Appropriating and Assessing Heuristics for Mobile Computing*

Enrico Bertini
DIS
University of Rome
"La Sapienza"
Via Salaria 113 - 00198
Rome, Italy
bertini@dis.uniroma1.it

Silvia Gabrielli
DIS
University of Rome
"La Sapienza"
Via Salaria 113 - 00198
Rome, Italy
gabrielli@dis.uniroma1.it

Stephen Kimani
DIS
University of Rome
"La Sapienza"
Via Salaria 113 - 00198
Rome, Italy
kimani@dis.uniroma1.it

ABSTRACT

Mobile computing presents formidable challenges not only to the design of applications but also to each and every phase of the systems lifecycle. In particular, the HCI community is still struggling with the challenges that mobile computing poses to evaluation. Expert-based evaluation techniques are well known and they do enable a relatively quick and easy evaluation. Heuristic evaluation, in particular, has been widely applied and investigated, most likely due to its efficiency in detecting most of usability flaws at front of a rather limited investment of time and human resources in the evaluation. However, the capacity of expert-based techniques to capture contextual factors in mobile computing is a major concern. In this paper, we report an effort for realizing usability heuristics appropriate for mobile computing. The effort intends to capture contextual requirements while still drawing from the inexpensive and flexible nature of heuristic-based techniques. This work has been carried out in the context of a research project task geared toward developing a heuristic-based evaluation methodology for mobile computing. This paper describes the methodology that we adopted toward realizing mobile heuristics. It also reports a study that we carried out in order to assess the relevance of the realized mobile heuristics by comparing their performance with that of the standard/traditional usability heuristics. The study yielded positive results in terms of the number of usability flaws identified and the severity ranking assigned.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems

*This work has been supported by the MAIS project (http://www.mais-project.it).

AVI '06, May 23–26, 2006, Venezia, Italy.

General Terms

Design, Human Factors

Keywords

Usability heuristics, mobile computing, heuristic evaluation

1. INTRODUCTION

Mobile computing presents formidable challenges in terms of design and evaluation methodologies needed to assist every phase of a system's lifecycle. In particular, here we take into account some main difficulties the HCI community is facing when it comes to evaluate mobile computing.

Evaluation methods in HCI have often relied on measures of task performance and task efficiency as a means of evaluating applications. While such primarily task-centric evaluation approaches may be applicable to the desktop computing paradigm, where tasks are often structured and relatively predictable, such evaluation approaches may not be directly applicable to the often unpredictable, rather opportunistic and relatively unstable mobile settings. Moreover, it is not trivial for evaluation methods to fulfill the pressing need to properly integrate real-world setting/context (or simulated context) during the evaluation process. Context entails aspects such as: location, infrastructure/resources (server and network capabilities and connections, applications), user (user data, usage patterns), environment, entities (people, devices, objects), and time (date, time of the day, season)[4].

As far as user testing is concerned, mobile computing demands not only: real users, real (or simulated) context and device interaction tasks; but also real tasks (or realistic simulations of the same). The design implementation or prototyping should therefore be able to support such real tasks in real settings (or their simulations), thus be robust and reliable enough to support the foregoing. This is not easy in mobile computing due to the fact that technology is often on the cutting edge [5], and new solutions need to be assessed 'early and often' [10].

Expert-based evaluation techniques (such as heuristic evaluation and cognitive walkthrough) are well known methods that often enable a relatively quick and easy evaluation. Heuristic evaluation, in particular, has been widely applied and investigated, most likely due to its efficiency in detecting most of usability flaws (75-80%) at front of a rather limited investment of time and human resources in the evaluation

(typically, 3-5 experts) [13] [12].

However, research in mobile computing points out that there are concerns on the capacity of expert-based techniques to adequately capture the multiple contextual factors that affect user-system interactions in real settings (for instance: [9], [7]). Inspection methods are often criticized for their ability to detect a fewer number of problems in total, as well as a very high rate of cosmetic ones [8]. Recent attempts to overcome some of these difficulties, in particular the 'context immunity' of these methods have been made [14]. Inspired by the foregoing, we do believe that heuristic evaluation can be enriched and adapted toward better taking into account some of the contextual factors that are relevant in mobile computing. Also, previous studies indicate that tailoring the heuristics to the mobile context reduces the problem of detecting cosmetic flaws or false positives in the data [15]. In this paper, we report an effort for realizing usability heuristics more suited for being used in mobile computing. The rationale of this work is to enable heuristic evaluation to capture contextual requirements while still drawing from the inexpensive and flexible nature of the method. Our initial findings indicate that when compared with traditional usability heuristics, the proposed mobile heuristics better supported experts in detecting usability flaws and providing adequate severity ranking of the same.

The rest of the paper is organized as follows: Section 2 presents a brief state of the art of past and current studies related to our work. In Section 3, the paper describes the methodology that we adopted to collect and analyze relevant mobile usability issues, as well as it presents the main findings obtained. In Section 4, we describe the process through which we came to develop a set of usability heuristics specifically dedicated to mobile computing evaluations. We also present a study aimed at assessing the benefits (if any) of applying the mobile usability heuristics proposed. In Section 5 we discuss our conclusions and highlights some future work in this research.

2. RELATED WORK

In recent years some attempts have been made toward adapting Nielsen's original heuristic evaluation method to better fit the requirements of specific application domains as well as to increase its ecological validity [1], [6], [15], [11], [14].

In [1] the authors propose a set of heuristics specifically designed to identify usability problems of groupware systems. The heuristics development process was based on the Locales Framework [6] a previous articulated approach to comprehensively evaluate collaborative environments and how they co-exist with the groups' everyday practices of communication and collaboration. In particular, the heuristics proposed try to focus experts evaluation on the system's capacity of supporting typical group activities such as communication, coordination, planning, monitoring, assistance, and protection. Although the approach adopted for developing the heuristics is theoretically and empirically sound, unfortunately no formal validation of the heuristics was provided by this study.

In [15] the Locales Framework was used as inspiration to the development of novel heuristics for evaluation of mobile use, by taking into account social aspects of mobility (such as awareness, presence, privacy, etc.) as a way of complementing the application of more traditional usability heuristics. In this study a first attempt to evaluate the new heuristics proposed was made, although experts applied them in the context of an hybrid technique (called Heuristic Walkthrough) that combined both aspects of Heuristic evaluation and Cognitive Walkthrough.

Another relevant work has been presented in [11] whose objective was to specialize Nielsen's traditional heuristics to the evaluation of ambient displays. Interestingly, in this study authors adopted a mixed and multi-step strategy to come up with the final set of heuristics proposed. The strategy was based on a series of internal and external refinement cycles operated on the set of heuristics, which involved both domain experts (like display designers) and usability experts not previously involved in their development.

A different but related stance was taken by [14] toward improving heuristic evaluation validity when applied to assess systems as they are used in real world settings. The authors did not modify Nielsen's usability heuristics [12], but experimented with two slight variations of the method, such as Heuristic Walkthrough (this time combining standard heuristic evaluation with scenarios of use provided to experts) and Contextual Walkthrough (entailing experts performance of the foregoing Heuristic Walkthrough in the field). As a main finding the study reports that both variations of the traditional heuristic evaluation performed better in helping experts to identify a larger number of usability flaws, although Contextual Walkthrough technique turned out to be very time demanding, thus it seems not to be adequate in retaining the low cost benefits of usability inspection methods.

As far as usability principles for mobile computing are concerned, there already are various efforts in that direction. In [2], Bertini et al. propose a review of commonly accepted standard usability principles [5] by analyzing how the principles apply in mobile computing and proposing appropriate revisions. In [3], Buchanan et al. propose a set of mobile usability principles which can be used by WAP service providers to improve user experience on small screen mobile devices.

By analyzing the contributions provided by the studies reported above we realized the need for a deeper investigation and adaptation of heuristic evaluation technique to make it suitable for application in the field of mobile computing. In fact, no previous attempts have specifically explored in depth this dimension.

3. METHODOLOGY FOR REALIZING MOBILE ISSUES

Toward developing usability heuristics for mobile computing, we found it worth analyzing usability issues in mobile computing. To that end, the first three authors of this paper worked as usability researchers at the following activities:

1. Each one of the 3 usability researchers was assigned a unique set of papers to analyze independently. The papers originated from the list used in [9]; a recent meta-analysis of hci research methods in mobile hci [1]. We updated the list with papers published in the period 2004-2005 and selected only those with elements

[1]The papers were selected from top-level conferences and journals like CHI, AVI, UIST, TOCHI, etc., see [9] for details.

of evaluation. The analysis entailed documenting for each of the papers, the appropriate values/inputs for the following dimensions: evaluation goal, evaluation method, evaluation setting, real device/emulator, location matters. Moreover, each of the usability researchers individually documented mobile usability issues that were indicated by (or evident from) each of the papers.

2. In the next step, the usability researchers came together and consolidated their individual realizations. This was done in the form of a spreadsheet. The usability researchers as a team then refined the usability issues and also merged similar issues.

3. Each of the researchers was then given the same realized list of mobile usability issues and asked to independently as an individual categorize (group or cluster) the issues. The usability researchers then came together, presented and exchanged their individual results of categorizations. With reference to the other categorizations, each of the researchers was requested to individually work further on his/her categorization by: eliminating redundant usability issues, clarifying the mobile usability issues, and also grouping the obtained issues to an abstraction level that would be appropriate for developing/generating heuristics. The usability researchers then came together and shared their individual realizations. Through brainstorming, they came up together and consolidated their work. They also as a team harmonized the terminology used to describe the issues.

4. 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.

4.1 Toward a set of heuristics

The brainstorming activity described in Section 3 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.

4.1.1 Phase 1

Each of the 3 usability researchers was provided with a table reporting Nielsen's traditional heuristics [12] 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 application?", that 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 3).

4.1.2 Phase 2

Each of the usability researchers compared her/his own table of heuristics proposed 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.

4.1.3 Phase 3

A new refinement process was started on the set of heuristics included into the three consolidated tables produced in phase 2. It involved: first, a discussion meeting among the usability researchers to arrive at a shared consolidated table of the three ones 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, for receiving feedback on the adequacy of the heuristics proposed. We contacted 19 experts: in person, by email or by phone calls. We managed to get feedback from 8 of them. In a following meeting, the 3 usability researchers discussed and compared the comments received from experts 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 the data should be taken to minimize loss. If the device is misplaced, the device, system or application should make it easy to find it back.

- Heuristic 2 - Match between system and the real world: Enable the mobile user to interpret correctly the information provided, 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 which 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 user's data are 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 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.

4.2.1 Experimental Design

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

Participants and Materials. The study involved enlisted 8 usability experts [2], 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 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 primary 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. [3]

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 [12], our proposed set of mobile usability heuristics (Table 1), and Nielsen's five-point Severity Ranking Scale (SRS) [12] (which is described in Table 2).

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 assigned to one of the foregoing two experimental conditions. They all had previous expertise in the HCI evaluation field and were familiar with both the application of traditional HE method and the use of mobile applications. Notwithstanding this, 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 evaluation method; the collected demographic data can be seen in Table 3. Most of the participants have a high level 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 of usability heuristics 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). While evaluating the

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

[3]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.

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

Table 1: Mobile usability heuristics

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

Table 2: Nielsen's severity ranking scale (SRS)

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 down by one of the evaluation moderators.

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

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

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

4.2.2 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 (reduced standard deviations for individual applications and for both applications evaluated using the mobile heuristics). By 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 overlappings (problems pointed out by more than one expert), which support the idea of inter-expert consistency when applying mobile heuristics.

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

Table 4: Number of usability problems identified

Severity of Flaws and Distribution. As depicted by Table 5 and Figure 1, 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 1, it does appear that Nielsen's heuristics do not reflect very high values for Minor and Major flaws when compared to the mobile heuristics. It does also seem that Nielsen's heuristics have a relatively well equally distributed 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 rather than those at the extremes.

Figure 2 can be used for further analysis on how the spe-

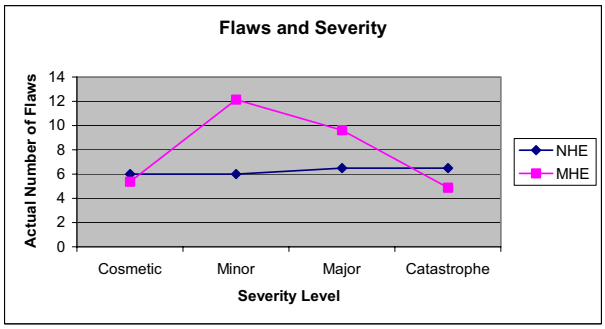

Figure 1: Actual number of flaws, severity and distribution

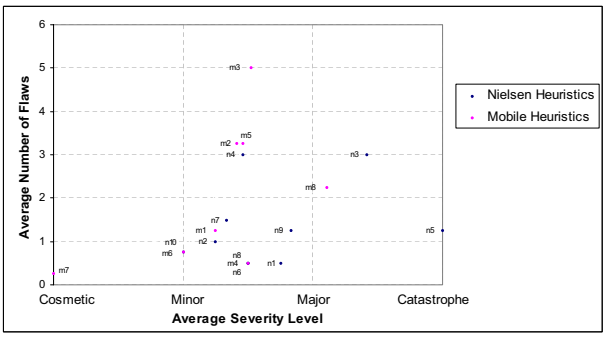

Figure 2: Comparison of the sets of heuristics: flaws and severity

cific heuristics from both sets fair with regard to average severity and average number of flaws. Figure 2 indicates that: mobile heuristics are more effective in supporting the detection of flaws, Nielsen's heuristics seem better suited to cover the case in which high severity flaws are present, and 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 2 are similar to those from Figure 1.

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 may specialize a bit the evaluation to the mobile sector instead of focusing experts' attention at a more general level (although the kind of setting we used in this study was promoting a more functionalities-based evaluation of applications than a contextual one). Moreover, the mobile heuristics could be applied when/where the extreme flaws have been addressed or are not an issue in the design. In case 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 1).

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[4]:

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

It is interesting to observe that foregoing Nielsen's heuristics (4 [Consistency and standards], 3 [User control and freedom]) are related to the foregoing 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 foregoing Nielsen's and mobile heuristics as the most violated or appreciated ones, although recur to them could be due to the particular type of evaluation/application(s) that was 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 4.1.3.

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, etc).

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 experts' less familiarity with these heuristics compared to Nielsen's heuristics. We tried to reduce (for it cannot be eliminated simply in an evaluation session) this familiarity issue by giving experts who were using the mobile heuristics some extra time at the beginning of the evaluation to study the mobile heuristics, to familiarize themselves with the mobile heuristics and to ask any questions on the mobile heuristics. 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

[4]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).

	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	

Table 5: Actual number of flaws and severity

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 6: Nielsen's 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 device losability/findability
6	3	Flexibility, efficiency of use and personalization
4	2	Good ergonomics and minimalist design
7	1	Aesthetic, privacy and social conventions

Table 7: Mobile heuristics and corresponding usability problems

heuristic evaluation is an evaluation technique strongly dependent on experts' previous knowledge and expertise with the heuristics, the application domain, etc.

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

Table 8: Time taken in minutes

As a general observation, it is worth mentioning that because our study adopted a between-subjects design, there is the 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 commendable.

5. 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 also 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 less 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. We are aware that these methods also entails risks, like the possible occurrence of false positives among the flaws detected in the experts evaluation. Inter-expert consistency, which was quite supported by our findings, is a valuable indication that can mitigate this risk, 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 3 and possibly consider more dimensions and at different levels of abstraction.

6. ACKNOWLEDGMENTS

We would like to thank those who were involved in the review of our initial proposal and those who participated in our study as evaluators.

7. ADDITIONAL AUTHORS

Additional authors: Tiziana Catarci (DIS - University of Rome "La Sapienza", email: `catarci@dis.uniroma1.it`) and Giuseppe Santucci (DIS - University of Rome "La Sapienza", email: `santucci@dis.uniroma1.it`).

8. REFERENCES

[1] K. Baker, S. Greenberg, and C. Gutwin. Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration. In *EHCI '01: The IFIP International Conference on Engineering for Human-Computer Interaction*, pages 123–140. Springer-Verlag, 2001.

[2] E. Bertini, T. Catarci, S. Kimani, and A. Dix. A review of standard usability principles in the context of mobile computing. *Studies in Communication Sciences*, 5(1):111–126, 2005.

[3] G. Buchanan, S. Farrant, M. Jones, H. Thimbleby, G. Marsden, and M. Pazzani. Improving mobile internet usability. In *Intl. World Wide Web Conf. (WWW)*, 2001.

[4] A. Dey. *Providing Architectural Support for Building Context-Aware Applications*. PhD thesis, College of Computing. Georgia Tech, December 2000.

[5] A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human-Computer Interaction*. Prentice Hall, 2004.

[6] S. Greenberg, G. Fitzpatrick, C. Gutwin, and S. Kaplan. 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*, 1999.

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

[8] C.-M. Karat, R. Campbell, and T. Fiegel. Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Conference on Human Factors in Computing Systems - CHI'92*, pages 397–404. ACM Press, 1992.

[9] 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.

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

[11] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic Evaluation of Ambient Displays. In *Conference on Human Factors in Computing Systems - CHI'03*, pages 169–176. ACM Press, 2003.

[12] J. Nielsen. Heuristic evaluation. pages 25–62, 1994.

[13] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *CHI '90: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256, New York, NY, USA, 1990. ACM Press.

[14] S. Po, S. Howard, F. Vetere, and M. B. Skov. Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap. In S. Brewster and M. Dunlop, editors, *Conference on Human Computer Interaction with Mobile Devices and Services - Mobile HCI'04*, pages 49–60. Springer-Verlag, 2004.

[15] F. Vetere, S. Howard, S. Pedell, and S. Balbo. 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*, 2003.