and petrol filling stations are as much a part of the driving experience as the car dashboard. Similarly, for the mobile user the infrastructure in terms of wireless connectivity, charging, and data synchronization is as much a part of the mobile experience as the usability of the device itself.

The heuristics presented here are focused primarily on the use of the device itself and only marginally refer to this mobile infrastructure. Given that the heuristics reflect the current literature, clearly there is need for research in this area, which then may lead to further heuristics or guidelines for mobile infrastructure.

The influence of infrastructure can be thought of as different kinds of connectivity: connectivity to networks, connectivity to power, connectivity to data and connectivity to location services.

Network connectivity is always of concern to mobile phone users and it is still not uncommon to see people hanging out of windows, or waving phones in the air looking for signal. Heuristic 1 notes the importance of giving information on "network status", and phone users become proficient at reading the signal bars on their phones. In related technologies this is less well managed and owners of digital radio sets often become confused as digital stations seem to appear and disappear without warning; whereas analog broadcasts degrade slowly with distance, digital broadcasts can either be interpreted perfectly from weak signal, or not at all. Similarly WiFi networks are seen as something akin to magic even by expert computer users, both in terms of how displayed signal levels correspond to actual access and in terms of the means to obtain connections through multiple levels of system property settings and authentication dialogues. Clearly we are still a long way from achieving even this simple goal of Heuristic 1.

The variability of network connectivity has been made a deliberate design feature in the notion of *seamful design* (Chalmers et al., 2004). Observations of network-based mobile games showed that players rapidly became aware that they could use patches of good or poor network connectivity in order to give them advantage during game play. This then led to games specifically designed using this notion.

Heuristic 8 on error management, while being partly general advice and partly about minimizing dialogue, is also indirectly related to connectivity. It is needed precisely because the user is far away from documentation, user guides, and expert help, and cannot rely on online help because of small screen size.

Heuristic 1 also notes the importance of "battery status". While battery technology has progressed remarkably, it is still one of the limiting factors for mobile devices, so much so that in some UK motorway service stations there are small racks of mobile phone lockers near the entrance where you can leave your phone to charge while you eat. The larger issues surrounding this are not mentioned in the heuristics, as a designer has little influence over them, but certainly standardization of power supply would seem an important step in reducing the plethora of power adaptors so many of us carry while traveling, as well as making public power-charging facilities easier to manage.

Interestingly, power is not unrelated to network connectivity, as mobile phones consume more power if they have to connect to more distant radio masts. However, few users are aware of these interactions and an application of Heuristic 1 would be to give users a better feel for these things.

Heuristic 1 also notes the importance of minimizing data loss. This is related both to privacy (not losing data to others) and to data recovery. Data synchronization has a long history, back to early systems such as the CODA file system (Kistler et al., 1992; Braam, 1998), but still seems to be only poorly managed in practice. While there are ways to synchronize data between mobile devices and desktop systems, the fact that devices are connected through mobile networks could be used more widely to seamlessly backup crucial information such as phone address books. As mobile devices increase in data capacity, devices such as USB sticks and MP3 players are increasingly becoming the vectors for synchronization, so that perhaps new classes of application and hence a need for new heuristics will arise in this area.

Much of the early research on mobile platforms was based around more substantial mobile computers and (often collaborative) remote applications. Issues of data synchronization over networks were crucial and in particular the problems due to network delays (Davies et al., 1996;

Dix, 1995). This is still a major problem; for example, few users are aware that SMS is an unreliable messaging medium and can have substantial delays, especially for international texts. Strangely this does not seem to be prominent in the current literature and hence is not reflected in the heuristics. Perhaps this reflects the belief (as has been the case for 20 years) that 'soon' everything will be quick enough, or perhaps simply that empirical work is usually carried out in areas of high connectivity.

For location-based services, it is important that users understand the accuracy and other features of the location estimates. Where information or other services are based on discretized regions confusion can arise at boundaries, rather like the digital radio example. Heuristic 1 applies again here, making not just location but data about uncertainty available. However, how this can be achieved in practice is still a matter for research, so more detailed general guidelines are not yet possible.

5.3 The context of mobile use

Mobile devices are used while walking, (with care) in vehicles, outside in the rain, on the beach in the sand; they are often held while trying to do other things: open doors, carry shopping, pay for the bus, and in environments with other people. This rich set of physical and social contexts is reflected in several of the heuristics.

Heuristic 4 in particular notes the importance of "Good ergonomics" so that devices are "easy and comfortable to hold/carry along" and also that they are "robust to damage" when they inevitably get banged or dropped. Heuristic 1's focus on losability/findability reflects the dynamic context where mobile devices may be put down while carrying out other tasks.

The social context is also noted in Heuristic 7, "Mobile interaction with the system should be comfortable and respectful of social conventions", and this interacts with the ability to (Heuristic 6) "dynamically configure the system according to contextual needs".

Avoiding embarrassing symphonic ring tones and similar context-sensitive adaptations has been the focus of much research and a mobile device may potentially have access to just the environmental information to make this possible. While still very much a matter of ongoing research, this is reflected in several heuristics: (Heuristic 1) "messages regarding … environmental conditions", (Heuristic 2) "the system should have the capability to sense its environment", (Heuristic 3) "user's conceptual model … consistent with the context".

Mobile devices are often used in far from optimal lighting conditions and while moving, making small fonts hard to read and so further reducing effective screen size. Both of these exacerbate the input/output problems of a small device discussed in Section 5.1. In addition, the user is often performing another task at the same time as using the mobile device, sometimes related to the mobile device's function (following a GPS map, or talking about a task on the phone) and sometimes unrelated (walking round the supermarket while talking to your mother on your mobile, or texting under the desk in a lecture). Heuristic 5 particularly picks up on these issues. When doing another task it is essential that the user can "get the crucial information from the system by glancing at (*the device*)" and avoid "the need for the user to use both hands".

5.4 The purpose of mobile tasks

Mobile applications can be split into two broad categories:

(i) Those where location matters

(ii) Those where it does not

Many of the heuristics apply to both categories of application, but some apply more to one.

The first category includes location-aware applications such as navigation, tourist information, targeted advertising, and even augmented reality. Heuristic 2 talks about the "Match between system and the real world" and says the system should "sense its environment and adapt the presentation of information accordingly". Heuristic 3 mentions the importance of consistency between users' actions and "the corresponding real tasks (e.g. navigation in the real world)". The ability of applications to achieve these aims is often dependent on hardware and environment, so additional strategies are used, for example the provision of several different landmarks to choose from, given an imprecise GPS location (e.g., Cheverst et al., 2000). Few location-aware systems include electronic compasses or gyroscopes, so directional consistency is particularly hard to maintain.

The second category is not simply the negation of the first, but the opposite: those applications that you specifically want to be able to do anywhere, such as being able to phone, access email, read electronic documents, write in a word processor. In these applications the aim is to unshackle the user from the need to be physically in a particular place. Some of these are where the user wants access to remote resources from anywhere, and in this case issues of data and network connectivity may be important. For other applications in this category the data is local, but issues of screen size, portability and dangers of data loss become more significant.

6. Conclusions and Future Work

In this paper, we have pointed out the benefit of expert-based evaluation methods and their need to capture contextual requirements in mobile computing. We have, in the process, described how we have analyzed mobile usability issues, and discussed our efforts toward realizing a set of usability heuristics that is relevant to mobile computing. Our study confirms previous observations that mobile heuristics detect fewer cosmetic problems and that, in any case, they should not be considered as alternative to user studies, but synergic. In particular, as often noted when speaking of inspection methods, we believe these are useful techniques to use when we are in early phases of design/prototyping, or when the low-cost issue is particularly relevant to the evaluation. As far as the false positives problem is concerned, the inter-expert consistency found when applying mobile heuristics may indicate that the flaws detected were not false alarms, although empirical evaluations with end users are the methods to uncover and solve this issue. As part of our future work, we intend to perform further literature analysis to the work reported in Section 4.1 and possibly consider more dimensions and at different levels of abstraction.

References

Braam, P. J. (1998). The Coda Distributed File System. *Linux Journal, 50*.

Chalmers, M. & Galani, M. (2004). Seamful Interweaving: Heterogenity in the Theory and Design of Interac-tive Systems. In *ACM Designing Interactive Systems – DIS'04*.

Cheverst, K., Davies, N., Mitchell, K., Friday, A. (2000). Experiences of Developing and Deploying a Context-Aware Tourist Guide: The GUIDE Project. In *MOBICOM'2000* (pp. 20-31). ACM Press.

Davies, N., G. S. Blair, K. Cheverst and A. Friday (1996). Supporting Collaborative Applications in a Heterogeneous Mobile Environment, Special Issue of Computer Communications on Mobile Computing, 1996.

Dix, A. J. (1995). Cooperation without (reliable) Communication: Interfaces for Mobile Applications.
Distributed Systems Engineering, 2(3): pp. 171-181.

Dix, A. J. (1999). Hazy, Crazy, Lazy Days are Over - Time for Designers to Think. (keynote) Designing Information for Mobile and Broadband Network Users, London, 15th December 1999, (available at http://www.hiraeth.com/alan/papers/mobile-and-b-99/crazy-lazy-ohps.html).

Cockton, G., Lavery, D., Woolrych, A. (2003). Inspection-based Evaluations. In J.A. Jacko & A. Sears (Eds.), *The HCI Handbook*, Lawrence Erl. Assoc., 2003, pp. 1118-1138.

Greenberg, S., Fitzpatrick, G., Gutwin, C., & Kaplan, S. (1999). Adapting the Locales Framework for Heuristic Evaluation of Groupware. In Conference of the *Computer-Human Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia - OZCHI'99*.

Johnson, P. (1998). Usability and Mobility: Interaction on the Move. In *First Workshop on Human-Computer Interaction with Mobile Devices*.

Kantner, L., & Rosenbaum, S. (1997). Usability studies of www sites: Heuristic evaluation vs. laboratory testing. In *International Conference on Computer Documentation - SIGDOC '97* (pp. 153-160). ACM Press.

Karat, C-M., Campbell, R., & Fiegel, T. (1992). Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation. In Proceedings of *Conference on Human Factors in Computing Systems - CHI'92* (pp. 397-404). ACM Press.

Kistler, J. & Satyanarayanan, M. (1992). Disconnected Operation in the Coda File System. *ACM Transactions on Computer Systems*.

Kjeldskov, J., & Graham, C. (2003). A Review of Mobile HCI Research Methods. In L. Chittaro (Ed.), *International Symposium on Human Computer Interaction with Mobile Devices and Services - Mobile HCI'03* (pp. 317-335). Springer-Verlag.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., & Davies, J. (2005). Evaluating the usability of a mobile guide: The influence of location, participants and resources. *Behaviour and Information Technology, 24(1)*, pp. 51-65.

Law, L. C., & Hvannberg, E. T. (2002). Complementarity and convergence of heuristic evaluation and usability test: A case study of universal brokerage platform. In *Nordic Conference on Human-Computer Interaction - NordiCHI '02* (pp. 71-80). ACM Press.

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM, 33(3)*, pp. 338-348.

Muller, M. J., & McClard, A. (1995). Validating an extension to participatory heuristic evaluation: Quality of work and quality of work life. In *Conference on Human Factors in Computing Systems - CHI'95* (pp. 115-116). ACM Press.

Nielsen, J. (1994a). Enhancing the explanatory power of usability heuristics. In *Conference on Human Factors in Computing Systems - CHI'94* (pp. 152-158). ACM Press.

Nielsen, J. (1994b). *Usability Engineering*. Morgan Kaufmann Publishers.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Conference on Human Factors in Computing Systems - CHI'90* (pp. 249-256). ACM Press.

Po, S. (2003). *Mobile usability testing and evaluation*. Master's thesis, University of Melbourne, Australia.

Sade, S., & Finland, O. N. (2002). Mobile ease-of-use and desirability through user-centered design. In *SSGRR-2002 Summer Conference on Infrastructure for e-Business, e-Education, e-Science, and e-Medicine*.

Simeral, E.J., & Branaghan, R.J. (1997). A Comparative Analysis of Heuristic and Usability Evaluation Methods. *STC Proceedings (1997).*

Vetere, F., Howard, S., Pedell, S., & Balbo, S. (2003). Walking through mobile use: Novel heuristics and their application. In *Conference of the Computer-Human Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia OZCHI2003*.

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R.L. Mack (Eds.), *Usability Inspection methods* (pp. 105-140). New York: John Wiley & Sons.