

Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field

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

Aalborg University, Department of Computer Science
Fredrik Bajers Vej 7E, DK-9220 Aalborg East, Denmark
{jesper, dubois, als, runethh}@cs.auc.dk

Abstract. Evaluating the usability of mobile systems raises new concerns and questions, challenging methods for both lab and field evaluations. A recent literature study showed that most mobile HCI research projects apply lab-based evaluations. Nevertheless, several researchers argue in favour of field evaluations as mobile systems are highly context-dependent. However, field-based usability studies are difficult to conduct, time consuming and the added value is unknown. Contributing to this discussion, this paper compares the results produced by a laboratory- and a field-based evaluation of the same context-aware mobile system on their ability to identify usability problems. Six test subjects used the mobile system in a laboratory while another six used the system in the field. The results show that the added value of conducting usability evaluations in the field is very little and that recreating central aspects of the use context in a laboratory setting enables the identification of the same usability problem list.

1 Introduction

In the proceedings of the first workshop on Human-Computer Interaction for Mobile Devices in 1998, researchers and practitioners were encouraged to investigate further into the criteria, methods, and data collection techniques for usability evaluation of mobile systems [8]. Of specific concerns to the development of such methods and techniques, it was stated that traditional usability laboratory setups would not adequately be able to simulate the context surrounding the use of mobile systems and that evaluation techniques and data collection methods such as think-aloud, video recording or observations would be extremely difficult in natural settings. These concerns have since been confirmed through a number of studies e.g. [5, 6, 7, 9, 16, 18].

In 2003, a literature study on mobile HCI research methods revealed that 41% of mobile HCI involved evaluation [10]. However, even though evaluations of mobile systems are prevalent, surprisingly little research has been published concerning the methodological challenges described above. Exceptions include studies comparing methods applied for evaluating mobile systems in e.g. [5, 7, 9, 11, 17]. Consequently, no agreed upon set of appropriate usability evaluation methods and data collection techniques yet exists within the field of mobile HCI.

While the literature study [10] also revealed that 71% of mobile device evaluation was done through laboratory experiments and only 19% through field studies, it seems implicitly assumed that usability evaluations of mobile devices *should* be done in the field [1, 5, 8]. But field-based usability studies are not easy to conduct. They are time consuming and the added value is questionable [6, 9, 14, 16, 18]. Motivated by this, it has been suggested that instead of going into the field when evaluating the usability of mobile devices, requiring mobility or adding contextual features such as scenarios and context simulations to laboratory settings can contribute to the outcome of the evaluation while maintaining the benefits of a controlled setting [4, 9, 11, 12, 17, 20].

More emerging mobile systems are being characterized as context-aware as *they incorporate the ability of an application to discover and react to changes in the environment* [21]. Abowd and Mynatt state that the strong link to the physical context of a context-aware mobile systems challenge even further the conceptions of usability evaluations as the *scaling dimensions that characterize context-aware systems makes it impossible to use traditional, contained usability laboratories* [1]. They continue by stating the effective usability evaluations require realistic deployment into the environment of expected use [ibid.]. However, we still have little knowledge about the relative strengths and weaknesses of laboratory-based versus field-based usability evaluations of context-aware mobile systems.

This paper has two purposes. Firstly, we want to compare the outcome of evaluating the usability of a mobile system in a laboratory setting and in the field in relation to identified usability problems and time spent on conducting the evaluations. Secondly, we want to describe two techniques used for 1) improving the realism of laboratory settings by including mobility and context, and 2) supporting high-quality video data collection when evaluating usability of mobile devices in the field.

2 Experimental Method

To address the above issues, we conducted a study involving two usability evaluations of a context-aware mobile electronic patient record (EPR) system prototype. The two evaluations involved a total of 12 professional nurses as test subjects conducting standard morning work routine activities. The first evaluation took place in a state-of-the-art usability laboratory where the subjects performed a series of assigned tasks while thinking-aloud. The second evaluation took place at the Hospital of Frederikshavn involving real work activities.

Studying usability of mobile systems for hospital settings takes the challenges of laboratory and field evaluations to an extreme. The users are typically highly mobile and the work pace is often intense and stressful. Furthermore, work activities are safety-critical (with errors potentially endangering the wellbeing or life of patients) and involve several ethical considerations such as privacy.

Recreating a healthcare context in a usability laboratory can be extremely difficult even impossible as such healthcare contexts integrate very complex work procedures, work situations, and tools. Recent studies on the usability of mobile information systems in healthcare have employed an indirect approach to data-collection about usability-

ity through interviews and questionnaires about user-friendliness and user-satisfaction with a prototype system [2, 23]. While overcoming some of the challenges described above, this approach does not provide first-hand insight into user-interaction.

2.1 The Context-Aware Mobile System Evaluated: MOBILEWARD

Based on evaluations of stationary electronic patient record (EPR) systems and field studies of mobile work activities in hospitals, we implemented a context-aware mobile EPR prototype called MOBILEWARD [22]. MOBILEWARD runs on a Microsoft PocketPC based Compaq iPAQ 3630 connected to an IEEE 802.11b wireless TCP/IP network. The system was programmed in Microsoft embedded Visual Basic 3.0.

MOBILEWARD is designed to support planning and conducting work tasks during morning procedure at a hospital department. The system is context-aware in the sense that the system presents information and functionality adapted to the location of the nurse, the time of the day, and the conditions of the patients. Based on the classification by Barkhuus and Dey [3], MOBILEWARD is an *active* context-aware system as it automatically presents information and adapts to the context.

Before visiting assigned patients for morning procedure, nurses often want to get an overview of the specific information about each patient. As this typically takes place at the nurse's office or in the corridor, the system by default displays the overall patient list (figure 1a). Patients assigned for morning procedure are shown with a white background and the names of patients assigned to the nurse using the system are bold-faced (e.g. "Julie Madsen"). For each patient, the patient list provides information about previous tasks, upcoming tasks and upcoming operations. The indicators TP (temperature), BT (blood pressure) and P (pulse) show the measurements that the nurse has to perform. "O" indicates an upcoming operation (within 24 hours), which usually requires that the patient should fast and be prepared for operation. At the top of the screen, the nurse can see their current physical location (e.g. "in the corridor").



Fig. 1. MOBILEWARD: Three different screen layouts of the context-aware mobile EPR system

The window in figure 1b displays information related to one patient including name and personal identification number of the patient, previous sets of measured temperatures, blood pressures, and pulses as well as notes regarding the treatment of the patient. To enter new data into the system, the nurse must scan the barcode identification tag on the patient's wristband using the "scan" function in the bottom of the screen. When the nurse enters a ward, the system automatically displays information and

functionality relevant to this location (figure 1c). Information about the patients on the current ward is presented, resembling the information available on the patient list displayed in the corridor, with the addition of a graphical representation of the physical location of the patient's respective beds. Data on each patient is available by clicking on the names.

In the evaluated prototype of MOBILEWARD, some of the contextual sensing functionality was simulated by means of a "context control centre" application. The control centre runs on a separate iPAQ connected to the wireless network. Through this application, an operator can trigger "context events" in MOBILEWARD, e.g. instructing the system that the user has entered a specific room.

2.2 Laboratory Evaluation

The idea of the laboratory evaluation was to evaluate MOBILEWARD in a controlled environment where we could closely monitor the use of the system. In addition to this, we also wanted to extend the standard experimental setup to include mobility and context. In order to achieve this, we modified the standard laboratory setup in a number of ways. The laboratory evaluation is described in detail below.

Setting. The usability laboratory was set up to resemble a part of the physical space of a hospital department (figure 3 and 4). This included the use of two separate evaluation rooms connected by a hallway. Each of the evaluation rooms were furnished with beds and tables similar to real hospital wards. From a central control room, the evaluation rooms and the hallway could be observed through one-way mirrors and via remotely controlled motorized cameras mounted in the ceiling.



Fig. 2. Wireless camera mounted on PDA



Fig. 3. Video images from furnished subject rooms

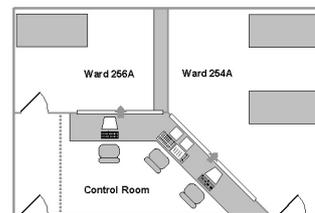


Fig. 4. Physical layout of the usability laboratory

Data collection. High quality audio and video data from the laboratory evaluation was recorded digitally. A tiny wireless camera was clipped on to the mobile device (figure 2), providing us with a close-up view of the screen and user-interaction. This was then merged with the video signals from the ceiling-mounted cameras (figure 2).

Test subjects. Six test subjects (four females and two males) aged between 28 and 55 years participated in the study. All test subjects were trained nurses employed at a large regional hospital and had between 2 and 36 years of professional experience. They were all mobile phone users but only one had experience with the use of hand-held computers. All test subjects were familiar with stationary electronic patient record systems and described themselves as experienced or semi-experienced IT users.

Tasks. All test subjects were given a series of tasks to solve while using the system. The tasks were derived from a field study at a hospital ward and covered the duties involved in conducting standard morning work routines. This primarily involved 1) checking up on a number of assigned patients based on information in the system from the previous watch, 2) collecting and reporting scheduled measurements such as temperature, blood pressure, and pulse, and 3) reporting anything important for the ongoing treatment of the patients should be taken into consideration on the next shift.

Procedure. Before the evaluation sessions, the test subjects were given a brief instruction to the system. This included the room-sensing functionality and the procedure for scanning patients' bar-code tags. The test subjects were also instructed on how to operate the available instruments for measuring temperature, blood pressure and pulse. The evaluation sessions were structured by the task assignments. The tasks required the test subjects to interact with all three patients in the two hospital wards, and move between the two rooms through the connecting hallway a number of times. The nurses were encouraged to think aloud throughout the evaluation explaining their comprehension of and interaction with the system. The evaluations lasted between 20 and 40 minutes and were followed by the test subjects filling out a questionnaire.

Roles. Each evaluation session involved six people. One nurse used the system for carrying out the assigned tasks. Three students acted as hospitalized patients. One researcher acted as test monitor and asked questions for clarification. A second researcher operated the context-control centre and the video equipment.

2.3 Field Evaluation

The second evaluation took place at the Hospital of Frederikshavn. The aim of this evaluation was to study the usability of MOBILEWARD for supporting real work activities at a hospital involving real nurses and real hospitalized patients. In order to achieve this, we adopted an observational approach combined with questions for clarification while the nurses were not directly engaged in conducting their work. The field evaluation is described in detail below.

Setting. The field evaluation was carried out at the Medical Department at the Hospital of Frederikshavn (figure 5 and 6). This included the physical area of seven hospital wards, an office with reception, a rinse room and a break-out area connected by a central hallway and involved nurses at work and patients committed to the hospital.



Fig. 5. Field evaluation at the hospital

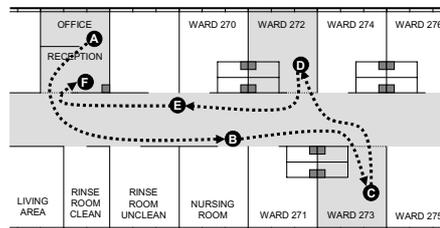


Fig. 6. Physical layout of the hospital wards

Data collection. Motivated by the challenges of capturing high-quality video data during usability evaluations in the field, we designed a portable configuration of audio and video equipment to be carried by the test subject and an observer, allowing a physical distance of up to 10 meters between the two. The configuration consists of a tiny wireless camera (also used in the laboratory evaluation described above) clipped-on to the mobile device (figure 2) and a clip-on microphone worn by the test subject. Audio and video is transmitted wireless to recording equipment carried by the observer (figure 7). In the test monitor's bag, the video signal from the clip-on camera can be merged with the video signal from a handheld camcorder (Picture-in-Picture) and recorded digitally. This allows us to record a high-quality close-up view of the screen and user-interaction as well as an overall view of user and context. During the evaluation, the observer can view the user's interaction with the mobile device on an small LCD screen and monitor the sound through earphones.

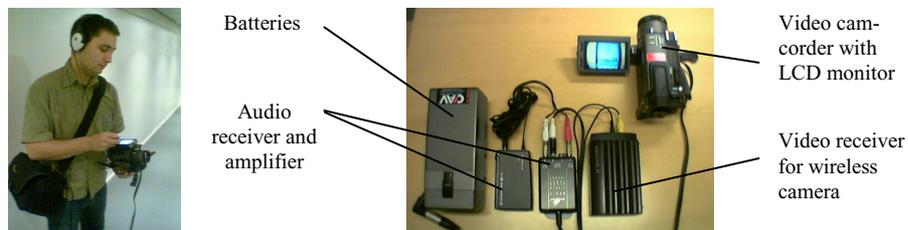


Fig. 7. Observer (left) carrying and operating portable audio/video equipment (right) for capturing high-quality data in the field.

For ethical reasons, we were not permitted to film the hospitalized patients allowing only the video signal from the clip-on camera to be recorded.

Test subjects. Six test subjects (all females) aged between 25 and 55 years participated in the field evaluation. All test subjects were trained nurses employed at the Hospital of Frederikshavn and had between 1 and 9 years of professional experience. They were all mobile phone users but novices with the use of handheld computers. All test subjects were frequent users of a stationary electronic patient record system and described themselves as experienced or semi-experienced users of IT.

Tasks. The field evaluation did not involve any researcher control in form of task assignments but was structured by the work activities of the nurses in relation to conducting standard morning work routines. As in the task assignments of the laboratory evaluation, the work activities of the nurses involved 1) checking up on a number of assigned patients, 2) collecting and reporting scheduled measurements, and 3) reporting anything important for the ongoing treatment of the patients.

Procedure. As in the laboratory evaluation, the test subjects were given a brief instruction to the MOBILEWARD system, including the room-sensing functionality and the procedure for scanning a patient's bar-code tag. The evaluation sessions were structured by the work activities of the nurses which involved interaction with three patients in different wards and moving between different rooms through the connecting hallway a number of times. The nurses were encouraged to think aloud when possible. The evaluations lasted 15 minutes on average and were followed by the test

subjects filling out a brief questionnaire. In order to be able to include a suitable number of nurses, the field evaluation took place over two days.

Roles. Each evaluation session involved six people. One nurse used the system for carrying out her work activities. One researcher acted as test monitor and asked questions for clarification while in the hallway. A second researcher operated the context-control centre application and the portable audio/video equipment. In addition, each evaluation session involved three hospitalized patients in their beds. Due to the real-life nature of the study, each evaluation session involved different patients.

2.4 Analysis

The data analysis aimed at creating two lists of usability problems identified on the background of the two experimental settings. The usability problems were classified as cosmetic, serious or critical based on the guidelines provided by Molich [13]. The two usability evaluations amounted to approximately 6 hours of video recordings depicting the 12 test subject's use of the system. All sessions were analyzed in random order by two teams of two trained usability researchers holding Ph.D. or Master Degrees in Human-Computer Interaction. Each team analyzed the videos in a collaborative effort allowing immediate discussions of identified problems and their severity (as adapted in [11]). As a guideline for the collaborative analysis, each identified usability problem would be discussed until consensus had been reached. The two teams produced two lists of usability problems. Subsequently, these two lists were merged into one complete list. Again, this was done in a collaborative effort, discussing each problem and its severity until consensus had been reached.

Resources spent on planning and conducting the laboratory and field evaluation respectively was calculated on the basis of a time log kept by the involved researchers.

3 Results

We identified a total of 37 different usability problems from the 12 laboratory and field sessions where eight problems were assessed to be critical, 19 problems were assessed to be serious, and ten problems were assessed to be cosmetic (see table 1).

Table 1 Distribution of total numbers of identified usability problems

	Laboratory (N=6)	Field (N=6)
Critical (N=8)	8	7
Serious (N=19)	18	10
Cosmetic (N=10)	10	6
Total (N=37)	36	23

Our study showed that the laboratory setting revealed more usability problems than the field setting. The six test subjects in the lab experienced 36 of the 37 usability problems whereas the six test subjects in the field setting experienced 23 of the 37 usability problems; this difference is extreme significant according to a Fischer's exact test ($p < .001$). 14 usability problems (1 critical, 9 serious, 4 cosmetic) were unique to the lab setting, whereas one serious usability problem was unique to the field.

Regarding the critical problems, the lab setting identified all eight critical problems and the field setting identified seven critical problems; this difference is not significant. Considering the serious problems, we find that the lab identified eight additional problems compared the field and this difference is strong significant ($p < .01$) whereas the difference in cosmetic problems is significant ($p < .05$).

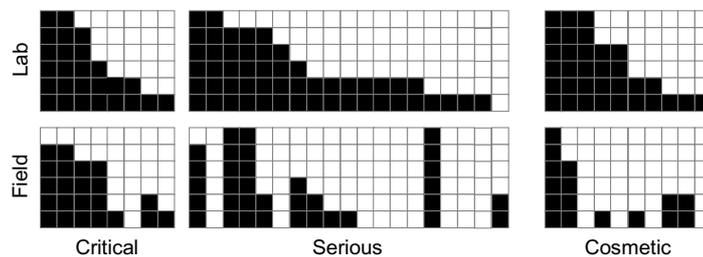


Fig. 8. The distribution of identified usability problems in the laboratory and in the field. Each column represents one usability problem associated the number of test subjects experiencing the problem (indicated by black boxes) for both settings.

Figure 8 summarizes the distribution of the identified 37 usability problems where each column represents one usability problem associated the number of test subjects experiencing the problem (indicated by black boxes) for both settings. Seven usability problems (two critical, two serious, three cosmetic) were experienced by all six subjects in the lab setting whereas three usability problems (two serious, one cosmetic) were experienced by all six subjects in the field setting; one usability problem (cosmetic) was experienced by all 12 subjects.

Looking across the distribution of the usability problems (in figure 8), we find that while the critical problems have a roughly similar distribution, the serious and cosmetic problems have rather dissimilar distributions where some problems were identified by all or nearly all subjects in one setting, but only identified by a few or none in the other setting. E.g. all subjects were informed to use either their fingers or the attached pen for device interaction, but only the lab subjects chose to use the pen and most of them experienced difficulties in placing the pen between tasks.

Analyzing the average numbers of usability problems identified per usability session, we find that the lab subjects on average experienced 18.8 usability problems ($SD=2.0$) and the field subjects on average experienced 11.8 usability problems ($SD=3.3$) and this difference is strong significant according to a Mann-Whitney U-test ($t=2.651$, $p < .01$). This is mainly explainable through higher average numbers of identified serious and cosmetic usability problems where the difference of identified serious problems is strong significant ($t=2.79$, $p < .01$) and so is the difference of cosmetic prob-

lems ($p=2.84$, $p<.01$). On the other hand, we found no significant difference between the numbers of identified critical usability problems. This perspective on our data supports the findings illustrated in table 1 on total number of problems identified by six subjects in each configuration.

Table 2 The average number of identified problems per test session (standard deviations in parentheses).

	Laboratory (N=6)	Field (N=6)
Critical	5.3 (1.2)	4.5 (2.2)
Serious	7.5 (1.0)	4.5 (0.8)
Cosmetic	6.0 (0.9)	2.8 (1.0)
Total	18.8 (2.0)	11.8 (3.3)

Summarizing the time logs, a total of 34 man-hours were spent on the laboratory evaluation. Roughly 50% of this time was spent on planning the evaluation and setting up the lab while the remaining 50% was spent on conducting the evaluation sessions. In comparison, the field evaluation required a total of 65 man-hours. While the actual evaluation sessions in the field took less time than in the lab, the difference between the two was mainly accounted for by larger overhead for planning and transport and by more time spent on setting up the portable AV equipment and configuring MOBILEWARD with real data.

4 Discussion and Conclusions

The aim of our study was to identify opportunities and limitations of usability evaluation in laboratory and field conditions. Based on the results above; the numbers of identified problems, the nature of the identified problems and the lessons learned from conducting the two evaluations, we present the following four key findings:

(i) *Little added value of taking the evaluation into a field condition.* Quite surprisingly, our study shows that when compared to setting up a realistic laboratory study evaluators achieve very little added value when taking a usability evaluation of a context-aware mobile device into the field. In fact, in our study the laboratory setting was able to identify the exact same problems as in the field except for only one. This particular problem was related to an uncertainty expressed amongst some of the nurses at the hospital about the validity of data entered into the system, and whether it had been correctly saved in the database. The identification of this issue in the field relates to the evaluation taking place during real work in a safety-critical use-context where errors cannot be tolerated. The fact that it was *only* identified in the field somewhat indicates a lack of realism in the laboratory condition.

The lack of added value of field evaluations contradicts the assumptions of more mobile HCI research studies, cf. [1, 5, 8, 14]. Here the general assumption is that

evaluation of mobile, context-dependent and nomadic software should be conducted in their natural habitat in order to generate appropriate findings. In practice, however, this assumption is not taken into account by most research studies on mobile HCI as these typically apply laboratory-based evaluations [10]. Our results indicate that this may not be such a huge problem after all, and that expensive time in the field should perhaps not be spent on usability evaluation if it is possible to create a realistic laboratory setup including elements of context [11, 12, 20] and requiring mobility [9, 17]. As in the case of the evaluated system, field studies may instead be more suitable for obtaining insight needed to design the system right in the first place. Our results furthermore show that recreating the use context in a usability laboratory, as e.g. outlined by Nielsen [14], can produce successful mobile system usability results.

(ii) *Lack of control undermined the extendibility of the field condition.* Our study showed that the lack of control in field-based evaluations makes it challenging for evaluators to conduct field evaluations in practice and to make sure that every aspects of the system is covered. In our case, none of the field subjects used the note taking facility of the MOBILEWARD system, leaving no chance for identifying usability problems in this particular system component. As we chose to have the actual work activities of the nurses directing the evaluation, we had no opportunities to force the use of the note taking functionality. This partially influenced the significant higher number of identified serious usability problems in the lab condition.

Issues often discussed in the usability literature are usability problem relevance and validity [13, 15, 19]. Artificial based evaluations e.g. think-aloud protocols in laboratory evaluations or heuristic evaluations may generate false positive problems that are not really problems in everyday use [13]. As a consequence, the higher number of identified problems in the lab condition could be a result of irrelevant usability problems; problems, which nurses would never experience when using the system in real life. However, our data does not exhibit whether this was the case or not. Finally, our field study was much more time consuming as it involved more preparation and travel cost; this is in line with findings of other research studies [6, 9, 14, 16, 18].

(iii) *Both the lab and the field revealed context-aware related problems.* For this particular study, we explicitly stressed the importance of context as the evaluated system was context-aware. Consequently, we would expect that in-situ evaluation could provide a different and perhaps more rich outcome. However, this was only vaguely the case. Both conditions identified all seven context-aware related problems, e.g. the problem of automatically updating information and functionality on the screen according to physical location was not always wanted by the subjects. Typically, they would either get confused or annoyed.

Surprisingly, however, all six field test subjects (but only one lab subject) got confused or did not understand why the system would automatically update information and functionality according to the physical location. So even though their use situation was in-situ and closely related to the context, they would still get confused of the system being actively context-aware (as defined by Barkhuus and Dey [3]). Analyzing this result, we find that their reluctance towards the automatic-update element in the mobile device may stem from the consequently decreased lack of control. Operating and working in a safety-critical environment like healthcare, the decreased level of control may not appear to support systematic work practices, but merely to compro-

mise the work activities. The feeling of lack of control is well-known to active context-aware mobile system [3] and should probably be investigated further. Summarized, for professional work activities, our results seem to contradict statements from other literature on where to conduct evaluations of mobile systems, e.g. [1].

(iv) *The clip-on camera facilitated high-quality data collection of mobile use.* Aspects of mobility and of use in field settings typically challenge evaluators' opportunities for capturing the interaction between the user and the system. However, our configuration with a wireless device clip-on camera allowed the capturing of high resolution images of the interaction, which was invaluable during the later data analysis. The mobile configuration allowed the subjects to move freely in the environment, i.e. the lab and the field, while at the same time still providing us with the opportunity to record a close-up view of the interaction. The portable configuration of audio/video equipment made it possible to capture this data in the field.

Other studies have also stressed the importance of capturing the user-interaction and screen images of the system being evaluated [19]. Generally, this have been found to be very difficult during mobile use [6, 8, 9]. Another way of dealing with this problem is to replicate screen images from the mobile devices on a laptop or stationary computer via a network connection and grab the images from here. However, this does not allow the capturing of situations where e.g. input is not registered by the system and does allow observation of user-interaction with the physical device. In laboratory settings, stationary cameras can be used to capture the screen of mobile devices too, but this approach is very sensitive to physical movements and typically requires the device to be held within a delimited area. In the field, video data is typically recorded by an observer with a handheld camera, continuously shifting focus between the mobile device, the user and the surrounding environment. However, this approach does normally not provide a very good view of the mobile device screen and user-interaction. Also, it requires the camera-operator to be in close proximity of the user and is highly sensitive to physical movement (which is, of course, prevalent during mobile use in the field).

Our study suffers from a number of limitations. First, the evaluated EPR system and the associated healthcare context probably influence the results of the study. Other domains may exhibit different characteristics where the link between the use of the system and the context may be weaker or stronger. Secondly, usability evaluations as applied in this paper provide only snapshots of intended future use. Other methods for understanding use and interaction like ethnographic studies can most likely provide different perspectives on context-aware mobile systems use. This could then in hand supplement or contradict the findings of our study.

References

1. Abowd, G. and Mynatt, E. (2000) Charting past, present and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction*, Vol. 7(1), pp. 29-58
2. Bardram, J., Kjær, T. A. K., and Nielsen, C. (2003) Supporting Local Mobility in Healthcare by Application Roaming among Heterogeneous Devices. In *Proceedings of Mobile HCI 2003*, Udine, Italy, Springer-Verlag, LNCS

3. Barkhuus, L. and Dey, A. (2003) Is Context-Aware Computing Taking Control away from the User? Three Levels of Interactivity Examined. In Proceedings of the UbiComp2003 conference, LNCS 2864, pp. 149 - 156
4. Bohnenberger, T., Jameson, A., Krüger, A., and Butz, A. (2002) Location-Aware Shopping Assistance: Evaluation of a Decision-Theoretic Approach. In Proceedings of Mobile HCI 2002, Pisa, Italy, Springer-Verlag, LNCS
5. Brewster S. (2002) Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing*, 6, 188-205
6. Esbjörnsson M., Juhlin O. and Östergren M. (2003) Motorcyclists Using Hocman – Field Trials on Mobile Interaction. In Proceedings of the 5th International Mobile HCI 2003 conference, Udine, Italy, Springer-Verlag, LNCS
7. Graham R., Carter C. (1999) Comparison of Speech Input and Manual Control of In-Car Devices while on-the-move. In Proceedings of the Second Workshop on Human Computer Interaction with Mobile Devices, Mobile HCI 1999, Edinburgh, Scotland.
8. Johnson P. (1998) Usability and Mobility; Interactions on the move. In Proceedings of the First Workshop on Human-Computer Interaction with Mobile Devices, Glasgow, Scotland, GIST Technical Report G98-1
9. Kjeldskov J. and Stage J. (2004) New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies (IJHCS)*, 60, 599-620, Elsevier
10. Kjeldskov J. and Graham C. (2003) A Review of Mobile HCI Research Methods. In Proceedings of the 5th International Mobile HCI 2003 conference, Udine, Italy, Springer-Verlag, LNCS
11. Kjeldskov J. and Skov M. B. (2003) Creating a Realistic Laboratory Setting: A Comparative Study of Three Think-Aloud Usability Evaluations of a Mobile System. In Proceedings of the 9th IFIP TC13 International Conference on Human Computer Interaction, Interact 2003. Zürich, Switzerland. IOS Press
12. Lai J., Cheng K., Green P., Tsimhoni O. (2001) On the Road and on the Web? Comprehension of synthetic speech while driving. In Proceedings of CHI'2001, ACM, New York, pp. 206-212
13. Molich, R. (2000) *Usable Web Design* (In Danish). Ingeniøren | bøger
14. Nielsen C. (1998) Testing in the Field. In Proceedings of the third Asia Pacific Computer Human Interaction Conference, IEEE Computer Society
15. Nielsen J. (1993) *Usability Engineering*. Morgan Kaufmann
16. Pascoe J., Ryan N., Morse D. (2000) Using While Moving: HCI Issues in Fieldwork Environments. *Transactions on Computer-Human Interaction*, 7(3), 417-437.
17. Pirhonen A., Brewster S.A., Holguin C. (2002) Gestural and audio metaphors as a means of control for mobile devices. In Proceedings of CHI'2002, ACM, New York.
18. Rantanen J., Impio J., Karinsalo T., Reho A., Tasanen M., Vanhala J. (2002) Smart Clothing Prototype for the Arctic Environment. *Personal and Ubiquitous Computing*, 6, 3-16.
19. Rubin, J. (1994) *Handbook of Usability Testing*. Wiley
20. Salvucci D. D. (2001) Predicting the Effects of In-Car Interfaces on Driver Behaviour using a Cognitive Architecture. In Proceedings of CHI'2001, New York, pp 120-127
21. Schilit, B. N. and Theimer, M. M. (1994) Disseminating Active Map information to Mobile Hosts. In *IEEE Network*, Vol. 8(5), pp. 22-32
22. Skov, M. B. and Høegh, R. T. (2004) Supporting Cooperation and Coordination in a Hospital Ward by a Context-Aware Mobile Electronic Patient Record. Submitted to the *Journal of Computer-Supported Cooperative Work*, Elsevier.
23. Urban, M. and Kunath, M (2002) Design, Deployment and Evaluation of a Clinical Information System which uses Mobile Computers and Workstations. In Proceedings of the 2nd conference on Mobile Computing in Medicine, Heidelberg