



Flexible Hardware Configurations for Studying Mobile Usability

Antti Oulasvirta

Senior Researcher
Helsinki Institute for
Information Technology HIIT
PO Box 9800
FIN-02015 TKK
Finland
aoulasvirta@acm.org

Tuomo Nyysönen

Researcher
Helsinki Institute for
Information Technology HIIT
PO Box 9800
FIN-02015 TKK
Finland
tuomo@aidem.fi

Abstract

The main challenges for mobile usability labs, as measurement instruments, lay not so much on being able to record what happens on the user interface, but capturing the interactional relationship between the user and the environment. An ideal mobile usability lab would enable recording, with sufficient accuracy and reliability, the user's deployment of gaze, the hands, the near bodyspace, proximate and distant objects of interest, as well as abrupt environmental events. An inherent complication is that the equipment will affect these events and is affected by them. We argue that a balance between coverage and obtrusiveness must be found on a per case basis.

We present a modular solution to mobile usability labs, allowing both belt- and backpack-worn configurations and flexible division of equipment between the user, the moderator, and the environment. These benefits were achieved without sacrificing data quality, operational duration, or light weight. We describe system design rationale and report first experiences from a field experiment. Current work concentrates on simplifying the system to improve cost-efficiency.

Keywords

Mobile usability, usability labs, field studies, hardware

Introduction

Science and technology go hand in hand. It is commonplace that paradigmatic changes in research are preceded by both conceptual and methodological advances. Of all methodological developments in usability studies, controlled laboratory-based usability evaluation may have had the most wide-spread impact on day-to-day operations. The two-room usability laboratory setup, deploying mirrors and multiple video cameras to record interaction on a desktop PC, was a necessary complement to the novel notions and operationalizations of usability that Jakob Nielsen put forward in his seminal 1993 book. Laboratory-based testing has become the de facto standard of usability practice worldwide.

However, nothing similar appears to have taken place on the side of mobile technology, although the research area has existed for over 10 years (the conference *Mobile HCI* started a decade ago as a workshop). One reason may be technical: no solution has been proposed that is technically reliable, cost-efficient, and flexible enough. In fact, the number of systems presented in the literature is in the order of a mere dozen. Figure 1 presents a sample of four of systems. These previous systems can be divided into three classes: those that allow recording from (a) both moderator-controlled and user-worn sources, (b) user-worn sources only, or (c) device-based sources only. (Sources are typically cameras and microphones but can also include logs collected on the phone.)

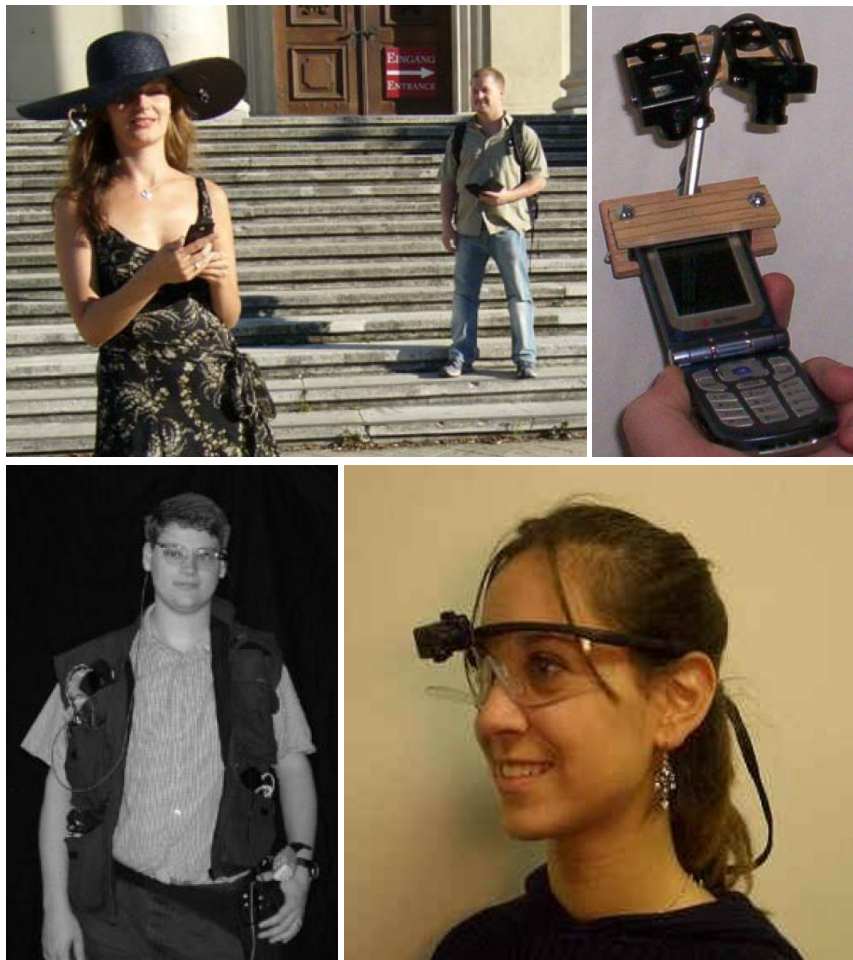


Figure 1. Previous systems (from left to right, top to bottom): Reichl et al.'s (2007) hat-worn system with moderator-controlled shooting, Schusteritsch et al.'s (2007) system for attaching

minicameras to a mobile phone, Lyons and Starner's (2001) vest-worn multi-camera system, and Applied Science Laboratories' (2008) forehead camera

In this paper, we extend on our previous work on mobile usability labs (Oulasvirta, Tamminen, Roto, & Kuorelahti, 2005; Roto et al., 2004). We present rationale and design of a "swiss army knife approach," operating with the principle of supporting multiple functions and configurations with one system. The equipment should support any of the abovementioned three configurations if the study at hand so demands. We argue that aiming for (a) modularity, (b) scalability, and (c) flexibility is crucial if the equipment is to be used across many studies. We argue that another reason for the absence of conventions in mobile labs may be that one cannot simply transfer the *thinking* behind laboratory-based setups to mobile conditions. The system should support capturing environmental events and the user's position in and orientation toward the environment in general. This ability is central in the class of context-aware applications, but also relevant in any use situation where the user's environment affects interaction.

In this paper, we describe our system Attentional Resources in Mobile Interaction version 2 (ARMiv2)¹. The ARMiv2 system supports both belt-mounted and backpack-mounted configurations of recording devices, as well as totally wireless ones. Wires and wireless transfer can be chosen according to mobility conditions. The system also supports environmental cameras. It integrates all video into a single stream that can be uploaded to a PC for analysis after a trial. In addition, it has the longest operational duration reported and is light weight for the user. The single most important quality, however, is its support for different configurations.

In the latter part of the paper, we report first experiences from a real-field deployment. Of the three desirable qualities mentioned above, our approach is targeted towards flexibility in particular. Modularity and flexibility are desirable from the perspective of experimental validity. However, although our system works, the solution is perhaps too complex to be cost-efficient as it places heavy demands on researcher training and system maintenance. We nevertheless believe these problems can be overcome and that the general solution represents a promising direction for mobile usability labs.

Rationale for field testing

We believe that usability studies should be conducted in the field if the key aspects of users' environment cannot be accurately simulated in a lab. For example, a student of mobile input devices may well get along with laboratory-based evaluations for most of the time. Real-world issues like timeout (due to interruptions), one-handed use (due to reserved modalities), application-switching (due to multitasking), or slow use (due to simultaneous walking) can be staged in the lab, given that they are known in advance. On the other hand, there are applications where this strategy does not work. For example, one cannot (easily) stage a whole city for a study of mobile maps or power relationships for a study of organizational use of mobile email.

We want to take this argument a bit further and present the following three conditions in which one should carry out usability studies in the field:

- The environment affects the interaction loop through the user. For example, users often interrupt themselves to respond to events perceived in the vicinity.
- The environment affects the interaction loop through the computer. For example, augmented reality and mixed reality interfaces have sensors that change the state of the interface according to environmental events.
- The user's actions are doubly-determined (i.e., they transform the environment as well as the interactive state). For example, when using location-aware mobile maps, a user's movement changes the state of the environment as well as the state of the map.

¹ The first system was built in 2004 (Oulasvirta et al., 2005).

There have been discussions around whether “it is worth the hassle” (Kjeldskov, Skov, Als, & Hoegh, 2004); that is, whether conducting evaluations in the field pays off in terms of increased ability to capture usability problems. The skeptics have based their arguments on experiments building on contrived operationalizations of mobility. For example, being in the field has been often operationalized by walking a pre-defined route. This, in effect, reduces environment to the role of a nuisance factor—a source of disruptive events and cognitive resource withdrawals. It is no surprise that the finding has been that field testing is less efficient in capturing usability problems. Environment is not fruitfully operationalized in such a way, rather one must think how environment may support or hinder interaction according to the three abovementioned conditions. Then, one must set up the testing situation so that such events can occur in a natural way and can be captured by the recording equipment.

Requirements

To demonstrate how varied the requirements for a mobile usability lab can be, and as a further argument for the swiss army knife approach, let us analyze the following three evaluation scenarios:

- *Usability testing of a mobile group media application.* Goal: Measure user performance in basic communicative tasks, comparing against a pre-defined baseline of errors and task completion times.
- *A real-world longitudinal study of a mobile collaborative awareness application.* Goal: Understand real-world use of the system in a team of information workers.
- *A comparative study of mobile map representations.* Goal: Assess which representation, 3D or 2D, is better for mobile maps in the task of locating buildings in a city centre.

For the first evaluation scenario, a setup that records only the user interface might be enough. The two latter scenarios require at least some track of what happens in the user’s environment. The third scenario is the most demanding in this respect. Ideally, it requires systematic second-by-second analysis of both bodily (turning of body, deployment of gaze, use of hands; analyzed post-experimentally from video tapes) and cognitive (verbal protocols) strategies, in addition to analysis of events at the interface and in the environment. Experience with several analogous studies of mobile systems has helped us gain perspective to mobile usability labs and spell out the following general goals:

- *Mobility.* The system moves with the user, capturing interaction reliably wherever and whenever it takes place both in indoor and outdoor contexts of use.
- *Captures embodied interaction.* The system captures both bodily and virtual components of interaction as well as environmental events that may have an influence on those aspects of interaction that are under scrutiny.
- *Unobtrusiveness.* The system does not itself bring about direct or indirect changes that would cause bias or distortion in ecological validity.
- *Multi-method support.* The system does not limit the researcher to one source of data.
- *Redundancy.* The system has multiple data capturing mechanisms, so if one method or source fails then the data is not lost.
- *Quality control.* The system allows the experimenter to be aware of the reliability of data captured both during and after an experiment. It can help the experimenter answer questions such as what caused missing data, what situations were data gathered, how reliably the data corresponds to the actual situations it comes from, and so forth.

Let us explain the background for these requirements. First, we took as our starting point that the core of mobility is that it foregrounds the changing relationship between the user and the environment. It raises new constraints and resources for interaction from context, and it allows users to involve and utilize new contexts. For the recording system it means that it should be able to record any significant action or event in context. In practice, depending on the situation, this may include near bodyspace of the user, distant and proximate physical objects that a user interacts with, ad hoc environmental events, as well as deployment of gaze. The view that the core of mobile human-computer interaction is in the triad *environment-user-computer* demands flexibility in the placement of cameras as well as their division among the environment, the user, and the device.

Second, a threat to *experimental validity* is that the system itself affects the phenomenon under study, for instance due to its (a) physical qualities—e.g., the weight of the system causing fatigue, (b) ramifications to the user's processing of the interface—e.g., a camera occluding the mobile device, or (c) social consequences—e.g., a hat not being acceptable indoors. Camera types, positioning, and form are crucial qualities that affect these problems. Another factor is the moderator. The presence of a moderator may also affect events and is not always desirable. This is yet another argument for modularity and flexibility for the selection of cameras and their positioning. The cameras should enable researchers to conduct studies without a moderator, preferably without sacrificing the ability to record in an environment. One solution is the utilization of surveillance cameras.

Third, the environment is not only of interest as such, but it also introduces noise, unexpected events, and it is the cause of technical unreliabilities. For example, due to a user walking very quickly, it may happen that the moderator is unable to reliably capture events in the front bodyspace of the user. These problems call for (a) redundancy in recording and (b) online quality control. The former can be addressed if the moderator can place "just in case" cameras to augment and back up the primary ones. The latter can be supported by providing a real-time copy of the A/V stream for the moderator.

Flexibility in practice

To address these requirements, we aimed for flexibility in the following four qualities of the system:

- *Camera types.* At the moment we have one model, "a sugar cube" (17 x 17 x 17 mm), but in the future a smaller "minitube" type (45 x 7 x 7 mm) will be explored.
- *Camera attachment devices.* The cameras attach using a pin, a shallow cell phone shell (Figure 4), a necklace (Figure 2), and a pole (Figure 2). With a pin we will be able to attach a minitube to a hat as in the LiLiPUT system (Schusterisch et al., 2007).
- *Wires.* Both wireless and cable transfer are enabled. Furthermore, by providing cables of different types and lengths, we enabled the positioning of user-worn, non-camera components either to a backpack or to a belt.
- *Offloading cameras to the environment.* We implemented an option to utilize environmentally placed cameras.

System Description

ARMiv2 is based on experiences and technology from ARMiv1 (Oulasvirta et al. 2005). It highlights the following features:

- *Belt-mounted, backpack-mounted, and wireless configurations are supported.* All non-camera equipment, except cables, attach to a belt worn by the user or can be placed in a backpack. A third alternative is to use wireless transmitters and let the moderator carry non-camera equipment. However, due to signal interference, this results in poorer video quality especially when outdoors.
- *Environmental cameras are supported.* Surveillance cameras can be assembled to different places of the study site and can be switched to *on* based on signal strength (Figure 2). These cameras offer a very high image quality and their number is not limited. This enables us to run studies where users can move about autonomously without a shadower. Alternatively, we can also run a study with moderator-controlled cameras (Figure 4).
- *Wireless and wired transfer are both supported.* All cameras can operate either with cables or wirelessly.



Figure 2. From left to right: All non-camera equipment attaches to a black leather belt worn by the user. In this setup the front camera is in a small black box (a necklace). (b) A lightweight pole attachable to almost any mobile device can host two minicameras; one camera capturing the face of the user, the other camera capturing the events on the display and keyboard. (c) Environmental cameras allow for higher quality third-person views. The receiver switches to the nearest camera automatically based on signal strength.

To our knowledge, these three are novel features in comparison to previous systems. The following are several other improvements that have been made in comparison to ARMIV1:

- *Higher quality integrated video.* New, compact MPEG-4 recorders provide the advantages of on-the-fly encoding, quick data transfer, and post-trial delivery via USB. Maximum recording times are about ten times longer than a cassette recording. Lower power consumption, price, and smaller size also promote these recorders. ARMIV1's original video *quad* was replaced by the smallest digital quad available. It was custom-built into a more compact and usable video hub with a smaller die cast aluminum housing and integrated power and camera signal connectors. With the integrated connectors on the hub and cameras it is easy to use one to four cameras and arrange them into the quad view simply by unplugging and plugging them into different inputs. This combination of a recorder and a hub results in 234 kbps AVI files, with a resolution of 640 x 544, a frame rate of about 17 fps, and a MS ADPCM 180 kbps 4 bit audio sound track.
- *Increased operational duration* (optimally up to 4 hours without a battery change). In practice we reached about 2.5 to 3.0 hours, which is a significant increase in comparison to previous systems. LiLiPUT (Schusterisch et al., 2007) and ARMIV1 both achieved about 90 minutes.
- *Lower weight.* Due to using hand-crafted casings and generally the smallest available components, the user-worn system part weighs less than 2 kgs. The previous version required the user to carry a backpack of about 4 kgs. LiLiPUT's user-worn part weighs about the same, but includes less functionality.

Figure 3 conveys the system architecture. The stabile part is marked with dark grey coloring; all receivers for A/V data can be either wireless or wired.

Cameras

As in ARMIV1, the cameras are the core of the system. In general, selecting camera elements are a matter of balancing performance, size, cost-efficiency, and versatility.

In ARMIV2, all miniature cameras were equipped with an automatic electric shutter and white balance. Camera elements were originally bought as miniature dome surveillance cameras that

were modified for pole attachment. A miniature audio amplifier and a microphone, originally from the same surveillance product, were also re-housed and modified. Images from these cameras on the pole (see Figures 2 and 6) are in the PAL format, still effective image quality is approximately 280 TV-lines or less. The necklace camera element performance is 350 TV-lines with dimensions of 30(H) x 30 (W) x 34 (D) mm and weighing 30 grams.

The environmental cameras provided full quality in PAL resolution, by one chip computer-controlled display (CCD) and programmability and versatility for all light conditions. When equipped with wide-angle lenses, the camera view can cover a space size of a meeting room. A practical limitation is posed by the fact that all video streams are integrated into one, which lowers the quality of any single stream. According to our experience, environmental cameras work best in small spaces like desktops and hallways. For outdoor experimenting, moderator-controlled camera operation is a better option.

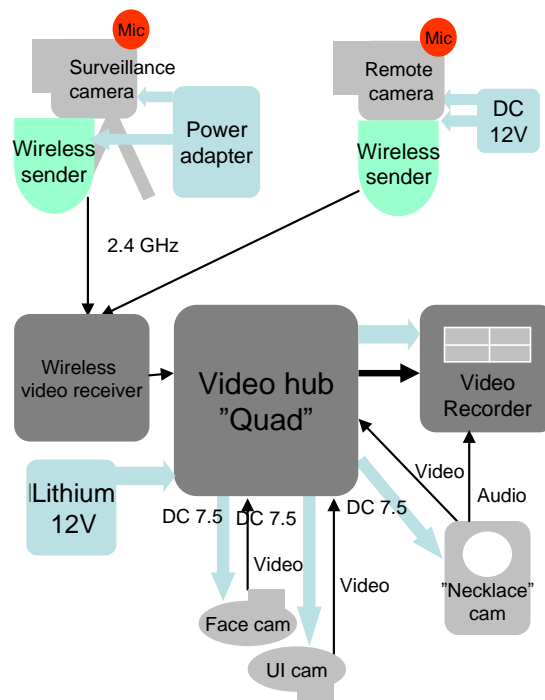


Figure 3. System diagram for the current setup

Current setup

Our current setup consists of the following three kits (see Figure 3 for the system diagram):

User's kit. The user-worn part (Figure 2) includes the following:

- One camera holder (pole) for attaching two cameras to the participant's device. One camera shoots the display, the other looks up to the participant's face.
- A necklace-camera with a microphone
- Leather belt, the following are carried on a leather belt:
 - A wireless 2.4 GHz video receiver
 - A video hub that integrates all the signals to the recording device and provides adjustable voltage for the attached cameras.
 - A Video HDD MPEG-4 recorder
 - Three battery packs (Lithium 12V)
 - Necessary cables

Moderator's kit. The environmental camera setup consists of the following:

- Three surveillance cameras
- Three wireless 2.4 GHz video senders
- Three 110-240 VAC to 12 VDC power adapters
- Video staves and fixing equipment

Alternatively, a moderator-carried camera (a Sony DCR-TRV30E PAL mini-DV digital handycam, equipped with a VCL-ES06 x0.6 wide-angle lens) can be used. In addition, we have previously explored the option of hiding the moderator-controlled camera in a cell phone shell to make it less visible in public places (Figure 4).



Figure 4. One of the earlier setups had the moderator-controlled camera built in a fake shell of a mobile phone. The wire goes through the sleeve to a recorder on the waist. This solution allows for quiet, non-visible recording in public places.

Maintenance kit. The following are tools for running and preparing the tests and maintaining the setup.

- Battery rechargers
- Wireless receiver for setting up the environment cameras
- Travel cases for the equipment

The kits fit into two aluminum briefcases that can be carried to the site of experiment (Figure 5). The total worth of the current setup, including labor and component costs, is about 10,000€.



Figure 5. The whole kit fits into two aluminum cases that can be taken to the site of study. This case includes environmental cameras; the other kit contains the user-worn components.

First Experiences

To acquire feedback on the system, we deployed a recent study using the system for mobile maps. In this field experiment, we asked 16 participants to carry out localization tasks in a city center, with either a 2D or 3D mobile map (Oulasvirta, Estlander, & Nurminen, in press). No environmental cameras were used in this ultra-mobile study where 16 different city sites were visited. A moderator-controlled, high definition camera with a wide-angle lens was used instead.

Figure 6 depicts output data after integration of the video output (on the left) with a reconstruction of the interface we automatically built from the logs. Using custom-made software, the log files from the mobile map software were integrated and then manually synchronised with the four-channel video.

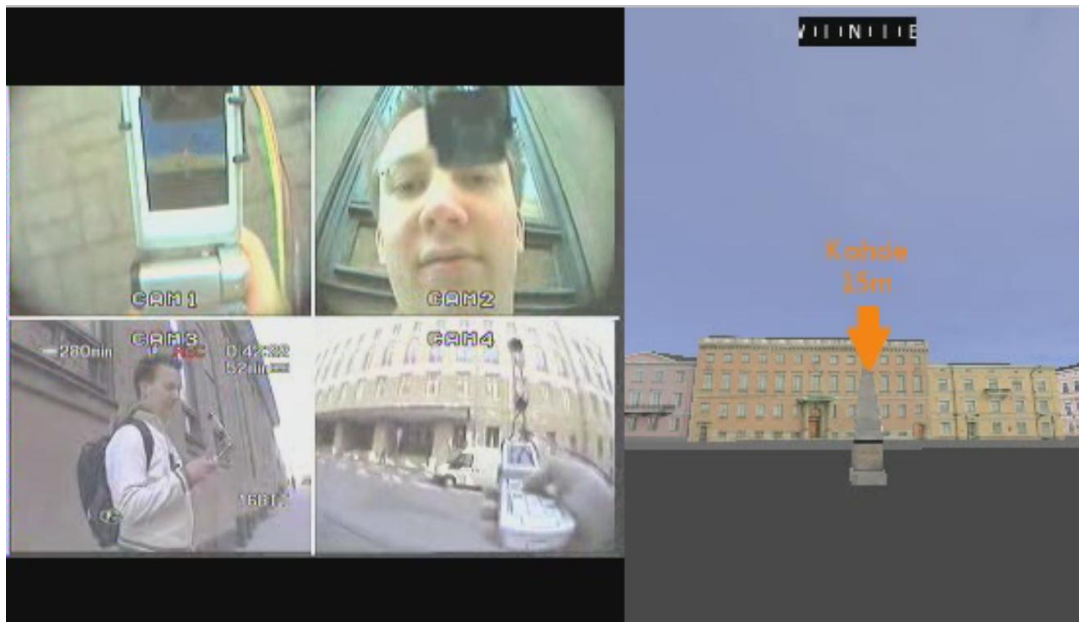


Figure 6. Sample data output from the first field experiment (Oulasvirta et al., in press)

The results of the experiment showed that 2D was better in both localization and navigation tasks, in all dependent measures we deployed. The mobile lab enabled us to gather unusually rich data from verbal protocols to subjective workload measures and from video observations to integrated interaction logs. These data afforded for going beyond typically measured performance measures, like task completion times and errors. Particularly, we were puzzled by the finding that 2D was associated with not only faster performance overall but lower cognitive load than 3D. The 3D maps were typically thought to lower cognitive demands by making it easier to match what was seen in the surroundings to what was seen on the map (ego-centric alignment). Finally, we came up with the hypothesis that 2D users can better use their body to find alignment between the map and the visual scene. We carried out a second-per-second analysis of the video tape that proved this hypothesis right. We found that 2D users often tilted the device in their hands, used their upper body more, and were generally more efficient in their deployment of gaze (i.e., they found effective cues quicker). Video tapes in combination with interaction logs and verbal protocols thus enabled us to find an explanation that we could have not found otherwise. These results were analyzed to draw implications to a new version of the 3D map that is functionally and representationally closer to the 2D map while retaining some advantages of the 3D map. The full results are presented in Oulasvirta et al. (in press).

We were generally happy with the achieved quality. The final integrated video files were 2000 kbps MPEG-1 streams, with a resolution of 520 x 320, a frame rate of 25 fps and a MPEG-1 Layer 3 64 kbps mono sound track. The sound track was passed through a 3000 Hz low-pass filter to dispose of a high-frequency, whining noise. For this data, we were able to carry out manual coding of events with good inter-coder reliability (Kappa .75), even for quite subtle behavioral measures (such as turning of head, turning of device in hand, body posture, and walking), that were coded with one second accuracy. The freeware video annotation software InqScribe was used for coding.

There were a couple technical difficulties. We had to change recorder and battery type, and consequently re-ran three more subjects. The more persistent problems were due to environmental conditions and accidental events that hampered the use of one or more of the minicameras. Direct sunlight to the face camera, shutter adapting excessively to large contrasts in camera image, necklace camera temporarily occluded by clothes, random compression artifacts, and rain effectively prevented coding of some of the variables—particularly when these

effected several cameras at the same time. These problems can be addressed in the future by camera selection, camera placement and attachment, and changing the video recorder.

We also noticed effects due to form and design. While our device-mounted parts include only two small minicameras weighing a few grams, they nevertheless affected how the mobile device was held in the user's hand, which created a bias to keep it more upright than normally. Second, while our setup attempted to minimize the visibility of cameras, particularly the pole on the device caught the attention of passers-by, although so sporadically that we did not assess this to pose a bias to the validity of the conclusions.

Recommendations

At the moment, we are focusing on improving and augmenting the system in several respects. We attempt to improve the system using the following approach:

- Increase the operational duration by exploring different battery and receiver combinations
- Implement synchronization of logging with recording events
- Improve technical reliability of the minicameras by trying alternative models
- Improve the ergonomics of the belt

We also plan to implement functionality for remotely triggering recording events via Bluetooth. This solution, enabling us to either start recording remotely or as a function of some context information, could save critical battery life.

Sensor extensions are also possible. Our collaborators at Helsinki School of Economics have developed a version based on ARMIv1 that integrated psychophysiological sensors.

As we have solved the most pertinent technical problems, the most significant problem at the moment is posed by the complexity of the system. Even though the system is able to support flexibility, training researchers to use it, constructing a new setup for a study, as well as maintenance and repair, pose such an overhead that the system's adoption even within the group has been slower than hoped.

Conclusion

This paper has proposed how the varying requirements of mobile usability studies could be addressed by a single mobile usability lab setup. In general, we find the approach of aiming for flexibility the key to the development of mobile usability labs in the future. The key endeavor in aiming for modularity and flexibility is the design the core of the system, in our case the belt-worn part. Flexibility can also be pursued in camera selection, attachment, and positioning, as well as their division between the user, the moderator, and the environment.

The system we developed has the following qualities:

- Support for both belt-mounted and backpack-mounted configurations, as well as totally wireless ones
- Environmental cameras, not supported by previous systems
- Wireless and wired transfer both supported and flexibility in cable positioning on the user's body
- Higher quality integrated video than our previous systems
- Increased operational duration, optimally up to 4 hours without battery change, which is more than previously achieved
- Lower or equal weight on the user than previous systems

Practitioner's Take Away

The swiss army knife approach to mobile usability labs centers around implementing a base system for non-camera equipment that allows enough flexibility to enable belt-worn, backpack-worn, and wireless configurations. For the camera equipment, the following goals are important:

- Flexibility in camera types, from small to large and from low fidelity to high fidelity
- Options in camera attachment devices (a pin, a shallow cell phone shell, a necklace, a pole, and so on)
- Options in cables and wireless transfer
- Offloading cameras to the environment by using signal strength as the switching criterion

At the moment the resulting system is somewhat expensive, around 10,000 Euros including hardware and craft, but it is on par with comparable non-mobile usability labs. The most significant challenge is to improve the user experience for the moderator by streamlining the process of using such a setup and by improving the interfaces between parts of the system.

Acknowledgements

We thank Sara Estlander, Vikram Batra, Peter Fröhlich, Kent Lyons, Virpi Roto, Antti Nurminen, Antti Salonen, and anonymous reviewers for pointers, discussion, and help. This work has been supported by the Academy of Finland project ContextCues and the EU FP-6 projects PASION and IPCity.

References

- Applied Science Laboratories (2008). Applied Science Laboratories–Eye tracking expertise. Retrieved on January 2, 2008 from <http://www.a-s-l.com>
- Kjeldskov, J., Skov, M.B., Als, B.S., & Hoegh, 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 *MobileHCI'04*. Springer-Verlag.
- Lyons, K. & Starner, T. (2001). Mobile capture for wearable computer usability testing. In Proceedings of *ISWC'01* (pp. 69-76). IEEE Press.
- Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Oulasvirta, A., Tamminen, S., Roto, V., & Kuorelahti, J. (2005). Interaction in 4-second bursts: the fragmented nature of attentional resources. In Proceedings of *CHI'05* (pp. 919-928). New York: ACM Press.
- Oulasvirta, A., Estlander, S., & Nurminen, A. (in press). Embodied interaction with a 3D versus 2D mobile map. *Personal and Ubiquitous Computing*, 13 (4).
- Reichl, P., Froehlich, P., Baillie, L., Schatz, R., & Dantcheva, A. (2007). The LiLiPUT prototype: A wearable environment for user tests of mobile telecommunication applications. In Extended Abstracts of *CHI'07* (pp. 1833-1838). New York: ACM Press.
- Roto, V., Oulasvirta, A., Haikarainen, T., Lehmuskallio, H., & Nyyssönen, T. (2004, August 13). Examining mobile phone use in the wild with quasi-experimentation, *HIIT Technical Report*, 2004-1
- Schusteritsch, R. & Wei, C.Y. (2007). Towards the perfect infrastructure for usability testing on mobile devices. In Proceedings of *CHI'07* (pp. 1839-1844). New York: ACM Press.

About the Authors



Antti Oulasvirta

Antti Oulasvirta is currently the co-leader of the Ubiquitous Interaction research group at HIIT. He received his PhD from University of Helsinki in 2006 and was recently a postdoc at the School of Information, University of California Berkeley. His research interests include mobile HCI and cognitive science.



Tuomo Nyysönen

Tuomo Nyysönen works as an entrepreneur in the area of camera and media technologies. He has also worked as technical coordinator for miniature camera and compact multi-camera setup development since 2000 within different companies and organisations such as Nokia, HIIT, and Aidem.