

Usability Evaluation of context-aware mobile systems: A review

Stavros Garzonis
Department of Computer Science
University of Bath, UK
Bath BA2 7AY
s.garzonis@cs.bath.ac.uk

Abstract

It is widely acknowledged that efforts to adapt the desktop PC's usability evaluation methods to the mobile environment are limited [4, 7]. In particular, there is a growing debate in the literature regarding the tradeoffs between lab and field usability of mobile devices and services. This review presents findings of research attempts to compare the two methods and draws conclusions from the lessons learned.

Mobile Usability Evaluation

Although the strengths and weaknesses of usability evaluation methods have been measured against desktop environments, research in the pervasive and mobile environment is very limited. Some researchers have addressed the challenges of interface constraints [e.g. 2], while others have addressed the dynamic context of use [e.g. 3, 10]. However, the rise of context-aware devices and services introduces a new level of complexity when it comes to usability evaluation.

When attempting to define the attributes of an evaluation method for context-aware mobile devices, the most significant choices are the method's where, who and how. The place of evaluation in HCI has traditionally been in labs as the control of environmental conditions is important for accurate measuring of performance variation. Other factors that also support the choice of lab evaluation are the ease of data collection, the equipment availability and the ease in the procedure of recruiting participants. On the other hand, ethnography has traditionally placed evaluations in the field, which is the actual environment where people will make use of the technology evaluated. Although there might be difficulties in translating the observation results into design suggestions, the field evaluation can provide a deeper insight in user requirements and issues with the everyday use of the technology.

Furthermore, evaluators need to decide whether the evaluation should include end user participation and then choose from a variety of methods, such as Heuristic Evaluation, Cognitive Walkthrough, Co-operative evaluation. Finally, the type of observation, the type of data recorded and the kind of analysis performed are important decisions for an evaluation.

When it comes to mobile usability evaluation, all these decisions need to be made under the particular characteristics of mobile use. First, users who are

mobile are more likely to have to divide their attention among elements of the environment (e.g. passing cars) and the technology at use. Secondly, the evaluation might be hindered by, or focus on, the dynamic context of use, such as ongoing activities, social and technological environment. Thirdly, the constraints of the mobile equipment in software and hardware capabilities are limiting researchers' options in evaluations.

Although research on evaluation methods of mobile services and devices are still limited, one can draw from several areas of the literature. For example, some researchers have focused on general guidelines on field evaluations [e.g. 1, 9] while others have evaluated novel interfaces and ways of interacting [e.g. 2, 8]. Perhaps one of the rare examples of adjusting an evaluation method to fit the mobile setting particularities was attempted by Vetere et al. [10]. They proposed a hybrid evaluation method comprised of cognitive walkthrough and heuristic evaluation, with added heuristics to capture the social aspects of mobility. However, the evaluation of the method showed that evaluators had no sense of the real world context and that current heuristics provide more concrete results and redesign implications than the extended mobile heuristics. The remaining of the paper focuses on examples of research done on the in situ – in vitro debate of mobile evaluation.

To field or not to field?

It has been reported [7] that only 41% of the surveyed research on mobile human-computer interaction between 2000 and 2002 involved evaluation of system designs and 71% of it was performed in labs. So, is it worth to spend the extra resources needed to perform the evaluations on the field just because this is the natural setting where the technology will be used?

Kjeldskov et al. [5] evaluated a mobile guide to support the use of public transportation with 4 distinct evaluation methods: field evaluation, lab evaluation, heuristic walkthrough [10] and rapid reflection. They categorised the usability problems as critical (stopped users from completing the tasks), serious (inhibited/slowed down the users from completing the tasks) and cosmetic (did not inhibit users from completing the tasks). The results showed that field and lab evaluations had a significant overlapping in critical and serious problems, although field was slightly more efficient in serious problems. However, field studies

were shown to be the least effective method in identifying cosmetic problems. The lessons reported by the researchers included that “the identification of critical problems depended little on the precise circumstances surrounding the deployment of a specific approach”, “laboratory-based approach drew attention to device-oriented issues” and “field approach drew attention to issues such as the real-world validity and precision of the data presented by the system and the ‘social comfort’”.

In a different study [6] six evaluation techniques were compared for their effectiveness in evaluating the use of a mobile device. One of them involved walking in a pedestrian street and the remaining five where in the lab: sitting at a desk, walking on a treadmill at constant or varying speed, walking at a constant or varying speed on a course that was constantly changing. The results indicate that the sitting down technique was the most effective but largely because of its effectiveness in identifying cosmetic problems. On the other hand, techniques involving movement were more effective in finding user interface layout problems. Overall, the authors report that “there were no significant differences between the techniques in terms of user performance” but the field technique exhibited significantly more mental workload in terms of perceived effort and overall workload.

Finally, a study [5] comparing field and lab evaluations for a context-aware mobile electronic patient record system prototype reported the lab evaluation as more effective. However, the difference of effectiveness of the methods was non-significant for critical problems. In particular, lab evaluation identified 13% more critical problems, 42% more serious problems and 40% more cosmetic problems, while both methods identified all context-aware related problems. We should, however, note that the evaluation time in the lab lasted twice as much as in the field per participant and the nurses participating in the field evaluation were less experienced.

Conclusion

Attempts to adapt usability evaluation methods for mobile devices and services are limited. One of the issues that has been addressed by the research community is the tradeoff between lab and field mobile evaluation. Although the studies carried out are limited and involve a small number of participants, they seem to agree on certain points. First, lab evaluations are more efficient in identifying cosmetic problems, which do not hinder interaction and user performance. Second, field evaluation is more likely to identify issues that are related to the real context of use, such as navigation and social comfort.

Perhaps it is time for a new hybrid approach to be introduced, in which users might perform cooperative evaluation sessions [8] in real world contexts, such as a café. Through this approach we would anticipate to find most of the device specific problems that arise from lab

evaluation but also preserve the evaluation’s ecological validity.

References

1. Borntager, T., Jameson, A., Kruger, A. and Butz, A., [2002], Location-aware shopping assistance: evaluation of a decision-theoretic approach., *in* Fabio Paternò (ed.), *Proceedings of the Mobile HCI 2002*, Springer-Verlag, 155-169.
2. Brewster, S., [2002], Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, **6**(3) 188-205.
3. Cheverst, K., Mitchell, K. and Davies, N., [2002], Exploring Context-aware Information Push. *Personal and Ubiquitous Computing*, **6**(4), 276-281.
4. Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S. and Davies, J., [2005], Evaluating the Usability of a Mobile Guide: The influence of Location, Participants and Resources. *Behaviour and Information Technology*, **24**(1), 51-65.
5. Kjeldskov, J., Skov, M. B., Als, B. S., and Høegh, R. T., [2004], Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field, *in* Brewster, S. and Dunlop, M. (eds.), *Proceedings of the 6th International Mobile HCI 2004 conference*, Springer-Verlag, 61-73.
6. Kjeldskov, J., and Stage, J., [2004], New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies (IJHCS)*, **60**(5-6), 599-620.
7. Kjeldskov, J., and Graham C., [2003], A Review of Mobile HCI Research Methods., *in* Chittaro, L. (ed.), *Proceedings of the 5th International Mobile HCI 2003 conference*, Springer-Verlag, 317-335.
8. Pirhonen, A., Brewster, S.A. and Holguin, C., [2002], Gestural and audio metaphors as a means of control for mobile devices, *in* Terveen, L. (ed.), *Proceedings of the CHI '02 Conference on Human Factors in Computing Systems*, ACM Press, 291-298.
9. Rowley, D. E., [1994], Usability Testing in the Field: Bringing the Laboratory to the User, *in* Adelson, B., Dumais, S., & Olson, J. S. (eds.), *Proceedings of the CHI'94*, ACM Press, 252 - 257.
10. Vetere, F., Howard, S., Pedell, S. and Balbo, S., [2003], Walking through mobile use: novel heuristics and their application., *in* Viller, S. and Wyethh, P. (eds.), *Proceedings of the OzCHI 2003*, 24-32.