A Field Laboratory for Evaluating In Situ

Høegh, Rune Thaarup; Kjeldskov, Jesper; Skov, Mikael; Stage, Jan

Published in:
Handbook of Research on User Interface Design and Evaluation for Mobile Technology

Publication date:
2008

Document Version
Early version, also known as pre-print

Please refer to the published version of this chapter:
Høegh R.T. Kjeldskov J., Skov M.B. and Stage J. (2008) A Field Laboratory for Evaluating In Situ. In Lumsden J (Ed.) Handbook of Research on User Interface Design and Evaluation for Mobile Technology. Pp. 982-996. IGI Global.

A FIELD LABORATORY FOR EVALUATING IN SITU

Rune T. Høegh, Jesper Kjeldskov [(*)], Mikael B. Skov and Jan Stage

Department of Computer Science
Aalborg University
Fredrik Bajers Vej 7E
DK-9220 Aalborg East
Denmark

Tel: +45 96358080
Fax: +45 98159889

Email: runethh@cs.aau.dk, jesper@cs.aau.dk, dubois@cs.aau.dk, jans@cs.aau.dk

[(*)] Corresponding author

A FIELD LABORATORY FOR EVALUATING IN SITU

ABSTRACT

Evaluating mobile technologies "in the real world" is hard. It is challenging to capture key situations of use, hard to apply established techniques such as observation and "thinking aloud", and it is complicated to collect data of an acceptable quality. In response to these challenges, we have developed a "field laboratory" for evaluating mobile technologies in situ. Facilitating high-quality data collection as well as unobstructed user interaction, the field laboratory allows a small wireless camera to be attached to a mobile device, capturing a close-up image of the screen and buttons. This chapter describes the iterative development of our field laboratory over 4 years of evaluating several mobile systems in field settings. It leads to a description of the current setup and how it is used, and explains the rationales for key decisions on technology and form factors made throughout its development.

INTRODUCTION

Studying peoples' use of technology is a key activity within the research field of Human-Computer Interaction (HCI) providing software developers with invaluable information about the usability and usefulness of their systems at different stages of the process from conceptual design to a final implementation. Traditionally, such studies have taken place in dedicated "usability laboratories" where users' interaction with computer systems can be observed in a controlled experimental setting providing video and audio data of very high quality. Studying the usability of mobile technologies, however, raises new questions and concerns. Mobile systems are typically used in highly dynamic contexts involving a close interaction between people, systems and their surroundings. Therefore, studying mobile technology use in situ seems like an appealing or even indispensable approach – rather than trying to recreate the use situation realistically in a laboratory. However, studying mobile technology usability "in the real world" is difficult. It is difficult to capture key situations of use, apply established usability techniques such as observation and "thinking aloud" without interfering with the situation, and it is complicated to collect data of an acceptable quality.

In response to some of these challenges, we have extended our stationary usability laboratory at Aalborg University's Department of Computer Science with a mobile counterpart, the field laboratory, which can be taken into the field when studying mobile system use and usability. Facilitating high-quality data collection as well as unobstructed user interaction, the field laboratory allows a small wireless camera to be attached to the mobile device, capturing a close-up image of the screen and buttons while a third-person view is captured by a handheld camcorder.

The purpose of this chapter is to communicate our experiences with developing and using the field laboratory for evaluating mobile technology use and usability in situ by taking the readers through four years of major iterations leading to its current configuration. By doing this, it is our aim to make practitioners, researchers and designers of mobile technologies able to set up and use their own field laboratories for evaluating mobile systems in situ. It is also our aim to inspire further development of even better field laboratory setups facilitating better, easier, faster, and cheaper use and usability data collection in the field. It is not the purpose of this chapter to discuss the relation between evaluating in the field or in the lab. We take our point of departure in the assumption that you have decided to evaluate in the field and focus on how you can collect high quality data while out there. It is also not our aim to present or discuss findings about the usability of the specific systems we have evaluated with our field laboratory (these can be found elsewhere). Instead, the purpose of mentioning these studies here is to illustrate how they functioned as vehicles for iterating on the field laboratory's configuration.

The chapter begins with a short summary of related work motivating the development of techniques for improving evaluation data collection in the field. We then describe three iterations of developing our own field laboratory. For each of these iterations, we describe our initial motivations and aims, the corresponding configuration of equipment, an example evaluation where it was used, and the pros and cons identified. The next iteration then describes how we modified the field laboratory configuration accordingly, and what we learned from using it in practice. Finally, we describe the current setup, outline some future trends within this area of research, and conclude on the work presented in the chapter.

BACKGROUND

In the proceedings of the first workshop on Human-Computer Interaction (HCI) 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 (Johnson, 1998). Of specific concerns 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 such as (Brewster, 2002; Esbjörnsson et al., 2003; Pascoe et al., 2000).

In 2003, a literature study revealed that 41% of mobile HCI research involved evaluation (Kjeldskov & Graham, 2003). However, even though evaluations of mobile systems were clearly prevalent, only 19% of these evaluations were carried out in the field while 71% were carried out in laboratory settings. Although the issue of how to study and evaluate mobile technology use and usability in the field has since received increased attention, no established set of usability evaluation methods and data collection techniques yet exists for field evaluations.

The research into field-based evaluations of user interfaces for mobile technologies can be divided into two overall categories of equal importance. The first category focuses on the methodological challenges of adapting traditional usability evaluation methods such as the use of the think-aloud protocol, as well as developing new ones, to suit the challenges and prospects of evaluating mobile user interfaces in the field. The second category focuses on the practical challenges of improving existing techniques for data collection in field settings and developing new ones. In this chapter, we focus on the latter: how to facilitate data collection better when evaluating user interface design for mobile technologies in the field.

One of the primary sources of data when evaluating the usability of an IT system is video and audio recordings of use depicting the system, the users' interaction with it, and the context in which this takes place. When evaluating in the field, the primary challenge of data collection is that these recordings can be very hard to make at a sufficient level of quality. Video filming evaluation sessions in the field with a handheld camcorder is seemingly an attractive approach because it is cheap and easy (figure 1 left). However, while suitable for capturing the overall use context of a field evaluation, capturing good close-up views of mobile device screen, buttons and user interaction can be quite difficult while moving (Kjeldskov et al., 2005). Furthermore, filming a good overview of a use situation with a handheld camcorder require a bit of distance while obtaining good close-ups and good sound requires that the cameraman stay relatively close to the test subject and interviewer. The latter often results in the so-called "bodyguard effect" (figure 1 right) where the test subject is practically isolated from other people in their surroundings, hence questioning the value of going into the field in the first place (Kjeldskov & Stage, 2004).


Figure 1. Usability evaluations of mobile technologies in the field using a handheld camcorder and note taking for data collection.

Within the "practical" category of improving data collection techniques for evaluations in the field, three specific approaches are particularly worth mentioning.

One approach has aimed at obtaining field data in a non-intrusive way through automatic logging of user interaction for later analysis. Through logging, researchers can accurately record a user's interaction with a system, such as clicks or keyboard entries, or even record the entire graphical user interface of the software being evaluated. One of the advantages of logging is that it does not necessarily require the presence of a test monitor, and involves a minimum of interference with the user's context. This makes logging particular useful for longitudinal studies of mobile technology usability. Logging is also an efficient method to obtain data in a cost effective way from a large population of users. One of the drawbacks of logging is that it does not usually record any information outside the mobile device. It provides no record of, for example, the physical surroundings of the user, and it does not record so-called "near-interactions" where, for example, the user fails to interact with the system (Waterson et al., 2002). Another quite significant limitation is that logging usually requires installation of dedicated software on the device being evaluated. This is not only cumbersome but also sometimes simply not possible. While highly suitable for generating large amounts of data for quantitative studies, logging does usually not provide good data for qualitative studies. A way of overcoming this limitation could be to combine automatic logging with, for example, video and audio recordings, interviews etc.

Another approach has aimed at bringing traditional laboratory setups into the field by means of a "portable usability laboratory" or "lab-in-a-box" (Kimber et al., 2005; Winters et al., 2001). The advantage of a portable laboratory is that it allows rich data to be collected using high quality equipment. Not being truly mobile, portable usability laboratories are, however, best used in field settings where the user remains semi-mobile within a delimited spatial area for a period of time – for example in a restaurant or on the bridge of a ship. Other drawbacks for this approach are that the equipment is often cumbersome to transport and setup and may be intrusive in the context (Rowley, 1994). Setting up large amounts of video and audio recording equipment in the field may also cause users and surrounding people to act differently, which, in essence, stand diametrically opposed to the purpose of evaluating usability in the field. As a final downside, it may be difficult to record video of users' interaction with a mobile device with standard camera equipment.

Taking its offset in the challenges of using portable laboratories in the field, a third approach has been aimed at developing more compact and mobile usability laboratory facilities that are able to record high quality video data from various sources in an un-intrusive way. Different configurations of such mobile usability laboratories have been described in recent literature (e.g. Betiol & Cybis, 2005; Kaikkonen et al., 2005; Roto et al.,

2004) and demonstrated at leading conferences within the field (e.g. Nyyssönen et al., 2002). The typical setup of a mobile laboratory makes use of a mini camera that can be attached to a mobile device for a good close up of the screen and user interaction. In some setups, such as the one proposed by Roto et al. (2004), additional cameras are used to capture views of the evaluation context. Images from these cameras are then mixed and recorded for later playback during analysis. While mini-camera approaches like this are highly promising – not only in field evaluations but also in laboratory settings – experiences from the deployment of mobile usability laboratories in the field also point out a series of challenges. Some of the issues relates to the quality of video and audio recordings when using wireless equipment, and how to best record multiple video sources and audio in sync. Other issues relate to battery lifetime and the weight of the equipment having to be carried around during the evaluation sessions.

In the following sections, we outline how we have dealt with these and other challenges through three iterations of setting up and using a field laboratory for in situ evaluations of mobile technologies.

CLOSE-UP VIDEO AND IMPROVED SOUND

Motivated by the challenges of capturing high-quality video data during usability evaluations in the field described in the literature and experienced in a series of evaluations carried out between 2002 and 2003, we decided to develop a portable configuration of audio and video equipment that could be carried by the test subject and an observer during a field evaluation. Our primary focus for the first version of our "field laboratory" was to enable close-up recording of the mobile device screen and user interaction. Inspired by commercially available products, such as the "mobile device camera" from Noldus (Noldus, 2005), we constructed a small camera-mount on which a mobile phone or PDA could be mounted with Velcro (figure 2 left). The camera-mount contained a wireless camera mounted on a flexible "gooseneck" as well as a 9v battery-supply. This allowed us to capture a detailed close-up view of the mobile device in colour (figure 2 right) and record this throughout the whole evaluation. Apart from recording close-up video of the mobile device, we also wanted to improve the sound quality of our data recordings to minimize ambient noise and ensure capturing all utterances made by the test subject and the interviewer. For this, we combined the camera on the mobile device with an off-the-shelf professional wireless microphone from Sennheiser; a lapel microphone with a belt-pack transmitter worn by the test subject and a belt-pack receiver carried by the observer.

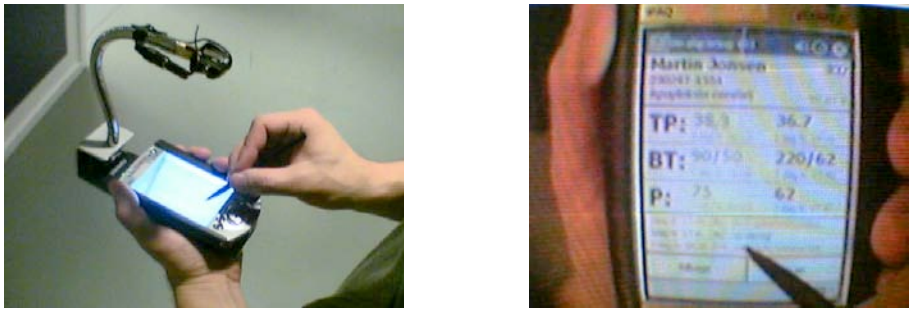

Figure 2. PDA on camera-mount allowing for close-up view of screen and user interaction.

Video from the camera on the mobile device and audio from the lapel microphone is transmitted wirelessly to receivers and recording equipment carried by an observer (figure 3). In the observers bag, the video and audio signals are recorded on a portable DV recorder, for

example a camcorder, set up to record from an external source. During the evaluation, the observer can monitor the user‘s interaction with the mobile device on a small LCD screen and monitor the sound through earphones.

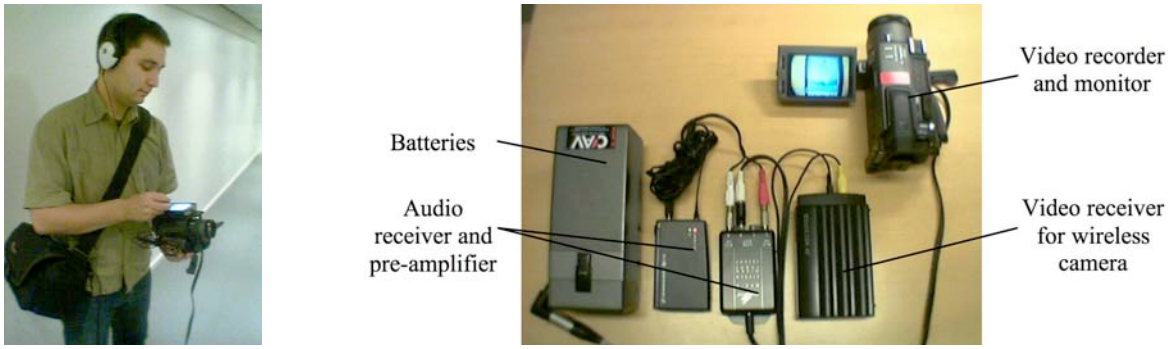

Figure 3. Observer (left) carrying and operating portable audio/video equipment (right) for capturing close-up view of screen and user interaction.

Using The First Field Laboratory in Practice

We used the first version of our field laboratory described above for an evaluation of a mobile information system in situ in 2003/04 (Kjeldskov et al., 2004). The evaluation focused on the use of a mobile, context-aware, electronic patient record system by nurses and doctors at a large regional hospital in Denmark. Six test subjects (all females) aged between 25 and 55 years participated in the field evaluation. They were all trained nurses with 1-9 years of professional experience.

Due to the real-life nature of the study, the field evaluation did not involve any researcher control in form of task assignments but was structured exclusively by the work activities of the nurses. The studied work activities were highly mobile, and involved interaction with assigned patients in different wards (i.e. collecting and reporting scheduled measurements), and moving back and forth between different rooms and hallways. As in a standard usability evaluation, the test subjects were given a brief instruction to the mobile system being evaluated and were encouraged to think aloud when possible. Each evaluation session lasted 15 minutes on average and involved three people. One nurse used the system for carrying out her work activities. One researcher acted as interviewer and asked questions for clarification while in the hallway. A second researcher operated the field laboratory. In addition, each session involved a number of hospitalized patients in their beds. For ethical reasons, we did not film the hospitalized patients. In order to be able to include a suitable number of different nurses as test subjects, the field evaluation took place over two days.

Lessons Learned From Using Field Lab #1

The field evaluation at the hospital highlighted a series of the challenges related to evaluating mobile technologies in situ. It was highly time consuming and complex to plan and execute the study, and it was difficult to capture key situations of use. However, in relation to data collection, the camera on the mobile device provided us with high-quality close-up views of the nurses’ interaction with the system being evaluated, while at the same time allowing them to move around freely in the environment and focus on their work. The use of a professional wireless microphone supplemented the video close-up recordings with a clear audio track capturing all the nurses’ utterances as well as enough ambient sounds to give a sense of context. During the later analysis phase, these video and audio recordings were invaluable

sources of data for identifying usability problems and suggesting opportunities for redesign. The video track allowed us to see exactly which parts of the system were perceived as problematic and where the nurses had problems with operating the interface. The audio track allowed us to hear the nurses' comments about their interaction with the system and provided us with context of use. When evaluating mobile technologies in a laboratory, this kind of data is very much standard. The first version of our field laboratory made it possible to capture the same kind of data in situ as well. It was lightweight, and it was relatively easy to operate.

On the downside, the first version of our field laboratory also had a number of limitations. First of all, the video recording only contained the close-up view of the mobile device and the user interactions taking place within 5-6 centimetres of the screen. It did not capture the users or their surroundings. During the data analysis phase, this proved to be very problematic at times where the use context was significant for understanding what the user was trying to do with the system. It was also hard to tell from the video track when the users were looking at the screen of the mobile device and when they focused elsewhere during the evaluation. Although the audio track did provide some information about context and the focus of the users, this information was often partial, ambiguous, and not conclusive. Secondly, the audio track only captured the voice of the interviewer if he or she was standing close to the test subject (who was wearing the microphone). In a stationary evaluation setup, this would usually not be a problem because the interviewer and test subject will be seated close to each other. However, when evaluating in the field it is most likely that interviewer and test subject will sometimes be physically separated by enough distance for directional microphones not to be able to pick up the voice of them both. In the field evaluation at the hospital this was often the case simply because the nurses were sometimes hard to keep up with by the interviewer and because the interviewer would sometimes have to stand back a bit in order not to interfere with the nurses' work tasks (i.e. attending to patients in bed). Thirdly, the mini camera was far from perfect. Although it was considerably smaller than commercially available alternatives, the gooseneck camera-mount clearly influenced the form factor of the mobile device being evaluated. It was too heavy, and made it impossible for users to hold the device the way they would usually do.

SMALL CAMERAS AND MULTIPLE VIDEO SOURCES

On basis of the lessons learned from the field evaluation at the hospital, we set out to improve our field laboratory in three ways. Firstly, we wanted to reduce the influence of the wireless camera attached to the mobile device being evaluated. We wanted to minimize the size and weight of the camera, and make it more flexible for use with different types and sizes of mobile devices. Secondly, we wanted to facilitate data recording from multiple sources of video allowing us to capture close-up views of the mobile device, close-up views of the user, 3rd person views of the user in context, and 1st person views of the surroundings as seen in, for example, Roto et al. (2004). Thirdly, we wanted to be able to capture audio from multiple sources independently (e.g. the test subject and the interviewer).

Minimizing the size and weight of the camera on the mobile device turned out to be surprisingly simple while at the same time also increasing its flexibility. Our solution was to simply strap the camera house on to a small plastic clamp with a flexible piece of plastic and a few cable strips. All items necessary to produce the wireless "camera-clamp" were purchased from a local hardware store for less than 20 USD. The clamp made it possible to mount the camera on almost any mobile device without interfering with its form factor (figure 4). The 9v battery powering the camera was simply attached to the mobile device with double-sided tape, wherever it would interfere the least with the user's grip of it. Using the same approach, we created other variations of the camera-clamp. One was also clipped-on to the mobile device but faced the camera towards the user (figure 4 right). Another one was

designed to sit on the user's ear (like a Bluetooth headset) capturing a first-person view of the surroundings. These additional wireless cameras allowed us to capture video data from multiple sources in parallel.

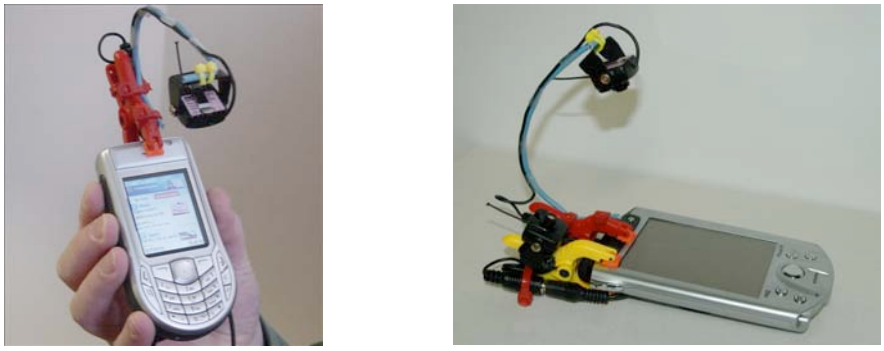

Figure 4. Lightweight camera-clamps attached to mobile devices

In order to capture a third-person view of the evaluation session, we decided to equip the observer with a handheld video camcorder. For better audio capture, we added a second wireless lapel microphone for the interviewer.

While reducing the size and weight of equipment carried by the test subjects (even though we added more cameras), the addition of more cameras and microphones significantly increased the equipment necessary to be carried and operated by the observer (figure 5). The additional lapel microphone required an additional belt-pack receiver. For each additional wireless camera we had to add another video receiver and 12v battery. In order to include the video signal from 3-4 different sources in one composite video recording, we had to include some sort of battery driven video mixing as well. For this purpose, we modified a stationary Panasonic WJ-MS 424 Quad display unit to run on batteries. In order to minimize the number of different batteries in use and avoid batteries running flat at different times, we custom-built a power supply, which could power all equipment from the same 12v battery source (apart from the camcorders which ran on their own batteries).

Figure 5. Equipment used for the second field laboratory (configured for two wireless cameras). Batteries and power regulators are not shown

Using The Second Field Laboratory in Practice

The second field laboratory setup described above was used in pilot studies preparing for a large-scale evaluation of a mobile information system in situ in 2005. The aim was to facilitate data collection about the use and usability of a context-aware mobile web site used by pairs of friends while socializing "out on the town". Hence, it was important to document both peoples' interactions with the device, with each other, and with their physical surroundings.

Figure 6. Example recording with multiple video sources

Lessons Learned From Using Field Lab #2

The pilot field evaluations in the city centre of Melbourne once again highlighted the complexity of evaluating mobile technologies in situ. However, this time we clearly got more out of our efforts to move from the laboratory into the field. The second version of our field laboratory made it possible to capture multiple video and audio sources in situ. As in the evaluations at the hospital, test subjects could move around relatively freely, and were undisturbed by the cameraman who could easily keep a distance of 5-8 meters while still capturing good images and sound. As we had aimed for, the second field laboratory provided rich data of high quality capturing both detailed views of the users, their interaction with the device, and their surroundings from several perspectives (figure 6). During the later analysis phase, especially the third-person view of the users in context provided an invaluable resource for contextualizing peoples' verbal utterances and their interaction with the system. Unlike our early evaluations in the field, where we were only using a handheld camcorder, the field laboratory allowed the cameraman to remain focused on the surroundings rather than having to zoom back and forwards between a third-person view and a close-up view of the mobile device. The use of two microphones resulted in a stereo audio track, which very clearly captured all utterances by test subjects and interviewer. Recording two separate audio tracks made it easy to separate between utterances made by different people during playback for analysis. It also made it possible to make post-evaluation adjustments of the relative levels of peoples' voices.

On the downside, however, it only took us two pilot sessions in the field before we realized that the current setup of the field laboratory had a series of fundamental problems and needed to be modified. While we were able to capture great data like never before, the cost of this was very high in terms of battery life, weight, and complexity of operating the equipment needed. We had been able to fit all the field laboratory equipment depicted on figure 5, as well as the necessary batteries and power supply regulators, into a large laptop bag with internal cabling. However, the total weight of the bag exceeded 10 kg, which turned out to be physically challenging for the cameraman to carry for more than a few hours. At the same time, the modified Quad display splitter and the video receivers ran the battery-pack of

four 12v motorcycle batteries flat in less than 1.5 hours. In effect this made back-to-back evaluation sessions impossible without recharging or carrying extra batteries with us into the field as well! While running all equipment on the same 12v power supply reduced the task of monitoring and replacing a lot of individual batteries for, for example, the audio receivers, we also found that the power regulators needed for doing this introduced noticeable noise to the audio recordings. Finally, the amount of equipment and the number of different video and audio sources made it highly complex for one person to operate the field laboratory in the (already) stressful conditions of an evaluation in situ.

On top of these problems, the number of different wireless technologies involved at this stage also resulted in problems with radio interference between equipment operating on the same or close frequencies. While we had no problems whatsoever with the professional wireless microphones, wireless video from multiple cameras turned out to be problematic. Camera signals sometimes interfered with each other, as well as with the wireless capabilities of the mobile device being evaluated (WLAN and Bluetooth). In fact using more than one wireless camera at a time sometimes completely disrupted the PDAs WLAN connection making parts of the evaluation impossible to carry out. At other times, the use of Bluetooth significantly distorted the images from some wireless cameras. Dealing with the problem of radio interference was quite a challenge. While we were to some extend able to modify our own use of wireless technologies during the evaluations to avoid problematic combinations of Bluetooth, WLAN and the wireless cameras, evaluating in the field of course made it impossible for us to control *other peoples'* nearby use of wireless technologies, which sometimes interfered with our equipment.

On the bright side, however, revisiting the field recordings quickly made it evident that collecting data from four independent video sources was not necessary in order to get a sufficient view of users, use, and context. The only sources we made any significant use of during the analysis of the evaluation sessions were the close-up view of the device and the third-person view of users and context. Hence, we could reduce our equipment.

MINIMIZING EQUIPMENT AND INCREASING BATTERY LIFETIME

Informed by the lessons learned from the pilot field studies described above, we made some significant changes to the field laboratory with the aim of minimizing equipment, reducing weight and complexity, and increasing battery life.

Our first major decision was to reduce the number of video sources to two: a wireless camera attached to the mobile device and a handheld camcorder operated by an observer. Reducing the number of wireless cameras limited the issue of radio interference and allowed us to make some significant reductions in the equipment to be carried by the observer. Firstly, the number of video receivers could be reduced correspondingly. Secondly, we were able to replace the battery-hungry Quad display unit with a much smaller Picture-in-Picture unit running on 12v (drivedata DPIP1). In return, these reductions made it possible to phase out a few heavy power regulators and run the field laboratory for almost 4 times longer on half the batteries. Replacing the wireless audio receivers with newer and more lightweight models (Sennheiser ew100 G2), we were also able to phase out an audio preamplifier and noise generating power regulators while at the same time improving the sound quality. We also replaced our portable tape-based DV recorder with a smaller and more lightweight 100GB AV hard disk recorder (Archos AV400). The third generation of our field laboratory is configured as schematically depicted in figure 7.

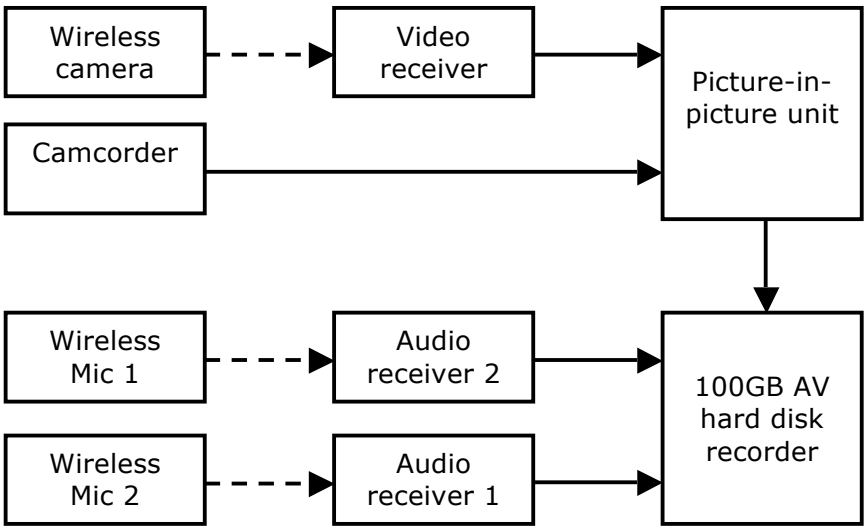

Figure 7. Schematic configuration of the current version of the field laboratory with two video sources and two audio sources recorded in one composite digital file.

Video signals from the wireless camera attached to the mobile device is sent to a receiver in a small bag carried by an observer where they are mixed on the fly with a third-person view of the users captured by the handheld camcorder. Ensuring high-quality sound, users and interviewer are wearing small directional wireless lapel microphones. Mixed video and sound is recorded digitally on a hard disk recorder in the observer's bag. This configuration of the field laboratory weights approximately 4 kg, measures 26x18x30 cm, and has a battery time of approximately 5 hours on two 12v batteries (figure 8).

Using The Third Field Laboratory in Practice

The third field laboratory setup described above was used in a large-scale evaluation focusing on the use and usability of a context-aware mobile web site facilitating sociality in the city centre of Melbourne, Australia (Kjeldskov & Paay, 2005). The field evaluation involved 20 people (grouped in pairs). All pairs of users were familiar with the location at which the evaluation took place and frequently socialized there together (figure 9).

Figure 8. The third field laboratory in a medium-sized light-weight camera bag

Figure 9. The field laboratory in action at Federation Square, Melbourne, Australia

With the purpose of being true to the real-life qualities of studying mobile technology use in situ, the field evaluations were not structured by tasks in a traditional usability evaluation sense of the term. Instead, the evaluations were structured by a set of overall prompts for use

of different parts of the system and a list of corresponding interview questions. The socializing activities studied were highly mobile and involved the users moving between several physical locations in the city; bars, cafés, museums, etc. Prior to the evaluation, the users were given a 10-minute introduction to the system and were allowed to familiarize themselves with it for 5-10 minutes. Inspired by the constructive interaction approach to thinking-aloud studies with more than one user, the groups were asked to talk among themselves about their perception of and interaction with the system interrupted only with questions for clarification. The evaluation sessions each lasted between 45 and 70 minutes and took place over several days.

Lessons Learned From Using Field Lab #3

The field evaluation of the mobile web site was very successful. With the third iteration of our field laboratory, we had reached a very useable and stable solution with a good trade-off between supported data sources, weight and battery lifetime. We were able to capture audio and video sources needed for studying the use and usability of mobile technologies in situ, and were able to do so in a quality that matched (and even sometimes superseded) our stationary usability laboratory. The field laboratory was small, light-weight, relatively simple to operate and had a battery lifetime allowing for 2-3 evaluation sessions in a row without worrying about recharging (at this point the weakest link was in fact the battery lifetime of the PDAs used to run the prototype system). It allowed the observer to effortlessly follow the participants and interviewer from a bit of a distance while filming them and their surroundings with the handheld camcorder. In turn, this allowed the interviewer to focus on the participants' use of the mobile system being evaluated without having to worry about data collection. Figure 10 shows an example of the video data recorded in the field.

While the third version of our field laboratory was already considerably smaller and lighter than any of our earlier ones, we have since been able to reduce the weight and physical size further through a fourth iteration of reducing cabling, battery supply, and optimizing the use of bag-space (figure 11). In our most recent design (version 4), the field laboratory has the same specifications for data capture as described above, but now weights only 2 kg and measures only 18x14x25 cm, making it highly mobile and very easy to bring into the field for longer periods of time. Powered by only one 12v battery, this configuration can operate for approximately 2.5 hours before the battery must be swapped with a spare one.

Figure 10. Video recording with third-person view of participants and close-up view of PDA. Note that the camera focused on the device screen is turned 90 degrees to optimize use of the Picture-in-Picture view.

Figure 11. Our most recent version of the field laboratory weighing only 2 kg and measuring just 18x14x25 cm - containing video and audio receivers, Picture-in-Picture unit, hard disk recorder, and battery.

FUTURE TRENDS

The future trends for developing field laboratories for evaluating mobile technology use and usability in situ focus primarily on improving the quality, reliability, and size of the cameras attached to the mobile device. As wireless video technology matures and becomes more widespread, we are likely to see an emergence of cheap high-end wireless video cameras matching the professional standard of the wireless microphones used in our current version of the field laboratory. Broadcast quality interference-free wireless video technologies exist today, but are still rather expensive and not sufficiently lightweight for our purposes.

Coming from another area of application, new camera technologies are also emerging within the field of video surveillance, which would allow video signals to be transferred digitally via wireless network connections rather than over an analogue radio link. Apart from offering much higher quality and stability, this approach is particularly interesting because it bypasses the use of any analogue video equipment, which is typically quite battery intensive. It also enables the development of field laboratories where all video sources are recorded digitally in separate, time stamped tracks avoiding the down-sampling of Picture-in-Picture and allowing for synchronised playback of multiple camera angles without any loss of quality.

A third emerging way of dealing with the camera problem is to replace it with a software solution that logs screen images from the mobile devices, or replicates them on a laptop or stationary computer via a network connection and then grabs the images from there. However, as discussed earlier in the section about automatic data logging, this approach does not capture the user-interaction with the physical device and situations where, for example, input is not registered by the system. Nevertheless, parallel data logging of the mobile device screen could be a very interesting way of complementing video and audio data captured through wireless cameras and microphones and should be investigated further. In a similar way, capturing video and audio data of user interaction could be an interesting way of enhancing the use of data logging when evaluating mobile technologies in the field.

CONCLUSIONS

In this chapter, we have described the iterative development of a field laboratory facilitating in situ evaluations of mobile technology use and usability. We have described a series of initial motivations, how we responded to these, and the lessons learned from deploying our field laboratory to a series of evaluations.

It is hard to evaluate mobile technologies in situ. It is difficult to capture key situations of use and it is complicated to collect data of an acceptable quality. However, by means of a field laboratory with small wireless cameras and wireless microphones, we have shown that it is possible to capture field data about the use and usability of mobile technologies in a quality that matches that of a stationary usability laboratory. Furthermore, we have shown that field laboratories can be made small, lightweight, and operational for hours before having to recharge batteries. Equipped with a field laboratory as the one described in this chapter, we believe that researchers and designers will be able to make more and better evaluations of user interfaces for mobile technology in the field.

REFERENCES

Betiol, A. H., & Cybis, W. D. A. (2005). Usability Testing of Mobile Devices: A Comparison of Three Approaches. In *Proceedings of Interact 2005* (pp. 470-481). LNCS: Springer-Verlag.

Brewster, S. (2002). Overcoming the Lack of Screen Space on Mobile Computers. *Personal and Ubiquitous Computing, 6*(3), 188-205.

Esbjörnsson, M., Juhlin, O., & Östergren, M. (2003). Motorcyclists Using Hocman – Field Trials on Mobile Interaction. In *Proceedings of Mobile HCI 2003* (32-44). LNCS: Springer-Verlag.

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.

Kaikkonen, A., Kallio, T., Kekäläinen, A., Kankainen, A., & Cankar, M. (2005). Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. *Journal of Usability Studies, 1*(1), 4-16.

Kimber, J., Georgievski, M., & Sharda, N. (2005). Developing Usability Testing Systems and Procedures for Mobile Tourism Services. In *Proceedings of Annual Conference on Information Technology in the Hospital Industry*. Los Angeles: Informational Technology Association.

Kjeldskov, J., & Paay, J. (2005). Augmenting the City: The Design of a Context-Aware Mobile Web Site. In *Proceedings of DUX 2005* (pp. 1-7). San Francisco, CA, USA: ACM.

Kjeldskov, J., Graham, C., Pedell, S., Vetere, F., Howard, S., Balbo, S., & Davies, J. (2005). Evaluating the Usability of a Mobile Guide: The influence of Location, Participants and Resources. *Behaviour and Information Technology, 24*(1), 51-65.

Kjeldskov J., Skov M. B., Als B. S., & 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 *Proceedings of Mobile HCI 2004* (pp. 61-73). LNCS: Springer-Verlag.

Kjeldskov, J., & Stage, J. (2004). New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies*, *60*, 599-620.

Kjeldskov, J., & Graham, C. (2003). A Review of Mobile HCI Research Methods. In *Proceedings of Mobile HCI 2003* (pp. 317-335). LNCS: Springer-Verlag.

Noldus (2005) Mobile Device Camera. Retrieved March 2, 2007 from http://www.noldus.com/site/doc200402054

Nyyssönen, T., Roto, V., & Kaikkonen, A. (2002). *Mini-camera for usability tests and demonstrations*. Paper presented at demo sessions of Mobile HCI 2002, Pisa, Italy, September 2002.

Pascoe, J., Ryan, N., & Morse, D. (2000). Using While Moving: HCI Issues in Fieldwork Environments. *Transactions on Computer-Human Interaction, 7*(3), 417-437.

Roto, V., Oulasvirta, A., Haikarainen, T., Kuorelahti, J., Lehmuskallio, H., & Nyyssönen, T. (2004). *Examining Mobile Phone Use in the Wild with Quasi-Experimentation* (HIIT Technical Report 2004-1). Helsinki, Finland: Helsinki Institute for Information Technology.

Rowley, D. (1994). Usability Testing in the Field: Bring the Laboratory to the User. In *Proceedings of CHI 1994* (pp. 252-257). ACM Press.

Waterson, S., Landay, J. A., & Matthews, T. (2002). In the Lab or Out in the Wild: Remote Web Usability Testing for Mobile Devices. In *Proceedings of CHI 2002* (pp. 796-797). ACM Press.

Winters, J. M., Story, M. F., Campbell, S., Lemke, M., Danturthi, S., Barr, A., & Rempel, D. M. (2001). Mobile Usability Lab: A Tool for Studying Medical Device Accessibility for Users with Diverse Abilities. In *Proceedings of RESNA 27th Annual Conference.*

KEY TERMS AND THEIR DEFINITIONS

Field laboratory – a configuration of laboratory equipment, such as video and audio recording devices, put together so that it can be taken into to field for data collection about the use and usability of mobile technologies in situ.

AV hard disk recorder – a video unit that records an external video and audio source directly onto a hard disk in a digital format that can be played back on a computer. The video recording is typically compressed when it is recorded resulting in manageable file sizes.

Picture-in-Picture unit – a video unit that inserts a video image over a part of another one. The inserted video image is rescaled and thus loses a bit of quality in the process.

Quad display unit – a video unit that merges four different video signals into one composite signal. All four video images are rescaled and thus lose a bit of quality in the process.

Lapel microphone – a small microphone that can be clipped on to a person's collar or revere. The microphone is usually connected to a small transmitter that can be carried in a pocket or clipped on to the belt.

Camera-clamp – a tiny camera that can be clipped on to a mobile device such as a PDA or a mobile phone. Camera-clamps can be either cabled or wireless. The latter require a battery supply and a video receiver.

Third person view – a video recording of the user(s) of a mobile device and their immediate surroundings during an evaluation of use or usability from the perspective of a third person observing from a distance.

Close-up view – a video recording of the screen and buttons o a mobile device, such as a PDA or mobile phone, during an evaluation of use or usability. Usually captured with a mobile devices camera attached to the device.