Comparing Usability Evaluations of Mobile Systems

Benedikte S. Als, Rune T. Høegh, Jesper Kjeldskov,
Mikael B. Skov and Jan Stage

Department of Computer Science, Aalborg University, Denmark

1. Introduction

Established approaches to design and evaluate usable systems are challenged by systems for wearable, handheld, or mobile devices. There are extensive methods and guidelines that describe how the usability of stationary computer systems should be evaluated [4,6]. This is complemented with experimental evaluations of the relative strengths and weaknesses of different techniques [1,3]. However, in relation to design and evaluation of mobile systems, such methods, guidelines, and experimental evaluations are yet to be produced.

Mobile systems are typically used in highly dynamic contexts and their use often involve several people distributed in the user's physical surroundings. Therefore, field-based testing seems like an appealing or even indispensable approach for evaluating mobile systems. Yet usability testing in the field is difficult. Firstly, it can be complicated to establish realistic studies that capture key situations in the use-context. Secondly, it is far from trivial to apply established techniques such as observation and think-aloud in the field. Thirdly, field-testing complicate data collection and limits control since users are moving physically in an environment with a number of unknown variables. When usability tests are conducted in a laboratory setting, control and collection of high quality data is not a problem, but one of the drawbacks is the lack of realism. Existing approaches to laboratory-based usability testing of stationary computer systems try to solve this problem by recreating or imitating the real use context, e.g. by office furnishing [6]. However, when mobile systems are tested in a laboratory setting, activities in the user's physical surroundings can be difficult to recreate realistically [5].

We explore laboratory and field-based approaches to usability evaluation of mobile systems through a number of comparative usability studies involving different experimental design. Two of these studies are illustrated below and the experiences from these studies are compared.

2. TramMate

In early 2003, we designed a context-aware mobile information service (TramMate) [2]. This service supports the use of the tram based public transport system of Melbourne by keeping track of contextual factors such as the user's physical location, upcoming appointments and real time tram information. The design is integrated with an electronic calendar. We designed and conducted two usability evaluations of an early prototype. The first evaluation was conducted in the field. The second evaluation was conducted in a usability laboratory. The two evaluations were identical in terms of tasks and the profiles of the test subjects. The users had to complete three tasks involving route planning prior to catching a tram. All tasks were realistic and achievable within the time frame. Five subjects participated in the experiments in the field and laboratory respectively. Half the users were male and the other half were female, balanced across the field and laboratory studies. Users were aged between 21 and 42 and were all frequent computer users and familiar with the tram system of Melbourne.
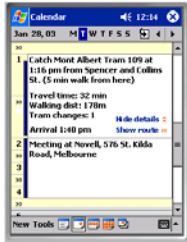

Figure 1: TramMate   Figure 2. Field evaluation   Figure 3. Lab evaluation

Field Evaluation

The field study focused on use of the prototype in realistic surroundings. In this study, the users had to both look up necessary information on the mobile device according to the tasks and then perform the tasks "for real" (e.g. catching a tram to a specific destination). The prototype accessed live timetable information via the Internet but GPS positioning was simulated. During the evaluation, three researchers observed the user: an evaluator encouraged the user to think-aloud, one took notes and one recorded the evaluation on a handheld camcorder (figure 2).

Laboratory Evaluation

In the second study, the user was only required interact with the prototype system. The user was seated at a desk, with the mobile device in his hand. An evaluator was seated next to the user and encouraged him to think-aloud. The usability laboratory facilitated video recordings of the display of the

mobile device and overall views of the test subject and the evaluator (figure 3). To ensure a good view of the mobile device, the user was requested to hold it within a limited area indicated on the table. Two researchers observed the evaluation through a one-way mirror. One took notes. The other operated the video equipment.

3. MobileWARD

During a five months project we designed MobileWARD, a context-aware mobile system running a PDA supporting work at a Danish hospital ward. MobileWARD is context-aware as it automatically keeps track of e.g. physical location of patients and staff, upcoming appointments and schedules. Physical location was simulated through a control unit operated by the participating researchers. We designed and conducted two different usability evaluations of the system. The evaluations were similar as they involved trained, registered nurses as test subjects, and they should conduct standard morning work routines. However, they were different in their data collection. The participating subjects were between 27 and 54 years old and they had diverse experiences with nursing. All of them were novices with PDAs.
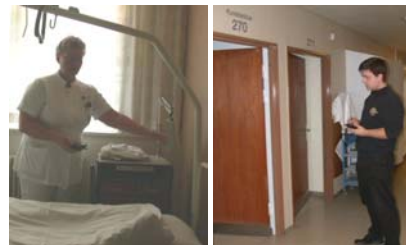

Figure 4: Laboratory evaluation   Figure 5: Evaluation at the hospital

Field Evaluation

The field evaluation focused on using the system in a realistic environment. The evaluation took place during morning procedure at a hospital ward. Prior the evaluation, we entered data on the committed patients at the ward. The use of the system was not controlled by task assignments. The test subject should merely conduct her standard morning procedure (figure 5). Three committed patients were involved in the morning procedure at the day of our evaluation. We conducted an interview with the test subject afterwards to identify opportunities and limitations of the mobile system.

Laboratory Evaluation

The laboratory evaluation took place at the usability laboratory at Aalborg University. The idea of this evaluation was to evaluate the mobile system in

an environment where we could closely monitor all actions and situations (figure 4). Three test subjects participated in the study and they were instructed through assignments and were told to think-aloud while using the system. Three students acted as patients for the evaluation. One researcher acted as test monitor while another controlled the equipment from the control room.

4. Lessons Learned

- Collection of sufficiently detailed and precise data in the field is difficult because of motion and work-related conditions, e.g. usability evaluators could not follow the nurses into the ward.
- Field testing requires several test monitors and loggers, it takes a considerable amount of time, and it is physically demanding.
- The largest number of usability problems is detected when test subjects are sitting down by a desk in the laboratory.
- The problems that are detected in the laboratory focus on interaction and interface design.
- The problems that are detected in the field focus on the relation between the system and the real world that it is supposed to have awareness about.
- A laboratory set-up that imitates the real context of work facilitates detection of problems that do not occur when sitting at the desk, e.g. when the system changes screen because of motion to a different context, the nurses think they have done something wrong.

References

1. Karat, C. M., Campbell, R. and Fiegel, T. (1992) Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation. In *Proceedings of CHI'92*, pp. 397-404. New York: ACM.
2. Kjeldskov J., Howard S., Murphy J., Carroll J., Vetere F. and Graham C. (2003) Designing TramMate - a context aware mobile system supporting use of public transportation. Proceedings of DUX 2003.
3. Molich, R., Bevan, N., Curson, I., Butler, S., Kindlund, E., Miller, D. and Kirakowski, J. (1998) Comparative Evaluation of Usability Tests. Proceedings of the Usability Professionals Association Conference.
4. Nielsen, J. (1993) *Usability Engineering*. Morgan Kaufmann.
5. Pirhonen, A., Brewster, S. A. and Holguin, C. (2002) Gestural and audio metaphors as a means of control for mobile devices. Proceedings of CHI'2002, pp. 291-298.
6. Rubin J. (1994) *Handbook of Usability Testing*. Wiley.