observing the mobile user experience

Proceedings of the 1st International Workshop. Held in Conjunction with NordiCHI.

Benjamin Poppinga, Charlotte Magnusson, Wilko Heuten, David McGookin, Niels Henze,
Ginger B. Claasen, Martin Pielot, Håkan Eftring, and Jörn Peters

October 17, 2010.

workshop on

observing the mobile user experience

table of contents

Methods for understanding the mobile user experience

Charlotte Magnusson
Certec, LTH, Lunds Universitet
Box 118, 22100 Lund, Sweden
charlotte@certec.lth.se

Kirsten Rassmus-Gröhn
Certec, LTH, Lunds Universitet
Box 118, 22100 Lund, Sweden
kirre@certec.lth.se

Delphine Szymczak
Certec, LTH, Lunds Universitet
Box 118, 22100 Lund, Sweden
delphine.szymczak@certec.lth.se

ABSTRACT

Evaluating the user experience is often done in a laboratory. Methods for observing what happens in the wild are nonetheless being employed because they bring results that the traditional methods of evaluation do not yield. In this paper we describe and discuss methods used at our lab for understanding the mobile user experience. These methods range from quantitative to qualitative evaluation, and encompass diverse aspects of the design process. Finally we argue the need for combining different methods to obtain a better picture of real mobile usage.

1. INTRODUCTION

Observing the mobile user experience is a challenge. Situations change, and outcomes of tests are highly context dependent –eg. a person sitting on a bus will use a mobile device differently to one who is cycling. The social context also matters, since usage will not only be influenced by what you are doing but also who else is present and what your relations are. In this paper we provide an overview of different methods and discuss experiences, pros and cons of the methods we have used in our lab. Given our experiences, we argue that no single method is enough, and suggest that one needs to make use of a "smorgasbord" of techniques – both qualitative and quantitative.

Observing in the wild usually takes more effort than doing lab studies. To assess the utility of this additional work, in [1] the authors compared the evaluation in the laboratory and in the real world. Although the evaluation steps were exactly the same, the field study gave different and unique results compared to the study in the lab. The benefit of getting unique information from a field study then justifies that researchers consider the trouble of observing outside of the controlled environment of their laboratory. In the following we describe and discuss different methods used at our lab to make observations of users in the wild or at least in more real settings.

2. LOGGING

In several studies such as [2] and [3], logging has been used to keep track of what is happening during the experiment. One can log queries made to the interactive device as well as values taken from sensors. It is also possible to add some processing to recognize specific actions or usages (context sensing). In our studies we have mainly used logging as a support for the qualitative observations made during the test, but some data such as time to complete or number of turns lends themselves well to statistical analysis. The advantage of logging is that it is automatic, while the main disadvantage is that it can be difficult to interpret the recorded data. Context sensing can potentially help, but for more complex activities it is a true challenge to implement.

3. SEMI CONTROLLED OUTDOOR TESTS

To get feedback on basic components of the interface we have done a kind of test we call a semi controlled outdoor test. This type of test has a more lab type setup, where one takes care to randomize the order tasks are performed in. The test is also done on a specific location which mirrors some relevant aspects of the real world. Quantitative measures are recorded (such as time to complete, number of turns etc) and analyzed statistically. In addition an observer walks alongside (but slightly behind) the test person in order to make qualitative observations of gestures and behavior [4].

The advantages of this type of test is that it is less time consuming than setting up and performing a full scale study of mobile use. Another advantage is that one can focus on a single interaction component in a more full scale study a more complete interface usually needs to be implemented.

Problems with this approach is to know how relevant the results really are for the real usage situation, and also the lack of control over external factors like weather. It is also difficult for a person that walks slightly behind to observe all aspects of the interaction. Logging may help to some extent, but it is hard to extract more complex gestures from logs of magnetometer or gps data.

When testing GPS based applications one also has to consider the problem of GPS accuracy. Even at the same location this can vary from day to day. A workaround that can sometimes be used is to avoid connecting the GPS positions to real locations, and instead focus on how well the user is able to reach a virtual position (specified by the GPS coordinates).

4. REAL TIME LO-FI WIZARD OF OZ

Another method, as in [5] is to have a person acting as the mobile device, and observing the interaction. The questions posed by the user as well as the system responses provide valuable input early in a design process. The advantage of this method is that it is very easy to implement (no technology development needed) while the downside is that results depend heavily on the performance of the person playing the system. An additional problem is that there is a difference between talking to a person and using a mobile device.

A particular issue in our tests done with this method, was how to record the dialog without disturbing the situation too much. We ended up recording sound with a mobile phone – something which was seen to work well.

5. SIMULATIONS

When looking at mobile behavior one can also consider making use of computer simulations. In a simulation it is possible to investigate the effect of different parameters without external disturbances, and it is also possible to run very large numbers of tests. Thus simulations can be a useful tool for analyzing test results, or provide initial recommendations for certain interaction parameters [6].

The downside is that the usefulness of the simulation depends entirely on how well it is implemented. Factors important in real life may be missing, and unless the simulation design is carefully grounded in observed usage one runs the risk of getting useless results.

6. INTERVIEWS (SITTING DOWN)

To gain an insight into what happened during interaction, as well as into the context of use (skilled or novice user, intentions when using the device...) we need to ask the users. Interviews can be controlled or more open, but the researcher should avoid questions that can lead to confusion or use too technical. We often use a semi-structured interview approach: we have a set of pre-defined questions, but allow for follow up questions and discussions depending on the user answers.

The interviews can be done both before and after use, to gain insight in the context of use, the background of the user, and to obtain reflections on the test.

Interviewing is a standard technique and has been used in most of our studies, and also in many of the studies made by other researchers as mentioned in [1].

7. INTERVIEWS IN MOBILE CONTEXT

Interviews can also be done in the mobile context. We have noted that answers given while on the move are often different than those elicited when inside in a laboratory or an office. For this type of interviewing it is important to consider the recording. Just as in the previous method mobile phones or small recorders may be suitable. Video is more disturbing, but may be necessary if actions are to be recorded properly. One strength of this method is that events in the environment may trigger the discussion – something which may also be a weakness in case the external events are disturbing.

8. FOCUS GROUPS

During focus group discussions the researcher moderates the discussion while the end-users bring in their ideas. The discussion can be open or semi directed. To avoid missing important topics, or to give more concrete ideas to the group, some technology samples or prototypes can be brought to support the discussion.

Just as for ordinary interviews, we have found that bringing such a group outside is very useful. The group may talk about more technical issues in an office and then switch their focus to more situated topics when outside in the real context. Again the environment is both beneficial and problematic – it can not only trigger useful discussions.

Just as for situated interviews the documentation needs to be thought through – video is valuable, but audio may be enough depending on the context.

9. USER WORKSHOPS WITH DEMONSTRATION WALK

In participatory design, design workshops with potential stakeholders are a commonly used type of activity. The workshops are usually centered on scenarios which form the context for the prototype use. We have carried out workshops in which the scenarios are the users themselves, and their wishes and needs. After they have designed their paper / lo-fi prototype, they have been asked to act out the functionality of the prototype, and since the prototype in all cases has been navigation devices, the acting has included walking while demonstrating. This has led to a richer and more detailed dialogue around the actual functions and at what times you are interested in what kind of information. A potential problem is that users are not designers – they may find it quite hard to generate good designs, and the activity needs careful design and also often a moderator to ensure a useful outcome.

10. DIARIES

One way to get more long term and rich information about how persons use technology, or what kinds of needs they might have, is to ask them to fill in diaries over a period of time. This has been explored by eg. Gaver et al., who used it together with other sampling material in the Cultural Probes that he described in [7]. We used diaries together with scenario walks, contextual interviews and workshops as one method among others, not as a stand-alone tool. The diaries were mainly to collect travel information and to ask users about technology they might or might not use when planning or undertaking a trip. Every day had preprinted data to be filled in, such as the number and nature of trips, plus one or two preprinted questions from a larger collection of questions and also additional space to fill in any comments. In one case, the diary was filled in between two meeting occasions, in the other case after a larger workshop. The answering frequency was 100% in the first case, and only 5% in the latter, which shows that it might be better to send out diaries to be filled in before a meeting, rather than after.

11. VIDEO OBSERVATIONS OF ACTUAL PRACTICE

To have an insight into what people are really doing, it is possible to go out in the real world and try to video tape examples of use of the targeted technology. At our lab we have used such observations to obtain a better understanding of how users use their mobile phones when biking or walking. Those methods give information about what is happening in the real life. One disadvantage is that it doesn't inform the observer about the use of devices that are not yet possible to use. Another problem is that it can be really hard to catch the person to ask him or her why they did what they did.

Ethical questions can also arise from this kind of observation, and the observer should ask whenever possible if the video recorded can indeed be used.

Such video clips are also useful for bringing developers and designers closer to the complexity of real use. This type of videos provide the kind of richness which tends to be lost in methods like personas [8].

12. SIMULATED USE IN THE WILD

Most of our work has been to evaluate some aspect of interaction with a prototype that has limited, but accurate functionality in those parts that we intend to investigate. However, we have also recently carried out an evaluation sequence with a simulated functionality in context, where the test users had to perform actions that were not part of the future interaction. The task was to compare different navigation image types and decide which was most preferred [9]. The prototype was entirely without navigation functionality; instead it was the user who flipped between navigation images cued by the test leader and observer, who followed the test person. The unnecessary flipping of pictures seemed not to disturb the users much, and they were able to walk with speed. Aside from the drawbacks mentioned previously, the simulated use and the observation by following made it doubly difficult to be able to know what information the user really received. It occurred more than once that the user flipped the image at an incorrect time or accidentally flipped twice.

13. DISCUSSION AND CONCLUSION

The above discussed methods probe different aspects of the mobile usage situation. On the whole we agree with what was already stated in [10] that one needs to make use of several methods in combination in order to obtain a good understanding of the user experience. Although longitudinal methods are good for existing technology, they tend to be hard to use in the design process due to the times involved. Instead one often has to probe potential future use by shorter tests and design activities. In doing so we have found it important to use a variety of methods, and to make use of both qualitative and quantitative approaches.

A problem common in many of our studies is how to observe what the user is doing. If you are walking a little behind (which you have to in order not to influence the test person) it becomes hard to observe everything that happens. The actual activity of having to walk outdoors also introduces some specific problems:

- It isnt possible to carry out tests in all weather types
- You cant expect people to walk very far, especially not when you are working with elderly persons or persons with mobility problems
- You need to find safe test environments for persons with visual impairments
- People have different walking speeds

One particular problem we have noted is the difficulty of observing the interaction if feedback is given through earphones or vibration. In several studies we have made use of the loudspeaker of the phone just to allow the observer to gain access to the same output that the user is experiencing – but this is for many use cases quite artificial, and it could be worth exploring to have the observer get the same feedback as the user through an external device. A possible setup would be if both users have mobile phones and the user phone sends messages to the observer phone to generate the appropriate feedback.

We also note that simulations based on observed user behavior can be quite useful. Since simulations take much less time than real outdoor tests, we have found them a valuable complement when it comes to understand navigational behavior. How useful it is of course depends on the type of interaction studied, but (just as [6]) we find simulations a tool which should be considered.

In any design process the role of the user study is also to allow the users to participate in the design process. Thus, methods need to be combined in such a way as to help give the users the appropriate concrete grounding (by allowing them to experience existing technology) as well as to give tem visions and suggestions of future solutions [11]. Most persons find it hard to know what kind of future technology they want and how they think it should be designed. In fact, when faced with the question what do you want the most common answer is what can I get. Thus, it is the responsibility of the researcher or designer to work together with the users in order to explore the future design space.

To conclude: there is no single best method observing the mobile user experience. Instead one has to put together a set of probes to try to obtain an accurate understanding of the situation and the usage. Which combination is used depends not only on the kind of usage studied, but also why it is studied – are we observing existing technology, or trying to understand how possible future technology is to be designed?

14. ACKNOWLEDGMENTS

We thank the EC which co-funds the IP HaptiMap (FP7-ICT-224675). We also thank VINNOVA for additional support.

15. REFERENCES

[1] C. M. Nielsen, M. Overgaard, M. B. Pedersen, J. Stage, and S. Stenild. It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, pages 272–280, New York, NY, USA, 2006. ACM.

[2] K. A. Hummel, A. Hess, and T. Grill. Environmental context sensing for usability evaluation in mobile hci by means of small wireless sensor networks. In *MoMM '08: Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia*, pages 302–306, New York, NY, USA, 2008. ACM.

[3] K. Church, J. Neumann, M. Cherubini, and N. Oliver. The "map trap"?: an evaluation of map versus text-based interfaces for location-based mobile search services. In *WWW '10: Proceedings of the 19th international conference on World wide web*, pages 261–270, New York, NY, USA, 2010. ACM.

[4] C. Magnusson, K. Rassmus-Gröhn, and D. Szymczak. The influence of angle size in navigation applications using pointing gestures. In *The fifth International Workshop on Haptic and Audio Interaction Design (HAID)*, September 2010.

[5] C. Magnusson, M. Pielot, M. Anastassova, K. Rassmus-Gröhn, K. Tollmar, and S. Roselier. The mobile oracle: a tool for early user involvement. In *MobileHCI '09: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–2, New York, NY, USA, 2009. ACM.

[6] J. Williamson, S. Robinson, C. Stewart, R. M. Smith, M. Jones, and S. Brewster. Social gravity: a virtual elastic tether for casual, privacy-preserving pedestrian rendezvous. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1485–1494, New York, NY, USA, 2010. ACM.

[7] B. Gaver, T. Dunne, and E. Pacenti. Cultural probes. *Interactions*, 6(1):21–29, 1999.

[8] J. Grudin and J. Pruitt. Participatory design and product development: An infrastructure for engagement. In *Proc. PDC 2002*, pages 144–161, 2002.

[9] H. Stigmar and K. Rassmus-Gröhn. Usability evaluation of navigation maps for visually impaired users. In *Workshop on Methods and Techniques of Use, User and Usability Research in Geo-information Processing and Dissemination*. University College London, April 2010.

[10] M. Kaulio and I. Karlsson. Triangulation strategies in user requirements investigations: a case study on the development of an it-mediated service. In *Behaviour and Information Technology, 17*, pages 103–112, 1998.

[11] Y. Rogers and V. Bellotti. Grounding blue-sky research: how can ethnography help? *interactions*, 4(3):58–63, 1997.

Mobile user experience beyond the laboratory: Towards methodology for QoE-QoS evaluation in natural user environments *(position paper)*

Katarzyna Wac
Carnegie Mellon University, US
katewac@cs.cmu.edu

Anind K. Dey
Carnegie Mellon University, US
anind@cs.cmu.edu

ABSTRACT
Inevitably, mobile applications and services on a growing scale assist us in our daily life situations, fulfilling our needs for leisure, entertainment, communication or information. However, user acceptance of a mobile application depends on the application's perceived quality of experience (QoE) and it also includes the criticality of the application to the user's context and situation at hand. Statistics for usage of mobile applications provided via 'app stores' show that more than 50% of these applications never reach a minimal required user acceptance level, and get removed from the store. However, despite the importance of understanding of the mobile user QoE, a sound methodology for evaluation of this experience, and of factors influencing it, does not exist. Moreover, this QoE relates to the level of quality of service (QoS) provided by the underlying service and network infrastructures, which usually is provided at 'best-effort' level. Therefore, in our research we aim to provide a set of rigorous and robust methodological steps to be taken to quantify a mobile user QoE in his natural environments and different contexts, and to analyze its relation with the underlying QoS. We aim to evaluate the applicability if the methodology in a large-scale mobile user study for a set of widely used mobile applications.

INTRODUCTION
Growing availability of diverse interactive mobile applications envisaged to assist us in different domains of our daily life, make their perceived Quality of Experience (QoE) increasingly critical to their acceptance. "If it's slow, I won't give my credit card number" refers to QoE expectations of a typical mobile commerce application user [1]. These expectations can be different given the user's previous experiences with an application or application's criticality to the user's task at hand. Yet, to date, due to lack of resources, the evaluation of QoE perceived by application users is mainly done with use of qualitative methods focusing on applications' usability [2] (as a component of QoE), and it is conducted in a limited time span in controlled laboratory environments - not resembling the users natural daily environments, where the application is to be used later on. Hence, the results of such evaluation may help to discover the mobile application's serious usability issues, but may not help to discover issues that are relevant to the world outside the lab.

The issues that are relevant to the world outside the lab relate to, *e.g.*, a non-deterministic quality of service (QoS) and particularly performance of the underlying network infrastructures supporting the application execution and mobile service delivery (Figure 1). The QoS can be quantified by delay, jitter and network capacity, and usually is provided at 'best-effort' level, *i.e.*, without any guarantees. Yet the QoS is critical to mobile user's QoE, especially for highly interactive mobile applications, that delivery depends on a frequent data transfers over the underlying network infrastructures.

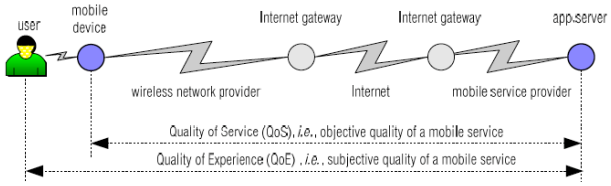


Figure 1: A concept of QoE and QoS in a mobile service delivery

Moreover, also a common practice for QoE provision is that mobile application designers use their own judgment and perception of application's experience and ease of use as a bellwether on application's perceived QoE by a mobile user [2].

The overall effect of that situation is that users whose QoE expectations are not satisfied, may simply stop using the applications or switch to another provider. For example, it is estimated that there are on average 200 new applications available daily in the online store for the Apple's iPhone platform, however, due to the low QoE perceived for more than half of them, they do not achieve a critical mass of user acceptance and are withdrawn from the store's offer.

The challenge is that rigorous and robust scientific methods, tools and engineered systems for application's perceived QoE evaluation in user natural daily environments do not exist [3]. Namely, there exist separate methods for usability evaluation in an HCI community [2, 4] and separate methods for evaluation of QoS and performance of the application's underlying network infrastructures in a data networking community [5-7]; the former methods are qualitative, while the latter are quantitative. Both methods' types can serve to acquire

quality results in their dedicated area of applicability; however, due to dichotomy between these two scientific communities, the scientifically proven methodologies and approaches combining both types of methods and quantifying user's perceived QoE in his natural daily environments and drawing conclusions upon applications acceptance, are non-existing.

OBJECTIVE AND APPROACH

Therefore, the objective of our research is to bridge this gap and develop a rigorous and robust, scientifically proven methodology for reliable real-time evaluation of interactive mobile applications' perceived QoE in natural daily user's environments, with variable QoS provisions.

In our approach we focus on already implemented and operational interactive mobile applications available to a typical mobile user; assuming that those applications underdid a cycle(s) of (re)design and usability tests in a laboratory environment, but we do not have an insights into that data.

Our approach is as follows. We will identify and analyse the existing as well as emerging qualitative methods for evaluation of usability and quantitative methods for evaluation of QoS and performance of mobile computing applications. Based on these methods, we will propose a novel methodology for a real-time quantitative evaluation of mobile applications' perceived QoE in user's natural daily environments. We have a long-standing successful history of research on measurements-based QoS and performance evaluation methods for interactive mobile applications [8, 9]. We have successfully used this methodology in a healthcare domain, *i.e.*, interactive applications for health telemonitoring and teletreatment [10-12].

To quantify the mobile user's QoE, the methodology will first require defining it as a set of variables called *dependent* (*i.e.* target) variables. Then the methodology will require defining a set of mutually exclusive and collectively exhaustive variables influencing this QoE; those variables are called *independent variables*, and they can include for example user context like location, time, social settings, *etc.* Both sets of variables must be defined based on the existing scientific literature and documented expert knowledge.

Furthermore, for a given interactive mobile application, the methodology will require to employ set of qualitative methods in order to derive *new* independent variables, not indicated in the HCI neither the networking communities so far, but important for a mobile user experience in his natural daily environments. A qualitative method, which can be used for this purpose is the *Experience Sampling Method* (ESM) [13]. The ESM is based on occasional user surveys; based on the elapsed time interval, an event occurrence or at random. Since we aim to evaluate user's perceived QoE while interacting with a mobile application, the ESM could be implemented in a form of a short, mobile-device based survey appearing to user after each use of this application. The survey will pose some open-ended questions to get the user's 'on-spot' real-time, spontaneous opinion on the mobile experience. New independent variables will be '*grounded*' as derived from the answers acquired from this user [14, 15]. The ESM method must be designed and deployed such that it does not influence the experience and behaviour of a mobile application user, but that it enables to gather information relevant and predictive for this user's QoE evaluation.

As the evaluation will be done in natural daily user's environments, the methodology will provide requirements and guidelines for the instrumentation of the mobile application and the QoS and performance of its underlying service infrastructure, such that a state of those variables (including the result of the ESM), is continuously and accurately logged in real-time in an automatic manner, *i.e.*, non-intrusive to the mobile user.

Moreover, having defined sets of dependent and independent variables, and having the system instrumented for measurements of those, the methodology will require reusing the exiting analysis methods for evaluation of variables relations and possible causality.

To analyse possible relations and causality between variables, the methodology requires occasional involvement of a mobile user into the data analysis process. Namely, a mobile user needs to be interviewed about their application's usage patterns and experience, and data must be matched to the data automatically logged in the application and service infrastructure. The interview will be based on the completion of detailed diary of the previous 24-hour period, as suggested by the *Day Reconstruction Method* [16], breaking the day into episodes described by activities, locations and times, and the mobile application usage and experiences during these times. During the interview, users will explain in more details their results for ESM, and these results will be compared to the state of other independent variables logged in the system. This way causalities and relations specific to this user could be pointes out, while any inconsistencies could be clarified.

The methodology will then provide guidelines on how to statistically analyze and interpret the acquired (qualitative) surveys data and (quantitative) measurement data for analysis within a user for one or multiple interactive mobile applications (*i.e.*, idiographic approach), within a population of users of one given mobile application and furthermore between a populations of users of different mobile applications. The data analysis will possibly include advanced statistical methods (*e.g.*, multivariate) and machine learning techniques for patterns recognition in data.

CONTRIBUTIONS AND EVALUATION

Our research will bringing together and expanding upon recent advances and methodologies in key scientific and

technical areas, like evaluation methods for human computer interaction, QoS and performance evaluation methods and tools for mobile computing, real-time machine learning and prediction. Along the realization of our approach, we will conduct research on critical issues like: (1) definition of QoE expected and required for interactive mobile applications; the definition must integrate multiple views: the application and its underlying infrastructure views (*e.g.*, interactions and provided QoS and performance) and the user view (*e.g.*, past experiences and expectations, current application's perception and its criticality to the task at hand), as well as user's context; the definition must also delineate a role of the user's affect in his perceived QoE; (2) reliable real-time capturing of user's perceived QoE and state of parameters influencing this QoE in his natural daily environments, including variable 'best-effort' state of QoS; (3) an automated and accurate inference of user QoE state (4) accurate and real-time recognition of QoE patterns based on data mining and machine learning techniques. These challenges become even more complex if the system is required to be accurate and operational in real-time and to generalize to novel situations (*e.g.*, novel applications, or novel user's interaction patterns).

With use of the proposed methodology we can gain deeper understanding of use of interactive mobile applications, quantify their user's QoE and relation with the underlying QoS, and point out areas of improvement for these applications towards their better usability and higher user acceptance. Our methodology will be evaluated with a set of widely available mobile applications, whose users tend to have high QoE expectations and hence to be sensitive to (changes in) the perceived QoE, like (a) streaming multimedia content, *e.g.*, YouTube or Internet-based radio; (b) highly interactive web browsing, *e.g.*, Google Wave, e-banking or e-commerce; (c) multiplayer online gaming; and (d) VoIP video-conferencing, *e.g.*, Skype.

CONCLUSIVE REMARKS

In this paper we have presented our research approach towards defining a methodology for quantifying a mobile user's experience (QoE) in his natural daily environments and relating of this experience to the performance (QoS) of the underlying service and network infrastructures. The methodological approach involves the user in the evaluation process twofold. First, it requires gathering 'on-spot' spontaneous information about the user's mobile experience by employing the Experience Sampling Method for interaction with the user directly after each mobile application usage. Second, it requires a retrospective analysis of the user's experience and of a state of factors influencing it, by employing Day Reconstruction Method for the last 24-hours recall. Our current work focuses on definition of the methodological steps; while future work includes a design of its evaluation in a large-scale mobile user study for a set of widely used mobile applications.

ACKNOWLEDGMENTS

Research conducted by K. Wac is sponsored by Swiss SSER (C08.0025) and Swiss NSF (PBGEP2-125917). This work is also partially supported by and the US NSF FieldStream project (0910754).

REFERENCES

1. A. Bouch, *et al.*, "Quality is in the eye of the beholder: meeting users' requirements for Internet quality of service," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems, The Hague, The Netherlands, 2000.
2. A. Dix, *et al.*, Human Computer Interaction: Prentice Hall, 2004.
3. J. Kjeldskov and C. Graham, "A review of mobile HCI research methods," in Lecture Notes in Computer Science, 2003, pp. 317–335.
4. K. Hornbæk, "Current practice in measuring usability: Challenges to usability studies and research," in International Journal of Human-Computer Studies, 2006, pp. 79-102.
5. K. Wac and R. Bults, "Performance evaluation of a Transport System supporting the MobiHealth BANip: Methodology and Assessment," MSc, MSc Telematics, University of Twente, Enschede, the Netherlands, 2004.
6. K. Salamatian and S. Fdida, "Measurement Based Modelling of Quality of Service in the Internet: A Methodological Approach," in Intl Workshop on Digital Communications: Evolutionary Trends of the Internet, 2001, pp. 158-174.
7. F. Michaut and F. Lepage, "Application-oriented network metrology: Metrics and active measurement tools," IEEE Communications Surveys & Tutorials, vol. 7, pp. 2-24, 2005.
8. ITU-T, Definitions of terms related to quality of service Recommendation E.800, 2008.
9. ITU-T, Communications Quality of Service: A framework and definitions, Recommendation G.1000, 2001.
10. K. Wac, *et al.*, "Measurements-based performance evaluation of 3G wireless networks supporting m-health services," in 12th Multimedia Computing and Networking (MMCN05), San Jose, CA, USA, 2005.
11. K. Wac and R. Bults, "Performance evaluation of a Transport System supporting the MobiHealth BANip: Methodology and Assessment," MSc Telematics, University of Twente, the Netherlands, Enschede, 2004.
12. R. Bults, *et al.*, "Goodput Analysis of 3G wireless networks supporting m-health services," in 8th International Conference on Telecommunications (ConTEL05), Zagreb, Croatia, 2005.

13. J. M. Hektner, *et al.*, Experience sampling method: Measuring the quality of everyday life: Sage Publications, Inc, 2006.
14. P. Y. Martin and B. A. Turner, "Grounded theory and organizational research," The Journal of Applied Behavioral Science, vol. 22, p. 141, 1986.
15. H. van Kranenburg, *et al.*, "Grounded Contextual Reasoning enabling Innovative Mobile Services," in 5th Workshop on Applications and Services in Wireless Networks (ASWN05), Grenoble, France, 2005.
16. Kahneman, D., A. Krueger, D. Schkade, N. Schwarz, and A. Stone, "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method, in Science, " 306 (5702), p.1776-1780.

UCD Research Methods for Mobile Industry Practitioners

Chandra Harrison
System Concepts
2 Savoy Court
London WC2R 0EZ
+44 (0) 2072403388
Chandra@system-concepts.com

Sam Medrington
System Concepts
2 Savoy Court
London WC2R 0EZ
+44 (0) 2072403388
Sam@system-concepts.com

Whan Stransom
System Concepts
2 Savoy Court
London WC2R 0EZ
+44 (0) 2072403388
Whan@system-concepts.com

ABSTRACT

In this paper, we introduce key areas of mobile device design that are barriers to creating a solid user experience for consumers including those with visual impairment. We also briefly review a variety of methods used in mobile device research 'in the wild' and the lab.

Following phases of a user-centred design (UCD) process, we review methods we have used in formative **research**, conceptual **design** development, **evaluation**, and post purchase experience following product **release**. The examples cover mainstream mobile handsets, mobile websites, design guidelines and out-of-the-box experience and include insight gained from attending mobile related practitioner and consumer events.

The examples offered are from a UCD practitioner's perspective of data gathering with consumers, consulting with designers, evaluating mobile products prior to release, assessing the post purchase experience, and attending mobile specific professional events, including events about accessibility of mobile devices.

1. INTRODUCTION

The UCD process should be an iterative cycle of research, design, evaluation and monitoring after release [1]. This process is applied to many different types of products within the mobile space including mobile phones, network providers' websites, and packaging of mobile devices. Having the right method available to gain the right insights at different phases is vital to successful product development. In practice using the right method can prove problematic due to the complexity of the mobile research space, restrictions imposed by the client, a lack of knowledge of methods, and issues around usability and accessibility.

The purpose of this workshop is to bring together people from industry and academia to exchange methods and experiences related to observing mobile device UX. Therefore, in this paper we briefly present specific applied examples of observing the mobile user experience in practice. Over the past two years we have been involved in many activities that provide insight into the relative merits of various methods and highlight the barriers to design.

In this paper, we discuss barriers particular to practitioners in the design of quality mobile experiences. We also present brief examples of applied research during the four phases of UCD (formative research, conceptual design, evaluation and post purchase). Due to non disclosure agreements we are unable to discuss particulars of the specific clients, but we do discuss the benefits and constraints of methods. These insights are based on our activities as practitioners in an agency over the past few years.

2. BARRIERS TO MOBILE RESEARCH

2.1 Complexity of Mobile Space

One of the key factors to researching and designing quality mobile user experiences is understanding the complexity of the space and the multiple factors that need considering. To design quality products we need to consider:

- the variety of users (e.g. able bodied and disabled)
- the hardware (e.g. screen size, button placement)
- the software (e.g. proprietary, open source)
- the content (e.g. websites, applications)
- the network provider (e.g. coverage, costs)
- the connection speed (e.g. 3G, wifi)
- contextual issues (e.g. lighting, glare, noise)
- functionality (e.g. storage

As well as the complexity of the space itself, one significant barrier to designing mobile user experiences is that as practitioners we rarely get to explore the entire space for one project. However, the variety of projects we do get involved with does provide some over arching clarity of the best methods for research, design and evaluation.

2.2 Client Relationships

Another significant barrier to designing quality mobile user experiences is the relationship between clients and practitioners. A client's location, short time frames, tight budgets, need for secrecy and lack of UCD knowledge can all have negative influences. If clients are based in Asia with a large part of their market in Europe then methods need refining to accommodate their location. The multicultural needs for scaling, communication within the design team etc also need to be considered. Decision makers are often not the team members that we see making it difficult to influence design decisions. Bound by non disclosure agreements practitioners are also severely limited in our use of case studies, which in turn restricts knowledge sharing. Clients also do not always understand or want the most appropriate method. For example, clients may ask for focus groups so they can see 16 people in one day, when in-situ observation of three people in one day would provide much better data. Improving the client relationship with quality results is often the only way to ensure the best methods are used.

2.3 Usability

In our research, one thing is apparent in almost all projects; basic usability is often overlooked in the design of mobile devices, content for these devices and the supporting websites and collateral that accompany the devices. While products such as the iPhone cash in on intuitive interaction because the actions are familiar, content is often poorly designed, and guidelines are

ignored. Design focus is also often on functionality rather than usability. Consumers seem to be willing to overlook usability issues because of the functionality. But basics of noise interference, lighting and glare issues, poor use of screen real estate, the ergonomics of handsets, web content providers using absolute values, connection speed and not designing specific mobile sites all seem to be overlooked.

As well as usability, access for all seems to be almost completely ignored. Mobile devices are difficult to use in a variety of different contexts and these factors are often over looked. As mobile devices are used in more and more varied locations the manufacturers and content developers need to consider access. For example in cold climates using cocktail sausages for touch interfaces rather than taking hands out of gloves, glare on screens, operating the systems in noisy environments. Many handset manufacturers seem to still be missing the point, designing separate handsets different demographics. Designing Fisher Price style phones for older adults is not respectful or tasteful. These oversights offer a huge space for improvement and gaining market share if clients are willing to spend the time and money.

3. RESEARCH METHODS

User-centred design follows an iterative pattern of research, design, evaluation and release as illustrated in Figure 1. Clients require us to become involved in research at various phases of the design life cycle for different projects. While ideally we would be involved throughout the life cycle, as agency practitioners we are often brought in for one phase or another rather than end to end. Here we present a variety of research methods that we have used and experiences we have had when conducting user research.

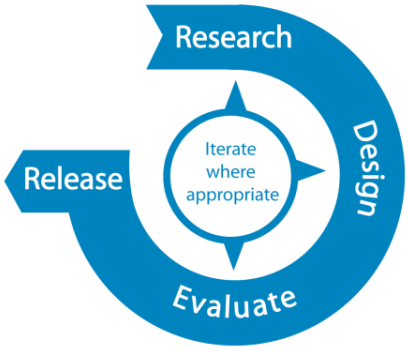


Figure 1: The User-centred Design Process

3.1 Formative Research

Formative research is necessary to gain insight into the needs and desires of the target market and to gain greater understanding of the context in which products will be used [1]. While formative research is highly valued in UCD for determining user requirements and setting release criteria, in practice it is rarer than we would hope as clients often mistakenly believe they already have sufficient insight to design their products. However, over the past two years when we have been involved in formative research we have used a variety of methods. These are mainly 'in the wild' methods with real consumers to gain insights into their behavior and context of use. For formative research we would definitely encourage 'in the wild' research. However, time and client needs often mean that this is not possible.

In one study for a mobile phone manufacturer exploring music consumption behavior on the move, **observation** was a key method used. The practitioner and one of the client design team observed participants in various settings including record stores, commuting on public transport, hanging out at home and university. The observation involved **shadowing** and a follow up **interview** after the session. The observation was augmented with participants completing **cultural probe** [2] type activities such as photographing significant moments influenced by music. The primary focus of this research was on contextual and behavioral aspects rather than fine detail of the interaction with mobile devices making these ethnographic methods ideal.

During another study conducted to better understand blind mobile phone users' needs, we used several other ethnographic techniques. We used an electronic **diary study** which blind consumers found difficult to complete, largely due to the time commitment involved which is a common complaint with diary studies. Participants mentioned that they would have preferred to use a Dictaphone to record their thoughts and activities.

Another tool used was a form of **experience sampling method** ESM [3]. At various points over a two week period participants were sent a text message asking them to perform a simple task using internet on their mobile phone, for example find a book on Amazon, and return a text message with details of how they got on. This technique was very successful with participants finding it much easier to respond immediately via text message than having to remember to note activities in a diary later.

Over the years we have also conducted numerous **one-to-one interviews** about mobile use, either in participants' homes, neutral locations such as cafés, or in the System Concepts' labs. One-to-one interviews 'in the wild' are particularly useful as they illustrate contextual issues around mobile use. For example, we would not see the difficulties a blind mobile phone user has when Talks software [4] is used in a noisy environment such as a café at a train station if the interview was in the lab. In this environment Talks users hold the handset up to ear to listen then move the phone down to press keys. In a quiet lab setting users wouldn't need to put the phone to their ear as the phone would be audible.

In addition, the banter and level of connection with someone in their own home is more relaxed than in a lab setting making it easier to discover facts more personal facts. In Figure 2 we spent several hours learning more about this man, his mobile and the environment he uses it .

Figure 2: One of our participants in his home showing us how he would use his mobile phone using Talks software.

Despite the major benefits of 'in the wild' research, interviews in the lab make it much easier to set up recording equipment

(discussed further in Section 3.3). Another problem with 'in the wild' research is that travelling to observe people takes longer and involves more transport costs. It is also necessary to have one person facilitate the discussions and another one to record the sessions increasing the resources and possible intrusion.

Relying on self report techniques such as diary studies can be problematic in general, but even more so for mobile research as the behaviors of interest in are often undertaken on the move when pen and paper are not handy. Therefore it is necessary to find ways of making the recording of events simpler for the participants.

What is clear from the formative research we have conducted over the years is that the location and method used greatly depends on the objectives of the research. If clients want the data recorded in great detail for later viewing then 'in the wild' research is more problematic. However, if the high level findings of the context and more general behaviors are more important, then 'in the wild' observation provides a richer picture.

3.2 Design

Once consumer insight and user requirements have been gained from the formative research, conceptual design follows. While designing the product is the focus of this phase, testing conceptual designs with potential consumers and comparing them to design guidelines can help ensure the success of a product. Clients can involve us in this phase as independent researchers to assess other people's designs or as consultants helping guide designers using insights gained in previous research, best practice knowledge and applying guidelines.

During one successful design consultancy project, we worked alongside a large online content producer who was adapting their online offering to mobile specific sites for a variety of handsets. They required **consultation** regarding usability and accessibility of the sites on various handsets. We conducted **expert reviews** of preliminary designs using **guidelines** [5; 6] and heuristics [7]. Once designs were coded they were evaluated and changes made. The designers were willing to learn as much as possible and we facilitated this through awareness training and allowing the designers to shadow us during the expert reviews. During this research we used a variety of handsets to test the different designs, but it was not possible to consider all variables in Section 2.1.

In many situations it is clear that mobile web content producers and practitioners are unaware of the guidelines that are available and the restrictions of mobile design. During a recent, UK Usability Professionals Association (UK UPA) event relating to mobile design, it was discovered that few practitioners used the W3C **guidelines** [6] and none had used the RNIB guidelines [5] when consulting. This is in part because guidelines are too specific and do not consider the interaction of the different factors listed in Section 2.1. In addition, few clients encourage use of guidelines preferring to look for innovation rather than solid design patterns. However, if clients can be convinced to involve practitioners who are aware of the guidelines and who can advise about the appropriate methods then better designs can result.

In another recent instance, a mobile manufacturer designing a new mobile phone content browser, wanted to explore how to present photos and video content. Following the technique of **Rapid Iterative Testing and Evaluation (RITE)** [8] for early design concept testing, we used low fidelity paper prototypes in the lab. We alternated between a day of testing and a day of workshops with the client to iterate the designs. In the final round of research we used prototypes of the visual design that weren't interactive to assess the branding and emotion. However, it was clear that participants were happier to criticize roughly sketched designs than to what appeared to be higher fidelity prototypes. They were often distracted by the detail or the specific content.

Encouraging clients to use RITE is a massive victory for practitioners and one that includes consumers early in the design process rather than just the evaluation. It is also clear that paper prototypes are much easier to change than higher fidelity prototypes and participants are more willing to be criticize them.

3.3 Evaluation

Once the conceptual design has been firmly established and higher fidelity prototypes are available evaluation against release criteria is often required by clients. This type of evaluation is usually to confirm that there are no major problems prior to release. Unfortunately clients often only bring practitioners in at this point to say they have considered usability rather than actually considering the user throughout the design life cycle. This often means that poor design decisions cannot be undone.

During a comparative study of a new proposition operating system with the android and apple operating systems, we used basic **usability metrics** to evaluate the products. In this comparative evaluation brand loyalty was a control variable with a focus on the usability of the new proposition operating system. The research was done in the lab, because there were a variety of tasks to cover with multiple handsets and it would not have been feasible to conduct this research 'in the wild'. Participants did not use their own phones, the tasks were contrived and not all functionality was available due to the prototype. However, client viewing of the evaluation was vital and large numbers of participants were tested and these were better facilitated in the lab.

One evaluation method that we have found extremely useful is automatic logging of behavior. During the evaluation of a media player application, an application was installed on participants' phones to record their activity with the phone and with the media player. In comparison to other diary studies we have conducted it is clear that automated works better.

Figure 3: Recording camera attached to device and the output of the remote high-zoom camera

The recording equipment used for evaluations (and research in general) is another issue to consider. 'In the wild' it is important to capture the behavior as naturally as possible. In the lab it is often important for the client to have control to focus on aspects they see as important. We have three different camera set ups which we use in different situations detailed in Table 1. We are lucky to have a bespoke camera that attaches to the phone which is much better for recording the interaction with the device in a

natural way rather than some solutions which require the phone to be fixed. However, it does not record the facial reactions, comments and contextual issues. Figure 3 shows the attached camera and the output from a remote zoom camera.

Table 1: Ranking of different cameras for viewing and recording mobile device interaction.

Camera Type	Freedom of participant movement	Lack of intrusion for participant	Client viewing experience
Attached to device	1	3	2
Suspended on Tripod	3	2	1
Remote high-zoom	2	1	3

Recently a group of UCD practitioners gathered for a UPA event. They ranged from freelance consultants specializing in mobile through to in-house practitioners at mobile phone manufacturer companies. Many of those present were aware of the W3C guidelines [6] and had used them for evaluation. Several practitioners stated that the guidelines had been augmented to included alternative wording, additional points to consider etc. None of these amendments seem to be fed back to the W3C or being shared which reduces the usefulness of the guidelines. However, of the approximately 50 people present only a handful were aware of the RNIB **guidelines** [5] and none of them had had an opportunity to use them.

3.4 Release

UCD consideration should not stop once a product has been released. Most research that we are asked to do about mobile devices post release are about the purchasing process and the out-of-box experience. It is rare that we are asked to do longer term studies into the learning and adaptation that is likely to take place over time.

We are most often involved with point of sale research and out-of-the-box experience. We usually evaluate the out-of-the-box experience using **expert reviews** assessing the packing, wires, user guide and set up user journeys. We also do expert reviews of handsets and mobi websites. Observing the experience at high street stores is difficult due to recording issues. Assessment might also have an environmental impact focus (reducing packaging and documentation) or a purely usability focus (using heuristics). We have also used **focus groups** to explore issues that consumers had with phones they had been using for some time.

For blind and visually impaired consumers the purchase and post purchase situation is dire, with little information available about the relative merits of different handsets at high street stores. While we have not done any specific research on accessibility needs post purchase, the RNIB recently organized an event to help members to choose a handset which we attended in an effort to gain further insight to share with members of the UPA. For the participants it was vital to have real hands on experience with the devices.

4. DISCUSSION AND CONCLUSION

What is clear is that quality research in each of the phases of UCD helps gain a greater understanding of the some aspects of the problem space. In addition, if the research is of a high quality then the practitioner earns the respect of the client and the relationship is improved. Practitioners can then be more assertive about which methods are preferred and can educate the client about becoming involved in earlier research and including more user involvement.

Because of the complexity of the mobile device and the contexts in which the devices are used, different research methods are better in certain situations. While 'in the wild' research has many benefits, lab based research can also offer useful insights and improve the overall client relationship by allowing them to participate more actively.

Many practitioners have never used guidelines which raises the issue of their effectiveness. Content of the guidelines are often too specific making them difficult to use. Many practitioners also do not know about specific guidelines. Many designers also do not know how to apply them and clients often ignore them believing that the functionality will make up for any lack. The complexity of the space makes guidelines too simplistic. More research into how to present the guidelines better may help. In addition, a variety of methods are available to conduct research but some are better applied at different points in the design process. Practitioners need to help guide clients as to which are the best.

There are a couple of take home messages from this snapshot of practitioner life. Firstly, there is no single right method for research as each situation is unique and the mobile space is complex. Secondly, the client still needs convincing to do quality research throughout the design life cycle. Long term relationships between UCD practitioners and business focused clients will help ensure that the best methods are used every step of the way. Finally, there is still a need to focus on core usability and accessibility when designing products and improve the use of guidelines to ensure quality mobile user experience is designed.

5. REFERENCES

[1] International Standards Organisation (2009). ISO 9241.

[2] Gaver, W., Dunne, T., and Pacenti, E. (1999). Design: Cultural Probes. Vol 6 (1) pp: 21 – 29.

[3] Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. New Directions for Methodology of Social and Behavioral Science, 15, 41-56.

[4] Nuance Communications. (2010). Convenient audio access to mobile phones. http://www.nuance.com/talks/

[5] Chandler, E., Dixon, E., Pereira, L., Kokkinaki, A., Roe, P. (2005) COST219ter: An evaluation for mobile phones. In P. Bust (ed.) Contemporary Ergonomics (2006) (Taylor & Francis, London)

[6] Rabin, J. and McCathieNevile, C. (2008). Mobile Web Best Practices 1.0. http://www.w3.org/TR/mobile-bp/

[7] Nielsen, J. (1994). Heuristic evaluation. In JL Nielsen and R.L. Mack (eds). Usability Inspection Methods. John Wiley & Sons: New York, NY.

[8] Medlock, M.C., Wixon, D., McGee, M., & Welsh, D. (2005). The Rapid Iterative Test and Evaluation Method: Better Products in Less Time. In Bias, G., & Mayhew, D. (Eds.), Cost Justifying Usability (pp. 489-517). San Francisco: Morgan Kaufman.

Understanding touch screen mobile phone users by Taxonomy of Experience (ToE)

Wen-Chia Wang
Brunel University
School of Engineering and Design
Kingston Lane Uxbridge
Middlesex
UB8 3PH UK
mepgwcw@brunel.ac.uk

Mark Young
Brunel University
School of Engineering and Design
m.young@brunel.ac.uk

Steve Love
Brunel University
School of Information Systems, Computing and Mathematics
steve.love@brunel.ac.uk

Ian Coxon
University of Technology Sydney
School of Design, Architecture and Building, Australia
bluearth@optusnet.com.au

ABSTRACT
The importance of user experience in the product design process has been discussed in recent decades. This paper introduces the methodology of Taxonomy of Experience (ToE) and its data analysis process of SEEing to obtain user experience with an unfamiliar interface on a touch screen mobile phone. To execute ToE, SEEing includes nine sequential stages to help researchers to deeply understand users' needs by transforming the user's verbal commentary into super-ordinary metaphors. This study presents the process of applying ToE-SEEing to understand users' experiences of trialing a touch screen mobile phone. The results can be used to establish extra design principles for touch screen mobile phones.

INTRODUCTION
User experience is one of the most important elements of product design and has often been discussed in the HCI community in recent decades. The general definition of user experience is beyond the usefulness and usability of a product [1, 2], and it might be affected by the 'user's internal state, the context, and perceptions of the product' [19, p.1]. Research into user experience endeavors to achieve users' pragmatic and hedonic level. It is, however, subjective, highly situated and dynamic in nature [19]. Therefore, efficient methodologies to obtain and to evaluate user experience accurately are essential for improving product design. Norman [3] indicates that an understanding of user experience should be able to evaluate the user's experience in a circumstance that is similar to the actual using situation to avoid the user imagining the experience. The data collection process should record the user's experience when it happens rather than rely on recalling the memory about the experience. Moreover, the user's experience should be understood through the user's subjective information about the experience [18]. Methodologies for evaluating experience have been established based on the user's attitudes and expectations [18], emotion [9], concept of the object [17], judgment of the product [16], and through comparing the user's reference to different interfaces [6]. These studies capture and analyze user experience by experimental pilots [18], emotion sampling [9], multiple card sorting [17], and repertory grids [16]. In addition, the approaches of diary [4], focus groups, surveys and competitive analysis are tools that are commonly used [14]. Whilst many researchers attempt to transform user experience to quantitative data, Coxon's [7] Taxonomy of Experience (ToE) and its analytic approach of SEEing, uncover an understanding of the user experience through qualitative analysis. The term 'SEEing' attempts to differentiate from the processes of thinking, but still associates with the thinking processes [7]. The nine steps of the SEEing process aim to clarify the user's experience. It begins by transforming the user's verbal commentary and ends in a synthesis, super-ordinary metaphors. This study applied the ToE to capture the user experience of trialing an unfamiliar touch screen mobile phone. The outcome of the SEEing analysis provides an alternative consideration for the interface design of touch screen mobile phones.

TAXONOMY OF EXPERIENCE (ToE)
The purpose of ToE is to understand the user's experience with a product via analysis of their verbal commentary to find the deep meanings hidden from the verbal commentary. It combines empirical [5] and academic [10, 11] perspectives. Previous studies that relate theory and practical concepts of user experience [15] provide a good foundation to establish this methodology [7]. The structure of the ToE is based on philosophy, methodology and design theory; thus the ToE provides a multi-layered method to understand user experience. In developing the initial concept to understand the experience of transportation vehicles, Coxon realised the importance of 'understanding the experience of those people already involved in designing and using the vehicle' [7, Ch.3, p.2]. Therefore, [7] discussed positivism, idealism, realism [13], and applied phenomenology as a methodology to understand individual lived experience [7]. Whilst Coxon [7] reviewed the video that recorded his own trial experience of an electric vehicle, it helped to recall the deep aspects of the experience that he was not particularly conscious of while personally trialing

the vehicle. The process of transcribing the sequence of the dialogue and other activities once again brought the experience more vividly into consciousness. Figure 1 is the framework for understanding an experience from four aspects. The experiential event impacts the user's cognition (thinking and acting), engages sensorial aspects, and impacts one's emotions and feelings. In the meantime, these elements are all taking place within an existential framework of temporality, spatiality, relationality and corporeality [7].

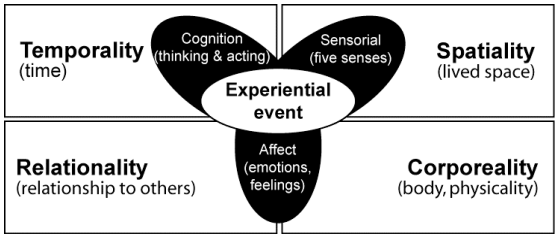


Figure 1. Framework of an experience

Coxon [7] thus develops the multi-layered analysis process of SEEing for understanding users deeply, which is also supported by [20]. According to [20], it suggests that when writing about 'lived experience descriptions', the descriptions should be able to a) describe the experience as it is lived without asking why; b) describe the experience from the inside, the feelings, mood, and emotions; c) focus on a particular example of the experience and to describe it; d) focus on an example that stands out, as it was the first time; e) aware how the body feels, smells, sounds and so forth; and f) avoid trying to beautify the illustration with flowery language or terms. Overall, the ToE-SEEing process brings an experience to a comprehensible and visible format rather than an abstract concept. It comprises nine steps, which can be seen as follow (details can be seen online [7]).

Step 1: Gathering data and establishing structures
It is important that the researcher 'gets to know' the experience, becoming familiar with the experience by understanding its 'language'. This stage emphasises that the researcher has to immerse themselves in the experience completely [8, 12]. The data of people's experience can be collected from observations, interviews, and contextual studies that are captured in creative ways. The information of images, sounds, samples or the other type of the information are collected that might be useful to the researcher to recall the user's experience and to write the descriptions for further analysis at later stage.

Step 2: Descriptive narratives
This is the process of transforming the data collected in *Step 1* into a textual format for analysis. This stage also breaks the experience into fragments as small as a single word, or a phrase, and refers them into the SEEing process.

Step 3: Sorting fragments into themes
This step includes meta-themes and sub-themes. Meta-themes in the SEEing process include somatic, affective, cognitive and contextual factors; the sub-themes include senses, positive-negative affect, internal-externalised cognition, and many contextual categories. Each theme has a collection of fragments, and provides the researcher with the feeling of the story that they are telling.

Step 4: Developing meaning(s)
This step requires the researcher to look at each fragment of the information carefully and to find other and deeper 'meanings' behind the fragment. This process helps to 'tease out' the text into different meanings. It is not yet the stage for the researcher to judge and to interpret what they think the meanings of the fragment 'should be' or 'could mean'. Instead, the aim of this process is to accept all 'possible' meanings that are contained within the fragment.

Step 5: Essential elements
This step helps to filter out the less important meanings. The researcher has to determine if the meanings in *Step 4* are incidental or vital to the essential nature of the experience. It is necessary to know the importance of the experience, whether if the element is essential to the experience, or the experience might be different without the element.

Step 6: Super-ordinary elements
This step distils the super-ordinary essence of the experience, i.e. the unexpected, novel and hidden aspects of the experience. Aside from the functional, form oriented, and everyday aspects; this stage isolates those elements of the experience that might not have been seen as an important part of the original design. However, those elements are still an important part of the experience. This process searches for the surprising elements, the unintended impacts of the experience.

Step 7: Weighting of super-ordinary elements
This is a weighting process to consider which super-ordinary elements are the more 'powerful' of the essential elements of the experience. The researcher evaluates the super-ordinary elements by his understanding of the language of the experience, to give a subjective numerical scale using a Likert rating (1-7, 1 is low) to determine a relative level of intensity.

Step 8: Super-ordinary summary words
The sorted super-ordinary elements in descending order provide a ranking of the essential super-ordinary elements of the experience by intensity. This stage uses word metaphors to synthesise *'what is the collective meaning behind these elements'*? For example, the super-ordinary element of 'no risk means no fun', could essentially be a statement about 'freedom to enjoy danger'.

Step 9: Summary word descriptions

The previous step summarised the super-ordinary elements; this stage focuses on 'explaining' the summary. It concludes the work of *Step 6-8*. One or two narrative paragraphs helps to represent the understanding of the experience to someone who does not understand the meaning of the super-ordinary words.

Design guidelines for the interface design for mobile phones have been well established by mobile phone manufacturers and include design principles for elements such as content, layout, colour, font size, text and terminology. However, it is still necessary to understand the users' requirements from the user experience aspect. As the market for touch screen mobile phones continues to grow, understanding experienced users' thoughts and novices' expectations of the touch screen mobile device is essential to providing a better design. This study applies the ToE and its analysis process of SEEing to generate deep understandings of users' experience in order to provide extra design principles for mobile phone interfaces.

EXPERIMENT DESIGN

The aim of this study attempts to understand the extra criteria of designing mobile phone interfaces from the users. This is the first trial of applying ToE-SEEing to mobile phone user experience, so this study tends to simplify the variables to make sure the collected data is clear and precise. Therefore, the participants were required to trial the touch screen mobile phone in the laboratory. Twelve participants were recruited from a British University. Half of them currently use a touch screen mobile phone, whereas the other half currently use a 12 keypad mobile phone. A Vodafone 541 mobile phone (Figure 2) is used for this study because participants from the other series of experiments would not have used this model beforehand. Operating this phone should be a whole new experience for most participants.

Figure 2. Vodafone 541

This model is the previous generation of touch screen mobile phone. The hardware and software are not advanced to compete with new generation phones, such as the iPhone. The aim of choosing this model was to push the participants to talk more about the using experience. Before starting the data collection, the observer demonstrated to the participant how to apply the approach of 'think aloud' by trialing a touch screen camera. Participants then were required to practice a 'think aloud' protocol by trialing the camera. The practice was intended to help the participants to get used to expressing their experience while trialing the Vodafone 541 mobile phone. Participants had five minutes to free trial the phone as they wished. Their interaction behaviour with the mobile phone was filmed for the ToE-SEEing analysis (the camera only focused on their hands and the mobile phone, and recorded their verbal commentary without showing their face).

RESULT

All collected data was transferred into *Step 3* of ToE-SEEing. As mentioned earlier, *Step 3* includes two layers of themes (meta-themes and sub-themes). Firstly, each participant's verbal description of the experience was coded into different themes. The meta-themes include the body-somatic experience (sensorial experiences, sound, touch-feel, sight, smell, taste, comfort-ergonomics, and appearance-aesthetics); the heart-affective experience (positive-negative emotions); the head-cognitive experience (conation-reflective thought-external-doing, conscious cognition-reflective thought-internal-thinking,); as well as a range of contextual factors (environmental, regulatory, social factors), and existential factors (time, space, corporeality body, and the relationship to others). Most of the participants' usage experiences with the touch screen Vodafone 541 mobile phone strongly relate to the sub-themes of sight and cognitive experiences. The following section presents the super-ordinary elements and the summary of participants' user experience with trialing the Vodafone 541.

Understanding-from the head

It is important to see that the 'graphic icon and its title are consistent, and represent the function clearly'. Clear feedback is given confirming whether or not the operation was successful.

It is essential to show instructions for unique features of the phone, maybe to demonstrate how to operate the feature, or to make it easy to get 'help' information.

Sensitivity of the touch screen is crucial, and should fit the user's pace when operating the phone.

The user would like to dominate, to trust the phone, and to fully understand the operation process before using the phone.

Experienced and familiar-from daily life and history

The way to operate the scroll bar on Vodafone 541 should be the same as using the scroll bar on a computer.

From previous experience of using a mobile phone with a 12 keypad, it would be good to see that the icon becomes highlighted when browsing the icon on the menu.

It will help to reduce mistakes if the phone can highlight what the mistake was, to detect the failed task automatically, and then provide help and instructions to complete the task correctly before the user has to ask for help.

Freedom-from the operation
The phone should provide links between different functions, rather than having to go to the menu to execute another function.

The size of the phone provides the freedom for the user to carry it all the time, and allowing the user to hold the phone in the hand easily without worrying that the phone might slip from the grasp.

The three super-ordinary elements above had the highest score from participants. The other super-ordinary elements were 'specific', 'share', 'intimacy', 'comfortable', 'enjoyment', 'flexible', and 'logic'.

The ToE-SEEing helps to transform and to categorise the raw meaning of an experience to find the meanings behind the user's commentary, to sort the importance of those elements, and to summarise super-ordinary elements of the experience. It provides an overview of the user's experience and describes whether it is the user's previous experience or the experience that was produced when trialing the object. The categories in *Step 3* help to clarify the key themes of users' experience, and to establish a good foundation for further analysis. In this case, the summarised super-ordinary elements not only reflect the user's expectation of Vodafone 541, but also highlight the components that the user cares about most. There is no doubt of simply following design guidelines to design a product, furthermore, this study suggests that it is helpful to include user experience as part of the design guidelines before executing 'design'.

CONCLUSION

This paper presents the process of executing the methodology of ToE-SEEing to understand user experience with a touch screen mobile phone. The validity of ToE has been examined with extensive observation data from video clips and interviews during the development process [7]. This method might be questioned due to its explicit subjectively; nevertheless, as mentioned earlier, the nature of an experiential encounter is subjective, situated, complex and dynamic. Therefore, the ToE-SEEing process is a useful tool for distilling the true meaning that lies behind the verbal description of such a complex event. This short paper emphasises the importance of understanding user experience before design begins. The result provides alternative considerations to achieve the goal of making things 'easy to use'.

REFERENCE

1. Alben, L. Quality of Experience: defining the criteria for effective interaction design. Interactions 3, 3 (1996), 11-15.
2. An evolving glossary of experience design. http://www.nathan.com/ed/glossary/
3. Anderson, R. Organizational limits to HCI: conversations with Don Norman and Janice Rohn. *Interactions* 7, 3 (2000), 36-60.
4. Bolger, N., Davis, A. and Rafaeli, E. Diary methods: Capturing life as it is lived. *Ann Rev Psychol* 54 (2003), 579–616.
5. Brezet, H.P.D., Vergraght, P.P.D. and Van der Horst, T. (eds), *Kathalys: Vision on sustainable product innovation*, BIS publishers, Amsterdam, 2001.
6. Comparing the user experience of search user interface Designs. http://www. cs.tut.fi/ihte/CHI08_workshop/papers.shtml
7. Coxon, I. Designing (researching) lived experience. Ph.D. Thesis. University of Western Sydney, 2007. http://www.bluearth.net/site/phd-dissertation
8. Csikszentmihalyi, M. Flow: the psychology of optimal experience, Harper Collins, New York, 1991.
9. Emotion sampling and the product development life Cycle. http://www.cs.tut.fi/ihte/CHI08_workshop/papers.shtml
10. Feenberg, A. from essentialism to constructivism: philosophy of technology at the crossroads. *Technology and the Good Life*, University of Chicago Press, Chicago, 2000.
11. Glanville, R. *Re-searching design and designing research*, School of Design, Hong Kong Politechnic University, 1999.
12. Hanington, B.M. Innovation and method in design research, *Design Plus Research Conference*, Carnegie Mellon University, 10, 2000.
13. Higgs, J. The context of qualitative research, in Higgs, J. (ed.), *Qualitative research: Discourse on methodologies*, Hampden, Five Dock, NSW, 1997.
14. Kuniavsky, M. Observing the user experience: a practitioner's guide to user research. Morgan Kaufmann, San Francisco, Calif. London, 2003.
15. Schmitt, B. Experiential marketing: how to get customers to sense, feel, think, act, and relate to your company and brands, The Free Press, New York, 1999.
16. The quantitative side of the Repertory Grid Technique: some concerns. http://www.cs.tut.fi/ihte/CHI08_workshop/papers.shtml
17. User experience: a multiple sorting method based on personal construct theory. http://www.cs.tut.fi/ihte/CHI08_workshop/papers.shtml
18. User experience evaluation with experimental pilots. http://www.cs.tut.fi/ihte/CHI08_workshop/papers.shtml
19. Väänänen-Vainio-Mattila, K., Roto, V. & Hassenzahl M. Towards practical user experience evaluation methods. *Proc. Workshop 2008,* GCOST (European Cooperation in Science and Technology 2008).
20. van Manen, M. Researching lived experience: human science for an action sensitive pedagogy, 2nd ed, Althouse Press, London, Ontario, 1997.

Into the Grey Zone: seniors, cell phones and milieus that matter

Kim Sawchuk
Communication Studies
Concordia University
Montreal, Quebec, Canada
01-514-848-2424
kim.sawchuk@sympatico.ca

Barbara Crow
Communications and Culture
York University
Toronto, Ontario, Canada
01-416-736-2100
bacrow@yorku.ca

ABSTRACT
This paper describes the project, 'Seniors and Cells,' a Canadian study involving over one-hundred and twenty users of mobile media technology who are aged sixty-five years and older. We recount the reasons for undertaking the research within a national context, the methodological choice to interview groups of senior users and non-users in their milieus, and the question of payment for time and other ethical and practical issues. We discuss how we came to make methodological adaptations and iterative decisions to better understand seniors' mobile usage within terms that make sense to them.

1. INTRODUCTION
Within the burgeoning literature on the everyday and innovative uses of cell phones and mobile technologies, there is a concentration of detailed statistical or ethnographic data on those who are young or middle-aged ([1]. [2], [3]). With the exception of a handful of articles ([4], [5]), much less attention, scholarly or otherwise, is paid to those who are fifty-five and over: this demographic constitutes a 'grey zone' literally and metaphorically [6]. Our research on 'Seniors and Cells' rectifies this absence and is intended to contribute, productively, to the discussion of the intertwining dimensions of age, technology, and the everyday practices of citizenship by differentiating between 'shades of grey': we highlight what they do, and try to make sense of it in their terms, rather than comparing seniors with more 'active' user-groups.

While we cannot claim, at this stage of the work, that we are in a 'truly mobile setting', our research has brought us into *milieus that matter* to our subjects: milieus are spaces of encounter and exchange, and not merely sites of data collecting and gathering. We discuss two sorts of milieu: intimate individual exchanges comprised of one-on-one conversation, and social interactions that break the isolation and loneliness often experienced by seniors. These milieus take shape within broader national contexts of telecommunications infrastructures and policies that influence and structure individual choices. We end with a discussion of some of the practical strategies we have adopted for engaging with these users from a perspective, which allows them to transform the research agenda [7]. In this paper we describe some of the broader lessons learned from our project to date, and reflect upon our research process and practice.

2. CANADIAN CONTEXT
Before examining our reasons for engaging in discussion groups in local settings, there are some data to consider when researching age in relation to wireless, mobile media in Canada's particular national milieu. Here we would like to clarify that we do not see our research in opposition to statistical research, but as complementary to statistical overviews. Large-scale quantitative studies provide a picture of how systemic conditions might impact the everyday practices of cell phone use in elderly populations.

2.1 Seniors and cell phones in Canada
Statistically, the number of seniors in Canada is predicted to double from 4.2 million at present to 9.8 million by 2038 [8]. This is a dramatic increase in the population, which has led to grave warnings in the Canadian media about the emergence of 'a grey tsunami' threatening to bleed dry the resources of the state, with rising costs for medical care, housing or other social services. There is a climate of fear around ageing reinforced by these discourses that our study's participants challenge overtly in their comments on encounters with ageism, and more subtly, through of the liveliness of their engagements with friends and family.

The mobile phone, or cell phone as it is called in Canada, is rapidly displacing the landline telephone for person-to-person communications here, as it is worldwide. Cell phones are also increasing in popularity in our country, which has tended to have lower adoption rates than elsewhere. On average, 72% of Canadians now own a cell phone, a steady increase since 1997. The lowest rates of ownership are in Quebec and amongst those fifty-five and over.

2.2 The Wireless Industry in Canada
As a recent report on the cell phone industry indicates [9], cell phone companies make enormous profits for services that many Canadians feel are overpriced and inadequate. Such media reporting on the industry has been backed by independent inquiries made by digital research institutes that confirm that Canadians are paying extremely high rates for their cell phones, compared to users in other countries, are often locked into draconian service contracts, and can experience punitive fees if they break these contracts [10]. The telecom regulation that historically guaranteed reasonable rates for landline phones in Canada has not been applied to cellular services. Landlines have been reliable and inexpensive, comparatively speaking.

These conditions influence seniors' responses to our study and to us; they must be accounted for in our analysis of the individual and household choices made on cell phone use. Hence, to quickly summarize the results of our preliminary data analysis, we have found that seniors tend to restrict their practices to a few functions, share the cell phone between spouses, use pay-as-you-go cards for monitoring minutes, and consider carefully who is given access to a phone number. These practices do not stem from mere 'fears' of entering into the brave new world of mobile technology. Instead, as we have seen, these 'restrictive' practices are logical choices given the infrastructural conditions in Canada. Understanding these systemic conditions that influence individual choice also makes us hesitant to use quasi-psychographic terms, based on survey research, to develop user profiles [5]. Such profiling does not account for individuals or social groups in the context of their milieus.

3. SENIORS' LOCAL MILIEUS

It is within this context that we are conducting our research. To date, we have held formal group discussions with over one hundred and twenty people who are sixty-five and over, accepting invitations into their community centres, legions, church halls, and homes. Over the past three years, we have engaged in a small number of early one-on-one interviews and countless informal conversations with retired individuals in shops, on the street and in cars on the subject of ageing and technological practices. We have received unsolicited emails from retired people, who have offered encouragement and their own testimonials upon hearing of our project. While technically, only the interviews have been approved by our University research ethics committee, all of these conversations constitute valuable source materials for understanding the digital desires and frustrations of senior users.

Our entry, albeit brief, into these local contexts provides crucial information about our subjects and their lives in relationship to the lifeworld that might not be shared in a survey or interview situation. Entry into these spaces gives us insight into the lives and mobile practices of both users *and* non-users. This latter group are particularly important to us. Just as we have been concerned with the reasons this population restricts use, our conversations with seniors indicates that the reasons for this 'non-use' are extremely complex, and need more attention. In this our project dovetails with the work of researchers such as Sally Wyatt [11], who see use and non-use as part of a longer continuum of practices.

4. ITERATIVE ADJUSTMENTS

In developing a multi-pronged research agenda, we have adopted methods of data collection that draw from our past experience in developing user-tests, guidelines and protocols for artists and engineers that are participatory and iterative in focus and in practice [12]. Participatory research design asks subjects to play a role in setting the terms of the research agenda. Iterative research design suggests a constant re-adjustment of the research strategies over time, as one learns 'in the field.'

While related to ethical issues in ethnographic research, including feminist empowerment research, these research strategies stress social change, and are based on ongoing dialogue and the researcher's accountability to participants at all stages and phases of the research plan. Unlike empowerment research, the demands we place on subjects to maintain involvement in our project is minimal. But then, we are not looking to institute change in a community, but to bring attention to those who have become invisible, but have other things to do. As word of our study has gone out into the communities we contact, we frequently find that we have more seniors wanting to talk with us than we have time to give.

From an ethical perspective, in a short research note on working in the field of gerontology, Alan Walker makes the crucial point that the 'older research subject' should ideally be an active participant in setting the research agenda for epistemological, ethical and political reasons [7]. This awareness and transformation of the research agenda is imperative when dealing with the elderly because of the rampant existence of age discrimination and social exclusion often experienced by this cohort. Given the lack of satisfactory studies with this cohort of users to act as a comparative benchmark, and our contextual approach, a pilot project with eight elderly intimates was critically important. During the pilot phase, we were able to test interview questions, develop a small survey, and most importantly negotiate our language and central 'concepts'. Given the lack of literature on seniors in media studies, we used these intimates to test initial hypothesis and intuitions, and to work out our own biases and presuppositions.

4.1 Local Help: from informants to mediators

Our local 'organizers' are seniors themselves and most often, our initial point of contact has been through family and friends. In 'Approaching the Elderly,' John Tulloch discusses openly the pitfalls and advantages of working with family members, which he sees as valid when working with populations that feel vulnerable [13]. We have also made contact with individuals in existing volunteer and local organizations for seniors: a guild of quilters; a resource centre; a swimming group; a community centre. After retirement, many seniors also get involved in volunteer work for others, and these organizations have been helpful for not only giving access, but as a reminder that active ageing is not only possible, but actual. Local helpers have also provided material comforts for our groups: coffee, juice and snacks. They play the critical role of mediators in the research process [14], and have not acted as informants in the classical anthropological sense.

4.2 The Old is Always Other

What constitutes a senior is a contested category (see [15]) and protocols for addressing this cohort are uncertain, given the range in ages from the recently retired 'young-old' to the 'old-old.' Ageing, we were reminded over and over again, was not only a demographic variable or a biological condition: it is also a question of perception, 'a state of mind.' What was interesting to hear was that no one sees themselves as old. The old is always other. Old is associated with a lack of agency and immobility. We learned that they felt more comfortable with the term 'senior' than with other identity categories associated with ageing.

4.3 Incorporating Critique

Our seniors were willing to engage in what are arguably insightful *critical* discussions of our research program, the current literature and presuppositions about age and ageing culturally. In some instances they have acted in a consultative role setting the research agenda, formulating initial questions and helping make contacts with others. In a more recent encounter, we have been

told that our ethics forms were too long and complicated and changed them in response. They have made suggestions to our survey, asked us to increase font size to make the text easier to read, and actively worked to set up interviews with their constituents. In discussion, they have corrected us when we have revealed our own ageist presuppositions: one early lesson for us was when we asked if they were 'still driving' which lead to a direct confrontation with our use of the word 'still,' which implied being incapacitated. Further, what this revealed is that the concept of mobility is about physical mobility, movement through space, as much as it is about a mobile device. The phone is not only a part of media ecology but a whole system for staying mobile and active: driving and public transport; exercise and walking; having and exerting agency. Conversation on other interests also made us initially attentive to emergent patterns, including pre-retirement and post-retirement work and the gravitation to particular digital devices.

4.4 Media Ecologies: phone alone?

One of the issues in studying any technology is that the focus of the study often isolates the one technology from other uses and practices. Many of our early participants wanted to talk not only about cell phones, and often diverted from the discussion of this technology to the question of other digital media. This pattern was repeated in all of our discussion groups, where we were reminded continually that technologies do not exist in the lives of individuals or households in isolation: that there was a 'media ecology' of multiple technologies for communications.

This led to the important finding that the 'restriction' or 'rejection' of the cell phone did not constitute a resistance to new forms of communications from these users. The cell phone exists as a choice among several options, including Skype, the landline phone, and email: choices made based on expedience and cost, the experience they had from pre-retirement occupations, and the demands of their interlocutors. This type of finding points to the need for long-term contact with participants to track the reasons individuals may adopt new practices, exchange technologies, or exit the cell phone scene altogether. In addition to tracking such changes over time, we have situated cell phone use in terms of income levels, for example, the pressure felt by seniors who balance a home budget and life on a fixed income to keep up with the costs of maintaining services, engaging in upgrades, or using multiple functions.

4.5 A space for non-users

Leaving space for the non-users involves accounting for those who would be left out of the conversation if we used more 'objective' means of gathering data on our subjects through a technological device (such as a tracking mechanism on their mobile phones). As we have argued elsewhere [16], one of the other biases our research addresses is the tendency in media studies to focus on the exuberant user of technology. In our study, the perspective of these non-users has become extremely critical, for it challenges the assumption that only active users or owners of a mobile device are affected by the transition to wireless, mobile means of communication. The increasing lack of public telephones is but one example of how non-users are affected by broader cultural shifts. It also means that instead of the research being about us 'getting information from them,' the discussion groups have also become spaces where non-users come to find out from users why they should or should not get a cell phone, what plans they might get, what options and features on the phone to look for, and tips about how it might best serve them.

4.6 We are not selling anything

This was particularly important when dealing with our seniors, as well as with cell phones: one of the important points of reassurance we had to offer was that we were not marketing researchers working for cell phone companies. We had to convince them that we were not trying to sell them something. This, we realized, is related to one of the main issues of this group: their distrust of telecommunications companies, and their sense of vulnerability as a population in relation to unscrupulous researchers and scam artists trying to get money from them.

4.7 From Focus Group to Discussion Groups

Initially we termed our research as 'focus group discussions'. During the course of our research, we have preferred to use the term 'group interviews' in order to stimulate a discussion amongst seniors, rather than simply read a list of questions. For this cohort, 'focus groups' imply that we are situated within the paradigms and parameters of marketing research, often affiliated with the much despised phone companies. But it also seemed as if we had a definite research agenda. 'Discussion group' reframed the terms of engagement as allowing for a much more meandering flow of conversation, guided not only by our questions, but by the participants' interests and needs, discussed within much less formal context.

4.8 Money talks

One critical aspect of recruitment, but also part of the politics of our project, that is rarely talked about is the issue of money. We have paid our subjects for their time: CDN$20 per hour, which usually translates into CDN$40 cash in-hand for each participant. We also know from experience that there are class divisions in the doing of research: doctors and lawyers are paid for their time in focus groups commensurate with their income and status. Survey research with 'ordinary' people frequently asks of time on a phone or an hour of two, but the pay scales are different. Paying our participants for their time was an ethical and political decision that benefitted us. Word got around about the compensation, which valorized their time as important and they loved this. Seniors are sometimes seen as people with 'time on their hands' and 'they like to talk' as if they have nothing better to do. Offering money for their time and talk was an affirmation that their insights were valuable.

But we also had insight into who they were because of what they told us they were going to do with the money. Given the significant socio-economic differences between our subjects, their responses to having their time acknowledged and their participation rewarded in tangible terms was telling. For one group, CDN$40 is a week's worth of groceries. For another group, this money represented a special lunch with a friend. For another group, our research was used as a fundraiser for their religious organization. For yet another group, this money was coveted as a way to purchase quilting materials for their favoured hobby.

5. CONCLUSION

Our foray into the 'grey zone' has revealed much to us, from methods of researching seniors to considering mobility and

mobile use in more complex and nuanced ways. While we have collected some qualitative data on our participants, our research largely relies on verbal testimony as well as our own observations of friends and relatives who are in this demographic and who have actively assisted us. In our understanding of what we are 'getting' in these conversations, we operate from within the perspective of repertoire analysis, defined as 'recurrently used systems of terms used for characterizing and evaluating actions, events and other phenomenon' [17]. As Joke Hermes explains, interpretive repertoires are 'a storehouse of possible understandings, legitimations, and evaluations that can be brought to bear on any number of subjects' [18]. We are in the midst of this analysis of our data, using the TAMs analyzer open source software program to systematically document emergent terms and themes from the volumes of data we have collected: the each group discussion is comprised of text ranging from forty to seventy pages in length.

Statistical portraits draw attention to larger discursive and social patterns, but this type of 'survey' can operate effectively in collaboration with qualitative data to offer insights and construct categories that are meaningful to the population studied. The technical collection of data from the devices themselves may also not be appropriate for this cohort. While there is much to be said for studies that track mobile users and do not require them to fill in details, instead relying on software programs (such as Mobitrak) and the phone itself to gather data, we are not sure if it would be either possible or desirable given the specificities of our group of participants. Considering that this cohort does not use or want many of the functions of the cell phone, imposing such a device will only give us access to some respondents. There are important cultural differences between this generation and younger users in their feelings about the need for privacy, as our discussions on these matters have indicated.

Our research suggests that the methodologies for mobile users and usage in the 'grey zone' are enriched when we engage with them in their milieus on their terms. As we enter into the next phase of the research, it is the insight gleaned through contact, conversation and entry into these *milieus that matter* that will guide the analysis, interpretations, and positions we inevitably must put forth as the authors of this study, seeking to make seniors matter within the ever-shifting terrain of mobility studies.

6. ACKNOWLEDGMENTS

Our thanks to the Social Sciences and Humanities Research Council of Canada for their support of this project (#410-2009-1553).

7. REFERENCES

[1] Ito, M. 2005. Mobile phones, Japanese youth and the replacement of social contact. In *Mobile Communications: Renegotiation of the Social Sphere*, R. Ling and P. Pedersen, Eds. Springer, London, 131-48.

[2] Caronia, L., and Caron, A. 2004. Constructing a specific culture: young people's use of the mobile phone as a social performance. *Convergence* 10, 2 (Jun. 2004), 28-61.

[3] Thulin, E., and Vilhelmson, B. 2007. Mobiles everywhere: youth, the mobile phone, and changes in everyday practice. *Young* 15, 3 (Aug. 2007), 235-253.

[4] Wong, C.Y., Thwaites, H., and Khong, C.W. 2008. 'Oh! My battery was drained because I forgot to press the end call button': In *Proceedings of the 21st International Symposium on Human Factors in Telecommunication: User Experience of ICTs* (Kuala Lumpur, Malaysia, March 17-20, 2008), 31-38.

[5] Lee, Y.S. 2008. Older adults' user experiences with mobile phones: user cluster identification. In *Proceedings of the 21st International Symposium on Human Factors in Telecommunication: User Experience of ICTs* (Kuala Lumpur, Malaysia, March 17-20, 2008), 39-47.

[6] Harris-Decima. 2008. *2008 Wireless Attitudes Study*. Technical Report. Canadian Wireless Telecommunications Association.

[7] Walker, A. 2007. Why involve older people in research? *Age Ageing* 36, 5 (Sep. 2007), 481-483.

[8] Statistics Canada. 2007. *A Portrait of Seniors in Canada*. Technical Report. Government of Canada.

[9] Nowak, P. 2010. Canadian wireless firms still tops in profit: report. *CBC News* (Jul. 19), http://www.cbc.ca/technology/story/2010/07/19/canada-wireless-profit.html?ref=rss

[10] Marlow, I. 2010. Cell phone contracts slammed. *Globe and Mail* (Aug. 10), http://www.theglobeandmail.com/report-on-business/cellphone-contracts-slammed/article1668239

[11] Wyatt, S. 2003. Non-users also matter: the construction of users and non-users of the internet. In *How Users Matter: The Co-construction of Users and Technology*, N. Oudshoorn and T. Pinch, Eds. MIT, Cambridge, 67-79.

[12] Crow, B., and Sawchuk, K. 2008. Shaking hands with the user: principles, protocols, and practices for user-integrated testing in mobile design. In *Mobile Nation*, M. Ladly and P. Beesley, Eds. Riverside Architectural Press, Toronto, 37-42.

[13] Tulloch, J. 1989. Approaching the audience: the elderly. In *Remote Control: Television, Audiences and Cultural Power*, E. Seiter, H. Borchers, G. Kreutzner and E.-M. Warth, Eds. Routledge, New York, 180-203.

[14] Latour, B. 2005. *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford University, Oxford.

[15] Riggs, K.E. 2002. The new, new deal. In Panel Discussion: Global Perspectives and Partnership on the Information and Communication Technology Divide. *Informing Science: InSITE – Where Parallels Intersect* (Cork, Ireland, June 19-21 2002).

[16] Crow, B. and Sawchuk, K. 2010. Embracing the restrictive user. *Cultures of Movement Conference* (Royal Roads University, Victoria, Canada, May 8-10, 2010).

[17] Potter, J., and Wetherell, M. 1987. *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. Sage, London.

[18] Hermes, J. 1996. *Reading Women's Magazines: An Analysis of Everyday Media Use*. Polity, London.

On Derive for Mobile Experience

Konrad Tollmar
KTH / Mobile Service Lab
Electrum 229, SE-164 40 Kista, Sweden
+46707337900
konrad@kth.se

Linus Harling
SU / DSV / K2Lab
Stockholms universitet, Forum 100, 164 40 Kista, Sweden
linus.harling@gmail.com

Robert Ramberg
DSV / K2Lab
Stockholms universitet, Forum 100, 164 40 Kista, Sweden
+46 8164914
robban@dsv.su.se

ABSTRACT
In this workshop note we describe a new approach we used for doing mobile user experience research. The method is build upon situationist theory and the Dérive method, and is used for inspire mobile storytelling. We outline some of the challenges we faced and the lessons learned based on these experiences and highlight what we see as the key areas to focus on, in terms of carrying out mobile user experience research on real use of mobile services in the future.

Keywords
Mobile user experince, Situationist theory, Storytelling

1. INTRODUCTION
This note will describe a new method for observing and understanding the mobile user experience. The method is build upon situationist theory and methods, the Dérive, and is used for inspire mobile storytelling.

The context for this work is an overarching research program - *Contemporaries* - where we study how writing of stories could become a part of everyday life and support participation in social, economic, political and economic life. The broad aim of *Contemporaries* is to facilitate the people's voices in what McLuhan would call the new global village. One the most important key here is: *Accessible to all*. Hence, we specifically focus on multimodal communication, using a variety of techniques and tools for the mediation of expressions. That is, how appropriate a tool and media is to present something, to illustrate and recreate expressions, its costs, reliability, and ease of use.

In this note we will describe a case where we have investigate how school children could use mobile phones in expressing themselves through storytelling and how to make this activity meaningful and valuable, but also ease of use, efficient and more accessible, i.e. the user experience of mobile storytelling. To observe and understand the user experience of mobile storytelling we have during the spring of 2010 run a serie of workshops with school children. Overall, +60 youths have been engaged in a total of 6 different workshops. We have also done a couple of pilots with university students (approax 30 participants), from the *University College of Film, Radio, Television and Theatre*, and young immigrants that mix Swedish language classes with work practice training, *Rekryteringsprogrammet* - city of Stockholm.

To facilitate the workshops we have used the *Dérive* (a walk and exploration of an environment without preconceptions, more bellow) method as an inspirational tool for initiate the storytelling. Starting with a dérive using a mobile blogging tools, the storytelling workshop continues using a web interface to the blogg tool on regular computer.

Understanding the overall user experience here is important [5]. Most often storytelling is used for document user experience, however in this case we have created a double loop where we also catch the user experience through the storytelling task.

In the rest of the paper will we talk a bit more about the methods we used. First will we elaborate more about how to use and motivate with the Dérive method, secondly will we discuss the outcomes and, last, sketch some ideas on how mobile storytelling could go hand in hand with observing and understanding the user experience.

2. ON DERIVE
A lot of our research is based on different methods that are anchored in action research where we as researchers rather than creating specific test groups prefer to enter into an already established structure and see how actual work is done. Our method is also influenced by, beside the traditional HCI methodology, an active community and city development.

However most of these methods often needed an injection to get started. Reason and Bradbury refer in their "Handbook of action research" [9] to the Scottish philosopher John Macmurray who note that " 'I do' rather than 'I think' " to initiate the learning process. Reason and Bradbury argue from this that the doing is the appropriate starting point for action research [ibid].

In this respect we have chosen to facilitate our workshops using the *Dérive* method. A dérive (drift) is an attempt at analysis of the totality of everyday life, through the passive movement through space [3]. This method has, for example, been used in studies of architecture by exploration of a built environment without preconceptions. Many situationist have also used derives for creating "psychogeographical maps". These maps are built from small snippets that form an understanding of bigger phenomena, something it described by Humber [10] as: "In discovering a small world we discover the whole world". In a similar way we also read our collection of stories as a part of a bigger and shared story that form a certain activity identity; it could be school work for some or looking for a job for others.

We use the dérive here to initiate a new moment within an ongoing activity. One example is a class that studied the industrialization period. In this case we talked about historical findings from the neighborhood and how these artifacts forms a shared collective memory that capture this place. The difficult part here is to explain the dérive method without being too specific.

We learned in the pilots that some participants get "lost" and need some more specific instructions. We choosed to providing some simple examples of how to do a non-planned movement through space, e.g. make up some non-deterministic rules such make turns on certain events. Furthermore, we briefly showed some examples of historical psychogeographical maps without going into the details.

The overall aim of the dérive was to inspire them to collect multimodal story snippets using their mobiles. We tried two different approaches, the first was to use Android phones and a specific app for mobile blogging, and the second was to ask them to use their own phones and standard tools, like SMS and MMS. We will return to a discussion about the trade-offs between these approaches later on in the paper. However we also learned that by sending a text messages, or take a picture on themselves, before the dérive has a positive impact on the experienced of the dérive as well as how much they used the mobile blogging tool.

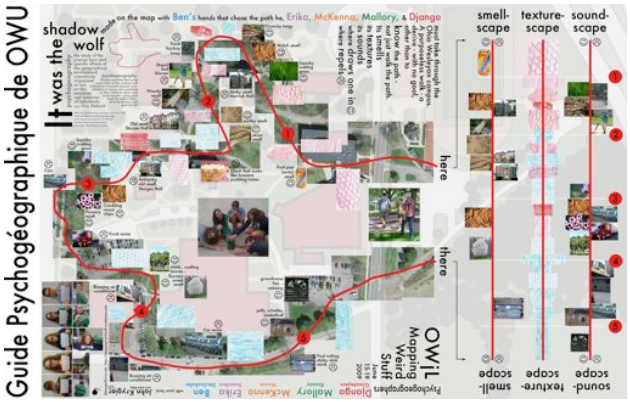


Figure 1. Example of psychogeographical map

Situationism has gained some recent popularity in the HCI community [6][8], mostly as inspirational tool to engage designers with modernist counter culture. Perhaps resembles our derive mostly with cultural probes [7]. This uptake from a wide variety of disciplines to understand and design is rather typical for HCI and its pros and cons are being discussed repeatedly. Some would argue that these methods are too often used without reflection and reference to their intent. We do agree and our solutions is to not use the derive to exclusively to understand the mobile experience but rather put a focus upon the stories that are generated and then use these sources to analyze and understand the mobile experience. In this was we also create something that we hope our participant's finds meaningful and real. This could also clearly been seen in the stories but the analysis is still work in progress and this note purpose is limited to reflect upon the method in this ongoing work.

3. STORYTELLING WORKSHOPS

As described the workshop was divided into two parts. The first part of the workshops, the dérive is described above. The second part of the workshop was the storytelling part, where the participants were introduced to the basics of WordPress and given access to personal blogs. This session allows us to have a flexible discussion about the dérive and the meaning of their stories, as well as the experience of the mobile tools and the use of multimodal media and expressions. This discussion, or conversation, consisted both of pre-prepared questions as well as free-form, open-ended conversation. The materials gathered during these workshops were of three primary types; voice-recordings from each session, the actual blogs produced and the answers to the survey questions.

The workshops conducted within *Contemporaries* included aprox 90 subjects, in the age of 14-25 with an equal gender balance and a mixed socio-cultural background.

Two different kinds of workshop formats were tested, the first with a longer dérive, followed by a later session in a computer room, editing and adding material to the blog. The other form restricted the users to only use SMS/MMS for input to the blog, and compressed both the dérive and compilation of material into a 3h workshop.

One substantial difference between these two types of workshops where what material the participants were allowed to make use of. The first type they could use of all photographs and videos collected, regardless if they had sent them or if they were stored on their mobile phone. In the other type they were limited to work with material sent and synchronized with the blog from their mobiles during the actual dérive. Additionally, in the second type, they did not have personal blogs, all posts were to a group-blog, with individual accounts.

First of all, we learned that introducing a tool like our mobile blogging tool for Android phones could fall short. We observed in the pilots that the mobile blogging tool often hindered the participants in their dérive with technical obstacles. The well-known SMS/MMS services took much less effort and enabled the participants to even use their own phones. Most important, this provided a much better result in terms of being able to create interesting stories. We also learned from the pilots that there is a need to provide a back-channel and better feedback through the SMS/MMS services that further engage the use. In this way we can partly mimic the online experience that otherwise is missing. Moreover, in our case we gave the participants vouchers to cover their cost but this worked less well due to the broad range of cell operators that our young group are using. In forthcoming studies we instead need to include some kind of premium SMS/MMS services.

However, at one point was the use of regular phones conceived less favorable. That is too follow and comment other work as well as search and connects the material with other resources. It seems that the integration between a mobile blogging tool and other online resources, e.g. social media and email, will become more critical as they services become more commonly used. Another observation is that some questions generated in the dérive faded quickly away, if they couldn't be concurrently explored.

Never-the-less, bottom-line here is that providing advanced handsets seems to work less well and most participants have sufficient advanced phones on their own. Some additional service could if needed also be hacked together on the server side.

Figure 2. Storytelling workshop

The shorter workshop format was a result of a perceived lack of connection between the gathering of data and manipulation of it. In the beginning of the project we thought that the storytelling work could improve if giving a chance to let the experience of the dérive sink in and reflect upon. This worked less well than expected. A lot of the ideas gained by the Dérive faded away rather quickly. The clear disconnection between these events made also the discussion about the mobile experience much less valuable. This follows our observation where it seems like it's a natural division between documenting and telling a story using a mobile device. Almost no one wrote text on their mobile device, they rather used other form of multimodal expressions and that tagged these with a few words that where later elaborated in front of a regular computer.

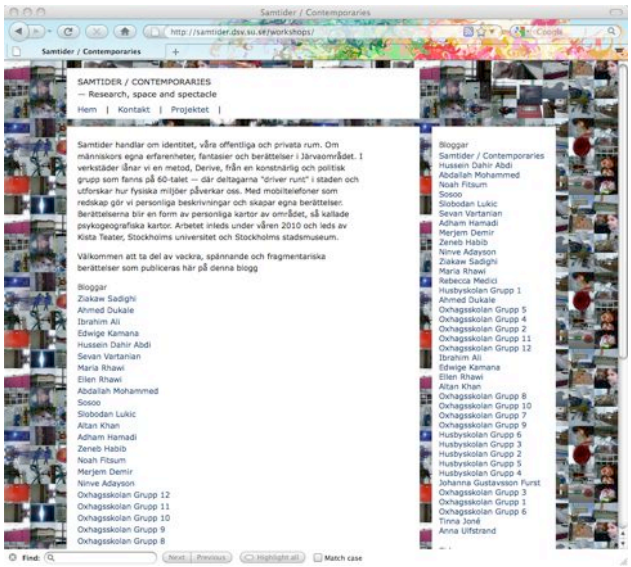
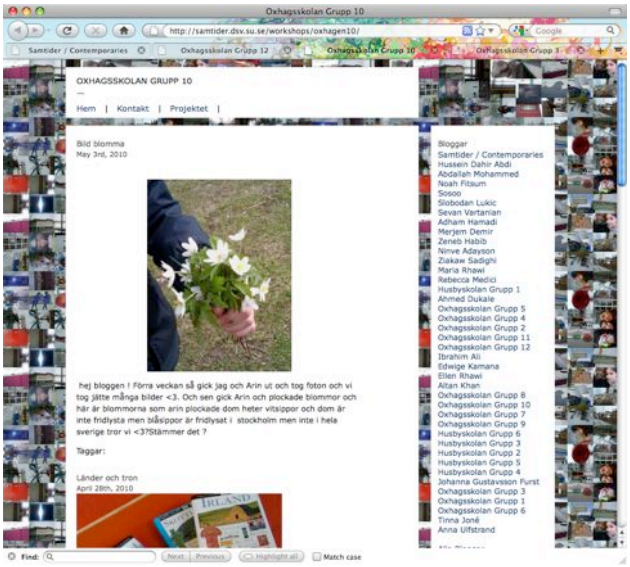


Figure 3. The *Contemporaries* web



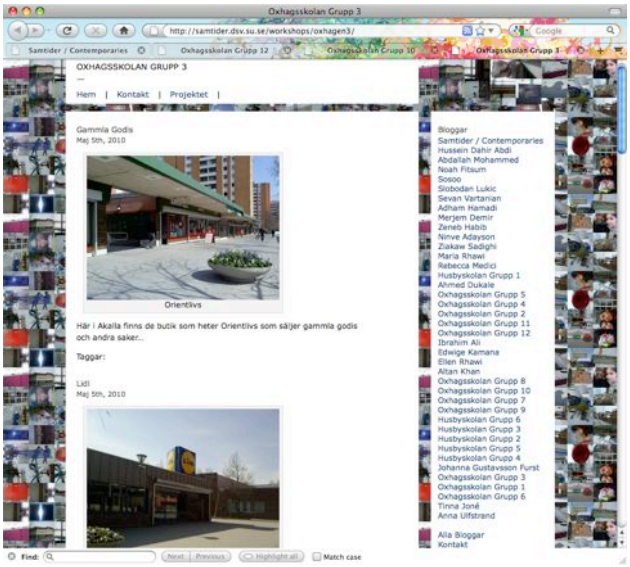


Figure 4. Examples from the *Contemporaries* web

The use of Wordpress as a blogging platform has worked well. We expected more problems here but almost all participants though it was very easy to work with Wordpress. This was unexpected for us and we had prepared templates that would simplify Wordpress authoring tools but these where not needed. However, based on input from the workshops, workshop participant experienced that they wanted to create and alter posts along a timeline, we also observed a lack of simple tools to create dynamic groups of users that more easily could follow each other blogs. This preliminary result points to the need for shared experiences, and an interplay between the group and individual development, and we are planning for forthcoming studies to build some new additional Wordpress tools along these lines.

4. CAPTURING THE MOBILE EXPERIENCE

Carrying out mobile user experience research is a difficult task. In order to observe and understand the mobile user experience we need to capture multiple aspects of what people do and feel about using mobile phones and services. Most often we develop mobile prototypes, deploy these applications on the personal handsets of real users and observe what happens. Shifting needs, contexts and the ubiquitous use of mobile phones makes it very difficult to observe naturalistic mobile behaviors and ask intelligent research questions about mobile user experience.

The dérive method overcomes some of these problems. First of all does the method allows a balance between flexible versus close instructions, and will hence constrain some aspects of contexts. Moreover do the method push for an open use of media in storytelling that will facilitate people's part of our social structure. Winograd and Flores among others argue that language is intrinsically tied to the situation. The context defines what the "words" mean as much as the "true" definition and composition of a sentence [11].

This lead us, secondly, to use storytelling generated by the dérive as a mean of analyze the user experience. There are a couple of different ways of measure the overall user experience. Most common is through various forms of self-reporting methods, e.g. diary methods [1] and Experience Sampling Method (ESM) [2]. In this case we have created a double twist and in the reasoning about the user experience we use a cyclical process, where understanding the user experience is a part of employment of the act the multimodal word stands for.

A derive is a situation-creating technique aiming at turning the city around. This "turning around" or détournment is a dialectical tool and in this context a method for argue and discuss the mobile experience.

5. REFERENCES

[1] Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. Annual Review of Psychology, 54, 579–616.

[2] Csikszentmihalyi M, Larson R. (1987). Validity and reliability of the Experience-Sampling Method. Journal of Nervous and Mental Diseases. Sep 1987;175(9):526-536.

[3] Debord G, (1981), Theory of the Dérive, Ed: Situationist International Anthology, Berkeley: Bureau of Public Secrets. http://www.bopsecrets.org/SI/2.derive.htm

[4] Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., and Landay, J. A. 2007. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In Proceedings of the 5th international Conference on Mobile Systems, Applications and Services (San Juan, Puerto Rico, June 11 - 13, 2007). MobiSys '07. ACM, New York, NY, 57-70.

[5] Forlizzi, J. and Battarbee, K. 2004. Understanding experience in interactive systems. In Proceedings of the 5th Conference on Designing interactive Systems: Processes, Practices, Methods, and Techniques (Cambridge, MA, USA, August 01 - 04, 2004). DIS '04. ACM, New York, NY, 261-268

[6] Gaver, W. W., and Dunne, A. Projected Realities. In Proc. CHI 1999. ACM Press (1999), 600-607.

[7] Gaver, B., Dunne, T., and Pacenti, E. 1999. Cultural probes. interactions 6, 1 (Jan. 1999), 21-29.

[8] Home, S. ed. What is Situationism? A Reader. AK Press, San Francisco, CA, USA, 1996.

[9] Reason, P. & Bradbury, H. (Eds.) (2001) Handbook of Action Research: Participative Inquiry and Practice, Sage: Thousand Oaks, CA, 512p

[10] Sadler, Simon. The Situationist City, Cambridge: MIT Press, 1998, pg 15

[11] Winograd T. and Flores F, Eds. 1985 *Understanding Computers and Cognition*. Ablex Publishing Corp.

-

Involving the user in the design of a mobile task-oriented travel guide interface

Lieve Laporte
IBBT-CUO, K.U.Leuven
Parkstraat 45
3000 Leuven, Belgium
lieve.laporte@soc.kuleuven.be

Bieke Zaman
IBBT-CUO, K.U.Leuven
Parkstraat 45
3000 Leuven, Belgium
bieke.zaman@soc.kuleuven.be

ABSTRACT
The Talos research project focuses on task-aware supply of rich content in mobile Location Based Services (LBS) environments. As a case study, a mobile travel guide application on iPhone is under development. This application has a task-oriented user interface, making it especially adapted to users' limited attention capacity and the great variety of tasks they have to perform on their trips. This paper focuses on (1) the characteristics of a task-oriented interface making it especially suited for use in mobile contexts, and (2) some specific methodological design issues including the involvement of users in the process of designing a mobile task-oriented user interface.

1. INTRODUCTION
The Talos research project focuses on task-aware supply of rich content in mobile Location Based Services (LBS) environments. In one of the case studies developed in the context of this project, a task-oriented user interface for an iPhone travel guide application had to be designed. The next section describes the concept of a task-oriented interface, and explains why it is especially suited for use in mobile contexts. Afterwards, an overview of some important parts of the design process is given, and some methodological issues and critical questions regarding the design of a mobile task-oriented user interface are put forward.

2. TASK-ORIENTED USER INTERFACES
Over the last years, mobile devices have become increasingly popular as a means to perform activities on the road. One of the most promising fields for the appliance of mobile applications in constantly changing environments is tourism. Substantial research effort has been devoted to mobile tourist services using Location Based Systems, though with different focus. A number of studies have focused on providing users with context-sensitive facilities or multi-modal interaction [[10]].

Others have explored continuous and data rich messaging across the mobile and desktop platform, augmented by complementary web services [[6],[4]].

Most of the prototypes that have been developed based on these studies, and most mobile interfaces in general for that matter, are organized from a domain viewpoint. Menu systems are structured hierarchically, linking the names of a category to the services in that category. If a user wants to catch the last train, for example, such a domain-oriented menu will guide him as follows: menu > latest information > traffic > train information > timetable > input start station name [[8]]. As this example shows, users have to follow the menu provided, and translate "what they want to do" to "name of the menu" before getting the mobile services they want. In other words, they have to learn the menu system to access the mobile services. This learning curve is one of the main disadvantages of a domain-oriented approach while handling a mobile device in a constantly changing environment, as it puts a lot of pressure on the user's cognitive skills. A user's ability to devote attention to several things at the same time is limited. Consequently, the use of computers impedes the user's attention resources with respect to other tasks and objects.

Especially in a mobile environment, the attention competition is complex and important, as the stimulations from other objects, events and tasks often require the user's attention. According to Kahneman's classical capacity model of attention [[3]], people will focus their attention first to objects they are interested in and familiar with, and to objects that demand less attention capacity. In mobile environments, users are more familiar with and interested in their daily tasks rather than in computers. Consequently, the tasks provided in a mobile guide system should be highly adapted and linked to the user's daily activities, using as few attention resources as possible [[10]].

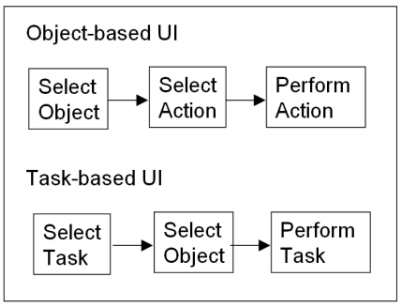
Furthermore, research on tourist behaviour [[1]] has shown that going on a trip encompasses a lot of potential activities and typical tasks, which do not always have a well defined goal. Instead, a great part of the actual enjoyment of being a tourist is in solving the problems they encounter on their trip (e.g. reading a map). Tourists are generally very flexible in planning and adjusting their activities, taking advantage of the changing environment (e.g. the weather). A seamless integration with the user's environment therefore increases the utility of a mobile device or service.

Because of the above arguments, namely the user's limited attention capacity and the enormous variety and flexibility of user tasks during a trip, more emphasis needs to be put on the integration of tasks within the context of the user's activity. Tourist systems need to provide task-based functionality and, equally important, task-oriented visualization.

A promising way of achieving this kind of mobile user experience, is the concept of a task-based interface (as opposed to a domain-oriented interface, as described above – Figure 1 shows the difference between both). Many definitions of task-based interfaces (TBI's) can be found [[9], [10], [11]]. We conceptualize a TBI simply as an interface where navigation is

based on tasks, allowing the user to concentrate on what he wants to do, instead of how to do it [[5]].

Figure 1. Domain (or Object) -oriented versus Task-oriented user interfaces.



For instance, imagine a tourist in Paris wanting to visit the Louvre museum. In an object-based city guide he has to look for a section on culture or museums and find the item for the Louvre to find its location and opening hours (the object). Then he has to look elsewhere for a section on public transport or a map to find out how to get there (the action). In a TBI, the tourist can simply select "visiting" (the task), followed by a selection of the object to visit: a museum, the Louvre. The user interface (UI) then offers him all the information he may need to perform the task (a description, a map with the route from his current location to the Louvre, the metro line) [[5]].

The advantage of TBIs over object-based UIs is that the first are less cognitively demanding because they do not require users to translate "what they want to do" to "name of the menu" before realizing a task [[8]]. Reducing the mobile system's learning curve to a minimum is especially pertinent for tourist situations which are typically complex and unpredictable.

3. DESIGNING A TASK-ORIENTED INTERFACE

Hardly any literature is available on designing task-based interfaces. For the design of the task-oriented user interface for the iPhone travel guide, we therefore made use of some well-known user-centered design (UCD) methods, but we adapted these slightly in order to focus on the task-oriented interface design.

3.1 Method

The design of the task-oriented user interface for the iPhone travel guide was the subject of a typical UCD design cycle.

In a first phase, we conducted a user and task analysis. Explicit focus, however, was put on the tasks part. Our goal here was two-fold: we wanted to get a view of (1) users' typical activities (or tasks) when travelling, and (2) the way they make decisions, the information they need to make these decisions, and the way they want to retrieve this information.

Users going on a vacation or a city trip were asked to keep a diary of their travelling activities. The diaries were structured, in the sense that they were explicitly designed to inquire after users' activities, the flow of these activities, and the way they came to certain decisions. For instance, one diary exercise contained an axis on which users had to put a range of activities, in the order that they were performed on their trip (e.g. looking for transport – search for opening hours – leave – read about a point of interest etc.)

Users were asked to send their diary to our department after returning from their trip. Each diary was then analyzed, and remarkable issues relating to their activities/decisions were marked. Afterwards, each user was invited to take part in an interview, using the travel diary as a steppingstone for further elaboration. Information from the travel diary and from the interview was put together, and a list of all tasks and task flows was extracted.

In a second phase, initial designs, based on the results of the task analysis, were made and, meanwhile, iteratively tested with end users.

In the final phase (which has not been conducted yet), a functional application prototype will be tested. These user tests will be three-fold:

(1) One group of users will receive the digital travel guide on an iPhone, together with a fixed test scenario, in which they are asked to do certain activities or tasks on the iPhone. This test will be conducted in a laboratory setting.

(2) Another group of users will also be provided with the travel guide on an iPhone, but, in contrast to the first group, they will not have to follow a fixed test scenario. Instead, they will use the application 'spontaneously' on a city trip, making use of a variant of the "Experience clip" technique as described in [[2]]. In this technique, a pair or a group of users is asked to test a mobile device. One of the users uses the mobile device/application to be tested, and the other user holds a camera to record the first user's actions and comments. This way, people are free to decide themselves what they want to record.

(3) A third group of users will participate in an experiment in which the task-based (mobile) interface will be compared to a domain-oriented (mobile) interface, in order to determine whether the task-based interface is more useful in mobile settings. The exact set-up of this experiment still has to be determined.

3.2 Design and Methodological issues

During the different phases of the UCD cycles, we encountered some specific issues and methodological problems, all of which are related to the fact that users had to be observed'on the road' and to the fact that such explicit focus was (had to be) put on tasks and task flows.

3.2.1 During the observation phase

(1) Users' tasks were collected using structured diaries. Our intention here was to have them use this diary as a kind of notebook, allowing them to note an activity or task on the moment they were doing it. However, although users were explicitly instructed to take the diary with them everywhere they went, most of them remarked that they filled out the diary at night in their hotel room instead of using it as a notebook. Diaries for studying mobile users should be made as concise as possible and other formats (besides questionnaires, paper exercises and the like) could be thought of.

(2) When designing a task-oriented interface, special attention is paid to the collection of user tasks. Compared to object-oriented or domain-oriented interfaces, the exhaustiveness and relevance of these tasks are much more important. Moreover, the task flow or

task sequence presented on the mobile devices screen should resemble the 'real life' task flow, as performed or experienced by the user on his trip. It should provide this same kind of 'naturalness' in its interaction with the user.

Doing a 'standard' task analysis did not seem to be sufficient to (a) obtain the necessary degree of task detail, (b) be able to decide whether the list of tasks in a particular domain (e.g. traveling) is exhaustive, and (c) deciding which tasks are relevant enough to uptake in the task-oriented interface.

(3) Other issues, related to observation methods in mobile contexts, come to mind here: the information used to derive the list of tasks, and the task sequences, is entirely based on the quality of people's memory. Due to the practical problems related to doing evaluations and observations in mobile environments (users on vacation can hardly be disturbed by researchers), 'real' and 'complete' observations are not possible.

These problems oblige us to conclude that our derived task model is an ad hoc one. Indeed, it is based on a particular set of users, and, hence, a particular set of tasks. Of course, it is almost impossible to list every possible task within the context of a UCD process.

Some research efforts have been done to formalize and/or automate the construction of task models [[1],[7],[8]]. In these studies, task ontologies are proposed, which are capable of supporting complex task definitions. However, although these models are often called user-oriented, the basic information needed to build these models or ontologies is often provided from an engineering viewpoint, and not via thorough user research.

3.2.2 During the evaluation phase

During the evaluation phase, similar problems occur: it is practically impossible to observe users when they are on vacation. We therefore want to combine a laboratory setting, in which fixed test scenario's will be executed, with a real life setting, in which users can make use of the application the way they want to. This combination will hopefully allow us to collect 'standardized' and uniform data (from the laboratory setting) on the one hand, and spontaneous user experience data (from the Experience clips in the real life setting) on the other hand.

Besides the typical user tests, we also want to make an explicit comparison between domain-oriented and task-oriented interfaces, and the use of both in mobile settings. An experiment will be set up to test the differences between both conditions.

4. DISCUSSION AND FUTURE WORK

The Talos project has shown that designing task-oriented interfaces for mobile applications poses serious challenges to include users in the design process by means of user observations and user testing. These difficulties are mainly related to the dependence upon the task analysis as input for the task-oriented interface design. Nevertheless, the whole concept of task-oriented interfaces is based on the precondition that users should not be disturbed in their normal activity flow, which challenges the 'traditional' task analysis methodology. One approach that seems promising to tackle this issue concerns the 'Experience clip" technique, as it allows tasks to be seamlessly integrated into users' activity while on the road.

Based on the experiences from the project, it is possible to list a number of issues that could be the subject of further research.

First of all, interesting ways of testing the usability differences between task-oriented and domain-oriented interfaces could be thought of. It would be a challenge to think of ecologically valid experiments in which the differences between these could be shown in a real life mobile setting.

Further research can also explore the interaction possibilities between abstract task ontologies on the one hand, and involving the user in the concrete interpretation of such an ontology for a particular domain (e.g. traveling). This kind of research probably is a difficult exercise in seeking a balance between formalization and automation on the one hand, and the much-needed "ad hoc" user input on the other hand.

5. ACKNOWLEDGMENTS

Our thanks to the partners in the TALOS project (FP7 Research for SMEs work program of the European Commission).

6. REFERENCES

[1] Brown, B. & Chalmers, M. (2003) Tourism and mobile technology.*Proc. of the Eight European Conference on Computer-supported Cooperative Work*, 2003, Helsinki, Finland.

[2] Isomursu, M., Kuutti, K. & Väinämö, S. (2004). Experience clip: Method for user participation and evaluation of mobile concepts. *Proc. Participatory Design Conference*, 2004, Toronto, Canada.

[3] Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

[4] Kenteris, M., Gavals, D. and Economou, D. An innovative mobile electronic tourist guide application. *Personal Ubiquitous Computing*, 13, 2009, 103-118.

[5] Laporte, L., Eyckerman, P. & Zaman, B. (2009). Designing a mobile task based UI for tourists. : ACM http://doi.acm.org/10.1145/1613858.1613946.

[6] Milic-Frayling, N., Hicks, M., Jones, R. and Costello, J. On the design and evaluation of web augmented mobile applications. *Proc. MobileHCI*, 2007.

[7] Naganuma, T and Kurakake, S. Task Knowledge Based Retrieval for Services Relevant to Mobile User's Activity. In The Semantic Web – *ISWC* 2005. Springer, 2005, 959-973.

[8] Sasajima, M., Kitamura, Y., Naganuma, T., Kurakake, S., Mizoguchi, R. Task Ontology-Based Framework for Modeling Users' Activities for Mobile Service Navigation. *ESWC Proceedings*, 2006.

[9] Thurrott P. How it works: Inductive user interfaces. Retrieved June 10, 2009: *http://www.winsupersite.com/showcase/longhorn_task-based_ui.asp*

[10] Yue, W., Wei-Ning, Y., Heng, W., & Shi-Hai, D. (2005). Multi-modal interaction in mobile computing. *Journal of Software*, 2005-01.

[11] Zimmerman G. Open User Interface Standards – Towards Coherent, Task-Oriented And Scalable User Interfaces In Home Environments. *Proc. IET IE07*, 2007.

From Workshops to Walkshops: Evaluating Mobile Location-based Applications in Realistic Settings

Matthias Korn
Aarhus University
mkorn@cs.au.dk

Pär-Ola Zander
Aarhus University
poz@cs.au.dk

ABSTRACT
Many open questions on how to best observe the mobile user experience remain – at the stage of design time as well as use time. In this paper, we are focusing on the stage of design time and describe our experiences from evaluating a mobile application for citizen involvement in municipal land use planning. Due to the problems and issues identified after conducting several user workshops in our exemplary case process, we propose "walkshops" as a complement to traditional workshops and prototype field studies specifically to evaluate mobile location-based applications (and similar context-aware systems). We report some problems with workshops and outline how a walkshop may be carried out. The first trials of the new method are promising and have generated valuable feedback, insights and discussions about using the mobile application within the intended contexts.

INTRODUCTION
How to evaluate the mobile user experience both at design time and use time poses many open questions. Specifically, conducting user evaluation with mobile location-based applications is difficult as most evaluation methods are not contextual and/or not suited for systems used in outdoor contexts. With this paper, we focus on a new technique for design-time evaluation of mobile location-based applications. Our purpose is twofold: 1) to illustrate situations where workshops, well suited for stationary computing, raise problems in a mobile context and 2) to show how this can be in part alleviated by, what we coined as "walkshops", given the right staging.

Methods for evaluating systems directly in the context of use exist. For example in prototype field studies the software is deployed and the use of the system over time somehow monitored or observed from a distance. They can be strong in their ecological validity, but in themselves they provide no access to how users think about the use.

Workshops address what field studies lack. The concept of 'workshop' as an evaluation activity has become an umbrella concept for a range of method prescriptions and activities involving groups of users who meet, where perhaps the participatory design workshop is the most well known type. Under the label of 'workshop' we find a number of evaluation activities that vary in how they are conducted, what they evaluate, and perhaps also their epistemological underpinnings. Workshops are, however, generally used in order to stimulate a discussion between users where the outcome is used in the next step of design. In the rest of the paper, we let the term refer to methods we have used throughout the project including future workshops, pluralistic walkthroughs and group discussions between users and designers facilitated by various design artifacts.

There may be differences between stationary use in a workshop and stationary use in practice in the field study. However, these differences are more severe in a mobile context, since mobile computing usually affords multi-tasking, and the physical conditions vary widely. Let us turn to walking as a methodological alternative that decreases these differences. Different walking approaches, where users would move about in the context of the application domain testing a system to be evaluated, have been used before, but a focus on walking as a stimulating activity has never been made explicit or analyzed systematically in any methodology to the best of our knowledge. For example, transect walks [4,5], a method from participatory rural appraisal (PAR), are used for understanding the local context (e.g. natural resources, landscape, land use etc.) by walking together with local informants through an area of interest (e.g. a rural village). In civil engineering and architecture, one researcher even spent an entire year walking the streets of Lisbon and Barcelona in order to understand the architecture of these places [8]. Ochoa highlights that "the physical walk allows the mental walk, stimulating the thought and making possible the contact of the body, as element of measure, with the space" [8]. Yet, both of these methods are aimed at understanding the environment and not the mediating technology.

Summing up, field studies do not provide the strength of workshops – to capture details in a user's sense-making and other cognitive processes. Workshops around a table do so, but sacrifice context. Walkshops enable the study of context paired with the micro-processes of sense-making. We apply walking (i.e. as in going for a walk) both as a tool for thinking and a tool for closer relation to the use context.

The forthcoming sections of the paper concretize this argument by examples from our research project. It describes how we developed that walking may stimulate reflection and that an increase of ecological validity can be gained by observing *sense*-making processes during walkshops. Finally it provides some lessons to be learnt.

RESEARCH CONTEXT

In this section, we describe the research setting where we employed our evaluations. This may give readers an idea of to what degree our findings generalize to their own evaluation tasks.

The evaluations have taken place within a project called "MobileDemocracy." This project has explored how citizens can participate in municipal planning in various ways. The approach is user-centered, and was conducted in participation with a municipality and some community-based organizations in western Denmark. A municipal plan is a document used in strategic planning that describes various visions and goals, but a key element is to relate the strategy spatially and to the existing physical infrastructure. The use of maps is frequent.

The municipality we interacted with has had problems in mobilizing its citizens and cooperated with us in order to better understand citizen involvement. The community-based organizations cooperated with us in order to make their voices heard to the municipality.

Our initial design idea was to create a mobile application that allowed citizens to suggest changes or to react to proposed changes, where these contributions could again be utilized in the planning process. The application was envisioned to be location-aware, and provides notifications when a user passed by a site of discussion. The municipality in our case is sparsely populated, so the number of discussions was estimated not to be occurring often enough to make most users turn off the notification. This mobile app was to be paired with a desktop interface, where people could engage more deeply in discussion. In sum, get people motivated (be it curiosity or indignation) through a mobile application, and provide room for deeper reflection at the desktop. The rest of this paper concerns the evaluations of the mobile prototype.

THE DESIGN PROCESS AND ITS EVALUATIONS

In this section, we highlight how we continuously evaluated our ideas in the design process, in order to arrive at an identification of some problems in the following section.

We explored these ideas in a user-centered system prototyping. The process was iterative, and we created a number of scenarios, storyboards, paper prototypes and refined a mobile prototype in a number of versions. The initial functions proposed in early versions of these design artifacts were based on a mix of our own ideas and empirical data from interviews and cultural probing not further described in this paper.

All of our design artifacts were put in contact with citizens and planners in workshops. Typically, we presented a scenario or storyboard, and discussed it with the participants. We moderated the discussions in order to get more concrete details or examples of actually ongoing planning situations, for the variety, and for barriers to appropriation of such systems. In one occasion, we assigned different roles to citizens, and asked them to discuss a fictive dilemma, and how dilemmas like this could be facilitated by IT. We did not show interaction on keystroke (or "tap stroke") level in these workshops.

We also conducted two pluralistic walkthroughs each with one user and one or two researchers in the panel [3]. The first was conducted on paper, where interactivity was emulated through Wizard of Oz [7]. The second walkthrough used an early version of the high-fi prototype on a mobile phone. The participants were given some tasks, where a problem a user could possibly relate to was introduced. They were then asked to solve or react on it through the prototypes. Following the steps proposed by Bias [3], each set of screens (on paper or the mobile phone) was looked at and notes written down individually. Afterwards they were discussed within the panel with the user going first. Some tasks given were aimed for use of potentially all functions of the prototype, and others were for a specific control inside the application.

Participants

There are two user groups in this project: citizens and planners from the municipality. The citizens were selected through "organizational belonging", and were therefore to some degree convenience sampling. However, we established contact with several organizations independently from each other, in order to avoid e.g. that the municipality chose citizens that would have opinions that fitted to their focus. The users from the municipality were chosen because they worked with municipal planning as key persons at various levels.

The citizens participating in our experience workshops were spread along common demographical spectra (age, gender, education, profession, computer literacy). The users in the most recent workshop to evaluate our latest high-fi prototype were chosen so that they fitted our final choice of a target user group – i.e. citizens from the more rural areas of western Denmark with medium computer, or rather mobile phone literacy.

METHODOLOGICAL ISSUES IDENTIFIED

On a general level, results from the workshops strengthened our design concept in making us sure about the motives in the activities users engaged in and in particularly interesting ideas for scenarios. Thanks to the Wizard of Oz nature of the first pluralistic walkthroughs, we got feedback of the interface at a stage, where we did not have to engage in time-consuming programming in order to implement it.

When we planned and later interpreted the results from the workshops, we experienced a number of problems with our method:

- We experienced a relatively formal workshop or meeting room atmosphere. No matter how we structured them it was mostly a bi-polar exchange between researchers and users.

- We could not utilize exploration by foot or vehicle in a natural manner, due to the physical scale of a meeting situation being too small. A user who was prompted by a position-dependent function had to be told "now you walked through the parking lot of your workplace" and then we made the mobile phone beep.

- Time constraints and stress on the user (e.g. for input with the onscreen keyboard) was observed to be totally different when sitting at a table (e.g. in a workshop where they posed no problems) or standing outside or even walking.

- We observed that in practice, typical workshop situations often proceed in a rather fixed setting, where one or two users continue to work on one phone. Although hypothetically possible, people do just not switch places that often.

- A meeting space is limited in the number of objects to interact with in ways that may be problematic. For instance, parts of tasks in our workshop included taking a picture. This resulted in arbitrary shots of e.g. the table instead of a suitable real-world photo. This includes e.g. problems of where to stand when taking the picture, or how the user would reason when the quality of the picture was poor. Similar issues arose when entering other types of content.

Our conclusion was that we wanted more realistic user conditions. At the same time, we wanted to keep the possibility to gain insights on the user's sense-making processes, which ruled out field studies with remote monitoring.

WALKSHOPS: EVALUATING MOBILE LOCATION-BASED APPLICATIONS IN REALISTIC SETTINGS

In this section, we further motivate and outline our proposed walkshop method, which we think alleviates some of the problems identified above. We also present experiences and results gathered from three walkshops we conducted with different user groups within the MobileDemocracy project.

The term "walkshop" itself has been used before – mainly by activist groups and in academia on topics like walkable cities as well as architecture and urbanism [10,9,6]. We adopt this term as it highlights the need to move part of the traditional workshops out of the meeting rooms and into the actual context of use. We stress both the in situ aspect and the aspect of walking as a thinking tool. The aim of this method is to evaluate prototypes in a more realistic or natural setting (i.e. within the context of use). Thus, the focus is on understanding the mediating technology, rather than the environment or context it is used in. With this, we strive to bring the evaluation into the context, rather than bringing the user's context into the evaluation situation.

Walking as a Thinking Tool

Neurologists have recently shown that walking as a rhythmic activity may possibly have a positive effect on our thought processes [2]. Similarly from the field of regional planning, Anderson proposes a method called "talking whilst walking", which suggests "that conversations held whilst walking through a place have the potential to generate a collage of collaborative knowledge" [1, p. 254]. While focusing on how an understanding of the knowledge and lives of individuals can be gained by wandering around aimlessly through place, he also again acknowledges that "the bodily movement of walking invokes a 'rhythmic relaxation' of both body and mind that 'frees the imagination'" [1, p. 258] as well as that "the rhythm of walking generates a rhythm of thinking" [Solnit in 1, p. 258].

We can thus argue that walking goes well together with talking and discussing the issues that surround us, and those we may be occupied with at that moment. We are aware of casual walks in the park with colleagues, friends or family, which occasionally lead to interesting and profound conversations. Back to our context, the activity of walking or wandering frees workshop participants from the fixed confines of the meeting room, table, and chairs making the atmosphere much more informal by allowing participants to move about freely and flexible.

Conducting Walkshops and Results

Our focus for the proposed method is on evaluating mobile location-based systems as their use cases are based on acquiring one or more spatial positions. Bringing these systems into the context allows for the creation of more realistic evaluation settings closer to the actual application domain (in terms of body movement, light conditions, distortion, etc.). Location and other environment variables can be incorporated more easily than in a spatially fixed setting.

Throughout the course of the MobileDemocracy project, we conducted three walkshops at different stages of the prototype and with different user groups. All walkshops took place outdoors. The first two walkshops were an integral part of workshops. One walkshop was conducted with planners from the municipality (three users), where the prototype only notified the user of topics at the locations he or she was currently walking and allowed him or her to retrieve details of these topics and see them placed on a map. The second walkshop was conducted with citizen users (four users), where we, in addition to the functionality above, allowed and asked participants to also create new topics with details, take photos related to these topics as well as view an augmented reality visualization of the future plan. The third walkshop was part of a preliminary project presentation again with planners and other interested parties from the municipality (six users). Here, we showcased in a hands-on (and foots-on) session the main functionalities and look-and-feel of our prototype implementation via scenarios and let the users react through the prototypes.

While one could imagine conducting walkshops as stand-alone, we deliberately chose to do them in conjunction with

user workshops in order to be able to work on different aspects of the project. In a three-hour session we reserved a timeslot of 45 minutes for a walk of approximately one kilometer. Before going out, the walkshop was introduced with a very short briefing of the prototype and followed up afterwards with a discussion. Here, created content (in our case topics and photos) could serve as a starting point and framing of the discussion. For the walk itself, we prepared real world points-of-interest along the route, of which our prototype would notify users and would allow them to view details and write comments. Users were also provided with more concrete problem-centered tasks and asked to respond to or rather interact in response to them. As it was our desire to understand the sense-making with such technology when used in context, we, as researchers, came along the walk. Our roles were, similar to those in workshop settings, those of facilitators (in terms of setting up the infrastructure and helping with usability issues), of observers (in terms of action research), and those of partners for informal conversations (in terms of soliciting, probing and discussing feedback and insights).

Through these walkshops we found several usability problems we hadn't identified before. These related especially to data input under stress (e.g. when standing or walking rather than sitting at a table), but also to ways of how and to what extent our system will and can actually be used in these (more realistic) settings (including what kind of content was created). Similarly, we experienced elaborate discussions and reflections of the users on how the system works, how it might be used, and which other opportunities it opens for the future. This may be in part due to the users interacting with the real environment rather than a staged one only provided through scenarios or similar. We believe that the real environment provided more graspable stimuli, which helped to fuel the users' imagination and thoughts leading to interesting discussions.

On a practical level, the walkshops allowed interacting with real-world objects and issues to create content from or take photos of. Furthermore, the walkshops afforded a flexible reconfiguration of usage situations between users. While also possible in workshop settings, with users already being on their feet and mobile, they simply moved around more and were free to engage with different other users, with the researchers or just explore the prototype on their own.

Apart from these findings, we are also of the opinion that going out into the context rather than bringing the context in is often the only meaningful way to evaluate a location-based mobile system with users. As our aim was to get an understanding of the sense-making process of users using the system, we decided not to put the system out into a field study and monitor it from a distance at this stage.

CONCLUSION

Based on our experiences with the workshops it seems that some things are problematic: Formality, stress constraints, exploration, and shortage of objects to interact with. It suggests that if such issues may be important for a user's experience, it is inadequate to rely too much on workshops for evaluation. Walkshops seem to mitigate some of these problems by intertwining the evaluation with the actual context of use. In conclusion, we observed users being more engaged with the software and the evaluation situation as a whole, but limitations e.g. on the use of paper prototypes persist. Therefore, walkshops are no silver bullet and we propose to integrate them into traditional workshops and complement them with other methods such as prototype field studies in later stages of a project.

REFERENCES

1. Anderson, J. Talking whilst walking: a geographical archaeology of knowledge. *Area 36*, 3 (2004), 254–261.
2. Babu, H., Ramirez-Rodriguez, G., Fabel, K., Bischofberger, J. and Kempermann, G. Synaptic network activity induces neuronal differentiation of adult hippocampal precursor cells through BDNF signaling. *Frontiers in Neurogenesis 3*, 49 (2009), 1-11.
3. Bias, R.G. The pluralistic usability walkthrough: coordinated empathies. In *Usability Inspection Methods*, J. Nielsen and R.L. Mack, Eds., John Wiley & Sons, New York, NY (1994), 63-76.
4. Chambers, R. The Origins and Practice of Participatory Rural Appraisal. *World Development 22*, 7 (1994), 953-969.
5. Dearden, A. and Rizvi, H. Participatory IT Design and Participatory Development: A Comparative Review. In *Proc. PDC 2008*, ACM Press (2008), 81-91.
6. Do Projects. *How to bring a Systems/Layers walkshop to your town*. (2010) http://doprojects.org/news/how-to-bring-a-systemslayers-walkshop-to-your-town (accessed August 25, 2010).
7. Kelley, J.F. An empirical methodology for writing user-friendly natural language computer applications. In *Proc. CHI 1983*, ACM Press (1983), 193-196.
8. Ochoa, R. The Importance of the Walk in the Analysis of Public Space. In *Understanding the Post-Industrial City: Metropolis, Urban Renewal, Public Space*, Joint PhD Seminar, Lisbon, Portugal (2009).
9. Osservatorio Nomade. *Walkshop: Aqueduto das Águas Livres.* (2009) http://www.osservatorionomade.net/lisboa/ (accessed August 25, 2010).
10. Walk21. *Walkshops at the Walk 21-VI 'Everyday Walking Culture' International Conference, Zurich, Summer 2005.* (2005) http://web.archive.org/web/20051223130401/http://www.walk21.ch/deutsch/programme/walkshops.htm (accessed August 25, 2010).

Observing the Context of Use of a Media Player on Mobile Phones using Embedded and Virtual Sensors

Jakob Eg Larsen Michael Kai Petersen Rasmus Handler Nima Zandi
Technical University of Denmark, DTU Informatics, Cognitive Systems Section
Richard Petersens Plads, Bld. 321, DK-2800 Kgs. Lyngby, Denmark.
{jel|mkp}@imm.dtu.dk

ABSTRACT
In this paper, we discuss how contextual data acquired from multiple embedded mobile phone sensors can provide insights into the mobile user experience. We report from two field studies where contextual information were obtained from N=21 mobile phone users in a 2–8 week duration, to derive information about participant context. In the second study our focus was on observing mobile interaction with a media player application over time and we discuss how the captured contextual data can lead to a better understanding of the context in which mobile applications and devices are used. We argue that this information can provide valuable insights to the design of mobile applications and user interfaces.

INTRODUCTION
Mobile phones have become ubiquitous and an integrated part of our everyday life. In the last couple of years smart phones have gotten increased attention with the availability of several new platforms enabling easy distribution of mobile applications. Present smart phones typically have a number of embedded sensors, which have been combined and utilized in interesting ways to create novel mobile applications. In particular the sensors enable location and context-aware mobile applications that are increasingly aware of the situation the user is in.

Designing and evaluating mobile applications introduce additional challenges compared to traditional desktop and web application development and evaluation. There is a set of design constraints due to limited size of the device, limited display size and resolution, and limited input-output capabilities compared to traditional computer form factors. Additionally the situations of use are inherently mobile, which makes testing actual use more difficult. Evaluations carried out in a laboratory setting might not be sufficient for applications where the use of the application is highly dependent on the context of use. However, obtaining data when carrying out "in-the-wild" studies of actual mobile application use can turn out to be difficult and resource demanding. This calls for methods and techniques to acquire information about actual mobile use in context to obtain a better understanding of the mobile user experience.

NordiCHI 2010 October 16 - 20, 2010 Reykjavik, Iceland.

In this paper we describe a software framework, which enable acquiring contextual data from the mobile phone embedded sensors during daily life use by a mobile phone user. The software runs silently in the background and is logging activities including data acquired from multiple embedded sensors, to describe and understand information about people and places, as well as application and media usage. Our focus here is on studying mobile phone use, which involves using the media player on the mobile device for music playback. The emphasis is on understanding the mobile user experience in the particular context in which it takes place. We hypothesize that contextual information obtained from a mobile device can offer useful information in terms of understanding the situations of mobile use involving the media player application. Furthermore, we suggest that such information can offer valuable insights for designers of mobile applications, where user interfaces for music recommendation is the present focus.

RELATED WORK
Kjeldskov et al. [7] discussed laboratory versus field evaluation of mobile applications, and discussed the issue of how much value field evaluation would add over laboratory evaluation. In their study of a specific mobile context-aware application it was found that not much was added by the field experiments and in [6] it was suggested that similar usability problems could be identified in a laboratory setting if the right use context is recreated there. On contrary, in a usability evaluation of a mobile application Duh et al. [2] found that significant more (and more severe) usability problems were identified in field experiments compared to laboratory experiments.

Bernard et al. [1] studied how users' performance changed under different contextual conditions, including varying the motion, lighting and task types. They found that the contextual changes had a strong impact on behavior and performance. Froehlich et al. [3] combined quantitative and qualitative methods for in-the-field collection of data about usage, including device logging of user context and environmental sensor readings, and in-the-field subjective user experience sampling (prompting for feedback on the mobile phone). Several field studies were carried out in order to study different mobile phone usage patterns.

Several studies of actual mobile phone use have been carried out, both in laboratory and "in-the-wild" settings. A recent example of a study of user experience evaluation of mobile TV was carried out by Obrist et al. [9] where the importance of studies within mobile context was emphasized in order to support the mobile user experience. Another example is Roto and Olasvirta [12] that studied mobile users on the move using web-browsers on mobile phones. The experiments were performed in a controlled manner by employing multiple cameras worn by the test participants to observe the use of the mobile application in a real-world environment. A moderator had to stay in proximity of the test participant to monitor the experiment. Interesting results were acquired in the study, such as the observation of shorter attention span when using mobile applications on the move compared to a laboratory setting [11]. Although the test participants are testing the applications in a real-world setting the setup is still artificial and does not necessarily reveal how the mobile user experience would be in a natural setting, which argues for a stronger emphasis on field experiments in HCI [10].

Generally such studies underline that although experiments in a laboratory setting might identify for instance user interface issues, it does not account for "in-the-wild" study of actual use in everyday life in context. The experimental setup might suffer from the fact that test participants are typically instructed to use an application they are not familiar with. This means that it might be the "learning to use" of an application, rather than "actual use" that is being studied. Such studies typically only capture use over a short period of time and only reveal little about actual use or use patterns over an extended time period where learning and habituation has taken place.

MOBILE CONTEXT TOOLBOX

In order to observe users while using mobile devices and applications in real-world settings we have created a context logging software framework for mobile phones. The Mobile Context Toolbox (MCT) framework for Symbian S60 mobile phones [8] aims to facilitate the process of developing context-aware applications as well as carry out "in-the-wild" experiments where acquiring data from multiple embedded mobile phone sensors is required in order to establish information about the context of use. The generic framework can obtain information from embedded sensors including accelerometer, GPS, Bluetooth, WLAN, microphone, call logs, calendar, and additional sensors can be added to the framework for specific experiments. The architecture of the framework is shown in Fig. 1.

EXPERIMENTS

We have carried out several experiments involving test participants carrying a mobile phone with our Mobile Context Toolbox software installed and report our findings from two of those experiments.

In our first experiment 14 participants were provided with a Nokia N95 mobile phone with our software framework installed. They were instructed to carry and use the mobile phone as they would normally use their own mobile phone. The software would silently and continuously acquire data from the embedded mobile phone sensors for the duration of the experiment. In addition to collecting mobile sensor data the framework can prompt the user for textual input on the mobile device, similar to the approach reported by Froehlich et al. [3]. Further details on this study is available in [8], and detailed analysis of the datasets acquired are provided in [4] and [5], but not discussed further here as that is beyond the scope of this paper.

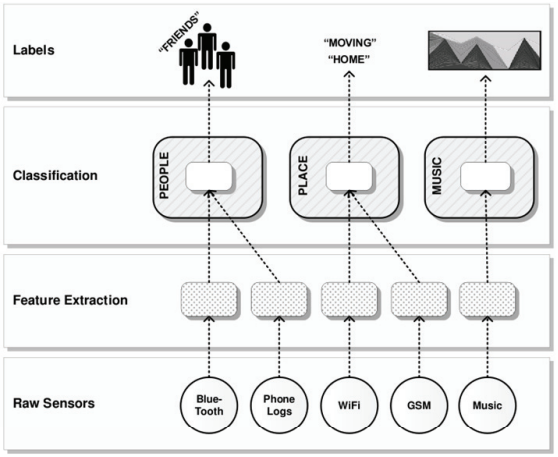


Fig. 1. Mobile Context Toolbox System Architecture

We extended the toolbox with a virtual sensor component capable of acquiring data from the embedded media player application on the mobile device. The component was capable of obtaining data including whether a music track was being played, the duration of the song, and current playback position. In addition we were able to acquire metadata from the particular song, including the artist and title. This was used in the second experiment where we focused on obtaining contextual information about people, places, and music [13]. As shown in Fig. 1, information was obtained from the Bluetooth sensor and phone log in order to extract features describing people related to the mobile phone user. Features describing places were extracted from information acquired from Wi-Fi and GSM cellular network information. Each of these features were translated into meaningful labels. From an application development perspective the intention is that applications built on top of the framework can utilize the contextual information inferred from the underlying system by means of the contextual labels acquired.

The experiment was carried out similar to the first experiment described above. This experiment involved 7 participants that were in a similar way provided with a Nokia N95 mobile phone with our software installed. They were instructed to use the mobile phone as their own on a daily ba-

sis for a two week duration. In addition they were told to use the mobile phone as their MP3 player device. The participants were also encouraged to upload their own music collection to the mobile phone, so that they could listen to the music that they liked and they would typically listen to on a daily basis.

ANALYSIS AND FINDINGS

An overview of the data acquired during the two week duration for the 7 participants in the experiment is shown in Table 1. As can be seen from the table the participants were fairly active in terms of using the media player on the mobile phone for music playback. In between 94 and 292 songs were listened to, which corresponds to 7—21 songs listened to on a daily basis on average. Also interesting to see is how many unique music tracks were listened to, indicating that some participants listened to a smaller set of tracks repeatedly. Each track played was logged with a time-stamp meaning that we could analyze the time of day where the media player was being used.

Participant	Tracks listened to	Unique tracks
1	160	85
2	153	100
3	190	48
4	292	68
5	110	58
6	167	124
7	94	65

Table 1. Overview of music listening for the 7 participants

Fig. 2 shows when music tracks were listened to by the 7 participants over two 3.5 hour durations on a random day in the experiments.

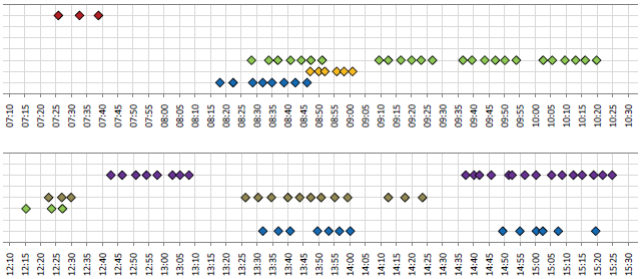


Fig. 2. Music listening patterns shown for the 7 participants on a random day of the experiments in two 3.5 hour periods. A dot corresponds to a music track played by the participant.

This information was coupled with the analysis of the contextual labels acquired from the logs of embedded sensor data. The GSM cellular information and Wi-Fi access points were analyzed in order to determine locations. Based on the analysis it was possible to determine the places in which the participants spend the most time. Thus it was possible to determine if a participant was at home, in a known place (a place where time was spent repeatedly), an unknown place, or in a transition between places (continuous changes in GSM and Wi-Fi data in minute size time windows).

The social relations were mapped based on the data acquired from correspondence logs (phone calls and SMS messages) in terms of who was calling and sending messages to whom. Based on Bluetooth device discoveries (of mobile phones) it was possible to map out when the participants were in physical proximity of each other. Furthermore it was possible to discover patterns in terms of the participants being in proximity of other people repeatedly. The inter-relations of the participants based on the mapping of the Bluetooth data is shown in Fig. 3. Participant 2 and 3 were not in physical proximity during the experiment, but we know they were related, as the correspondence logs showed that they called each other during the experiment. The numbers on the edges denote the number of Bluetooth discoveries indicating the time spent in proximity.

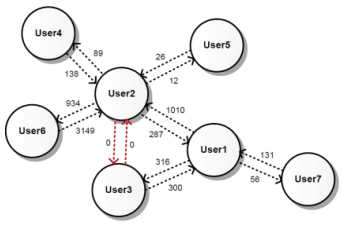


Fig. 3. Social relations of the 7 participants mapped based on Bluetooth device discoveries (physical proximity)

Based on this data it was possible to establish the context of use of the mobile media player on the mobile phone. It was possible to determine the time and places in which the music was being played. Furthermore it was possible to determine the people present when the media player was being used for music playback.

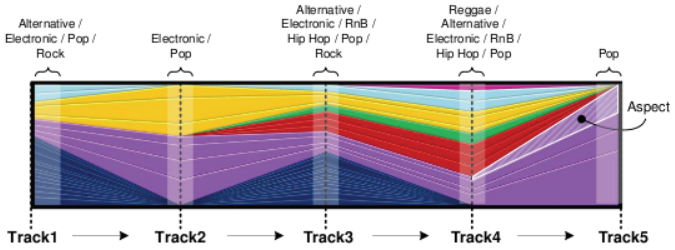


Fig. 4. Example five track sequence genre signature. Each color corresponds to a unique genre (obtained via last.fm)

The analysis of the music used the track metadata that was acquired from the media player during playback as the starting point. Based on the artist and song title the collaborative tagging of music tracks available from Last.fm was used to establish the music genres for each track being played (a genre signature). Furthermore we considered at least three songs played in a row to belong to a track sequence. In a similar manner a genre signature for these track sequences was calculated based on the individual genre signatures of each song, as illustrated in Fig. 4. This allowed us to consider and compare which genres of music were being played over time by the participants and the particular context in which they were played.

DISCUSSION

When studying the usage patterns we found that the genres of music that were listened to over time highly depended on the context of the user. In some places one set of genres were typically played, whereas in other places or in transition between places a different set of genres were being played. Furthermore, the level of interaction (such as skipping songs) also depended highly on the context in which the mobile application was being used. Transitions between places were characterized by frequent interaction (skipping and choosing songs) on the media player, whereas in known places the interaction was less frequent (less skipping of songs), meaning that participants would play a longer sequence of songs uninterrupted. An example of being in transition between places could be riding a bus, where the participant could have time to interact more frequently with the media player application, which could explain our findings from the data acquired in the experiment. However, it must be underlined that our findings are based on only two weeks of data acquired from 7 participants. Thus further experiments must be carried out in order to establish whether the findings mentioned above can be generalized. Nevertheless, we find that studying the mobile user experience in context has the potential to be a valuable source of inspiration for designers. For instance in terms of suggesting alternative user interfaces for navigating and selecting the content in the media player. Thus the mobile context could potentially play a much more prominent role in mobile applications, such as the media player. An obvious example is for recommendation systems that not only recommend music based on music similarity, but also contextual similarity.

As for evaluating the mobile user experience we find that the point in the process where the evaluation takes place has profound implications for the method to be chosen. A mobile application under development can probably benefit from simulating different contexts of use in a laboratory setting as proposed in the literature. However, if a mobile application has already been deployed we find that contextual information as discussed in this paper can offer useful information and provide valuable insights to the use of the application. As found in the case studied here we discovered how the context of use had implications for the interaction with and the content chosen in the media player.

CONCLUSIONS

Our mobile context toolbox for mobile phones has allowed us to carry out several experiments, where we have observed mobile phone users using a mobile phone in real-world settings. Based on the logged information from multiple embedded mobile phone sensors we have been able to establish information about the time and context of use of the media player application on mobile phones and we have been able to identify how the context has implications for the use of the application. We conclude that contextual information can offer valuable insights to the where and when of mobile use and provide valuable insights on the aspects having implications for the mobile user experience.

ACKNOWLEDGMENTS

We would like to thank the participants that took part in our experiments. Additionally thanks to Nokia for devices.

REFERENCES

[1] Barnard, L., Yi, J.S., Jacko, J.A. and Sears, A. (2007) Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing 11(2)*, 96.

[2] Duh, H.B.L., Tan, G.C.B. and Chen, V.H. (2006) *Usability evaluation for mobile device: a comparison of laboratory and field tests*. Proc. of the 8th conf. on HCI with mobile devices and services, 186

[3] Froehlich, J., Chen, M. Y., Consolvo, S. and Harrison, B. and Landay, J. A. (2007) *MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones*. Proc. 5th Int. conf. on Mobile systems, applications and services, 57-70.

[4] Jensen, B. S., Larsen, J., Hansen, L. K., Larsen, J. E. and Jensen, K (2010). *Predictability of Mobile Phone Associations.* Int. Workshop on Mining Ubiquitous and Social Environments.

[5] Jensen, B. S., Larsen, J. E., Jensen, K., Larsen, J. and Hansen, L. K. (2010). *Estimating Human Predictability From Mobile Sensor Data*. IEEE Int. MLSP Workshop.

[6] Kjeldskov, J. and Skov, M.B. (2007). Studying usability in sitro: Simulating real world phenomena in controlled environments. *Int. Journal of HCI 22(1)*, 7-36.

[7] Kjeldskov, J., Skov, M.B., Als, B.S. and Høegh, R.T. (2004) *Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field*. In Proc. Mobile HCI, 529-535.

[8] Larsen, J. E. and Jensen, K. (2009). *Mobile Context Toolbox: an extensible context framework for S60 mobile phones*. In Proc. EuroSSC 2009, Springer LNCS 5741, pp. 193-206.

[9] Obrist, M., Meschtscherjakov, A. and Tscheligi, M. (2010). User Experience Evaluation in the Mobile Context. *J. of Mobile TV: Customizing Content and Experience*, 195-204.

[10] Oulasvirta, A. (2009) Field experiments in HCI: promises and challenges. *Future Interaction Design II*, 1-30

[11] Oulasvirta, A., Tamminen, S. and Roto, V. and Kuorelahti, J. (2005) *Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI.* Proc. of the SIGCHI conference on Human factors in computing systems, 928.

[12] Roto, V. and Oulasvirta, A. (2005) *Need for non-visual feedback with long response times in mobile HCI.* Special interest tracks and posters of the 14th international conf. on World Wide Web, 775-781.

[13] Zandi, N., Handler, R., Larsen, J. E. and Petersen, M. K (2010). *People, Places and Playlists: modeling soundscapes in a mobile context*. 2nd Int. Workshop on Mobile Multimedia Processing (WMMP2010).

What can we get “help” to observe when it comes to mobile use and mobile user experience?

Stina Nylander
Swedish Institute of Computer Science
Box 1263, 16429 Kista, Sweden
stny@sics.se

ABSTRACT
Mobile devices and mobile services have been around long enough for the research community to start thinking about the next step in studying them: larger user groups and longer periods of time. Strictly quantitative methods are not very useful when it comes to studying user experience so we need to find scalable ways to support our qualitative methods to be able to take this next step. This paper reflects on automatic gathering of context data as one such way.

INTRODUCTION
Cell phone use is nowadays so pervasive in many parts of the world that we can no longer consider it a new technology. It is a highly integrated part of many people’s lives and should be studied as such. Until now, many studies of mobile use has been conducted on rather small user groups (e.g. [3, 13, 15]). I believe that we now need to conduct longitudinal studies of large groups of participants in order to fully grasp the role of the mobile use and the character of the mobile user experience.

Most of the existing research on mobile use and user experience is qualitative work using methods such as self report and interviews. Simply extending those studies in time and including more users would not work. That would be too demanding for both participants and researchers. However, studying mobile use and mobile user experience only with strictly quantitative methods would miss many aspects of the activities.

My main interest in this is to explore scalable methods that can help us gather as much data as possible about the mobile use situation. What can we add to our qualitative methods that can help us study larger groups and still keep some of the qualitative aspects in our work?

EXISTING WORK AND THEIR METHODS
Various aspects of mobile use have been studied in the HCI domain in the last decade. Most of the published work is qualitative, and the predominant methods are different sorts of self report. For example, Palen et al. [17] used voice diaries, i.e. participants called a voice mail service to report the use of their new cell phones and Isomursu et al. [6] used experience clips where participants videotaped each other. More traditional paper diaries were used to explore text messaging among British teens [3], internet use from cell phones [11, 13], and mobile video watching [14]. Self report data can be unreliable since participants forget to report or choose to report some parts of the relevant data. However, it allows researchers to collect subjective data such as motivation and purpose for the mobile use that is not possible with strict direct observation or the use of logging software. Moreover, since mobile use takes place in a number of different places at various times of day, self report is a feasible option to direct observation that in many cases is impossible. There are examples of direct observation of mobile use though. Oulasvirta & Sumari [15] observed how Finnish information workers managed their devices when moving while working. However, their observations were mainly conducted indoors in office buildings.

Logging software is another way of gathering data on mobile usage which has become feasible as mobile devices get more powerful. One example is Kane et al. [9] that installed logging software on participants’ smart phones and computers to compare their web surfing and email use patterns between the devices. Karlson et al. [10] provides an interesting study that did not exactly use logging software but software that sent a screen shot to the researchers every time the participant got interrupted when using the device. The screenshot image provided extra context to participants’ own recollection of events.

Quantitative studies of mobile use are still quite rare. This is probably due to the difficulties to install logging software on a large number of cell phones, or acquire other types of quantitative material such as log data. The proliferation of cell phone brands, models, and operative systems make it very cumbersome to create and deploy logging software, and ISPs are usually reluctant to provide log data of any

kind. There are a few examples though, Kamvar & Baluja [8] conducted a large scale study of mobile search queries, and Hård af Segerstad [5] created a corpus of more than 1100 text messages to study the language characteristics of Swedish teenagers' messaging. New repositories like AppStore make it possible for researchers to act as service providers and spread and application to a large user population that can be studied and will probably be a common data source in the future. McMillan et al. [12] provides maybe the first example of this, distributing their game Hungry Yoshi through Apple Store, using the game to gather data. However, the drawback of data from logging software, ISP logs or service providers' logs are that they are quite decontextualized. They tell us very little about users' motivation to do a certain thing, if the accomplished what they wanted, or how their experience was.

To compensate for on one hand the potential loss of data in self report and on the other hand the lack of context and subjective information from logs, virtually all studies described above complement their data gathering methods with interviews. The interviews make it possible to get subjective information from participants, such as motivation, preferences, or how they experienced their use. Interviews share many of the drawbacks of self report though, such as memory loss and unwillingness to report information that is unflattering for participants themselves.

DISCUSSION OF EXISTING METHODS

Here, I will discuss the above mentioned studies from my two main points of interest, studying large user groups over time.

In the examples given in the Existing Work section, few studies have more than 30 participants. The exceptions are Lee et al. with 75 participants, and the corpora based studies [5, 8]. The methods used are heavily based on qualitative data gathering through various kinds of self report and interviews. These methods are time consuming both for participants that need to keep diaries or use other tools to report their use and experience, and to researchers that need to oversee data gathering and analyze the material. Thus, they do not scale well for large numbers of participants even though they might offer better alternatives than for example direct observation. Mobile use takes place in many locations, sessions are often short and occur when users have a moment to kill [13] and spread over the day from the moment participants wake up until after they go to bed. It is close to impossible to observe the mobile use of a large group of participants without spending an insane amount of work hours and make huge intrusions in their lives.

When studying mobile use and mobile user experience, time is an important aspect. It takes time for users to learn new applications and find out how they really want to use them, and the novelty factor can make people use a service for a short time while they in the long run stop using it. The studies described in Existing Work were rather short, ranging from a few days [15] or one week [3, 13] to a month [11]. Studies of how new applications are received by end users also typically last for a month or shorter (e.g [7, 18, 19]). Longitudinal studies raise problems that are related to those connected to studying large user groups. It is cumbersome for participants to self report their use for long periods of time and that also generates a lot of material that is time consuming for researchers to analyze. In addition, users might drop out of the study or report their use poorly during the study.

To be able to conduct longitudinal studies of mobile use and mobile user experience in large groups of users we need to find new methods or new combinations of methods to avoid killing both participants and researchers.

LOOKING FORWARD

There is no such thing as a free lunch, so we will probably not find simple or automatic methods that can gather high quality data that is easy to analyze from many users during long periods. However, we should explore the possibilities to combine our existing qualitative methods with automatic data collection since that provides us with structured data that is easy to handle large amounts. For example, context strongly impacts mobile use and use experience and can thus provide valuable information. Here, I believe that we should take inspiration from other areas such as context aware services where automatic detection of for example position [7], ambient sound and movement [4], or proximity of fellow motorcyclists [2] has been used to create service functionality. Moreover, there are examples of services that are not strictly context aware but still automatically collect context information and can inspire: the Affective Diary system [18] that serves as a diary where users can add notes and pictures during the day and records messaging activity, Bluetooth presence and body metrics to add more content to the diary notes; the Ubifit system [1] that automatically recognizes various exercise activities, logs them, and presents them to the user. These examples show us that is is possible to collect meaningful data automatically and should inspire us to go further.

It is also important to simplify the user part of self report. Good examples of this are the voice diary from Palen et al. [16] and the sending of screenshots from Karlson et al. [10] even though they do not provide data that is structured and easy to handle. Providing participants with simple and efficient report methods is essential for longitudinal studies.

A more farfetched thought might be to to combine self report data with automatically gathered context information and try to predict the self report data, assuming that there are contextual situations that reoccur and calls for the same self report data. It is perhaps not so likely, but an interesting idea.

CONCLUSION

Automatic collection of context data will not make it easy to conduct longitudinal studies of mobile use and mobile

user experience on large user groups, but it might be a helpful tool. I believe that we need to find fruitful combinations of qualitative and quantitative methods to continue to study mobility.

REFERENCES

1. Consolvo, S., McDonald, D.W., et al. Activity sensing in the wild: a field trial of ubifit garden. In Proceedings of CHI, (2008), 1797-1806
2. Esbjörnsson, M., Juhlin, O., et al., Motorcyclists using Hocman - Field Trials on Mobile Interaction. in *Mobile HCI*, (2003).
3. Grinter, R.E. and Eldridge, M. y do tngrs luv 2 txt msg? In Proceedings of European Conference on Computer-Supported Cooperative Work (ECSCW), (2001), 219-238.
4. Håkansson, M., Gaye, L., et al. More than meets the Eye: An Exploratory Study of Context Photography. In Proceedings of NordiCHI, Oslo, Norway, (2006).
5. Hård af Segerstad, Y. Language Use in Swedish Mobile Text Messaging. in Ling, R. and Pedersen, P.E. eds. *Mobile Communications - Re-negotiation of the Social Sphere*, Springer Verlag, 2005, 313-333.
6. Isomursu, M., Kuutti, K., et al., Experience Clip: Method for User Participation and Evaluation of Mobile Concepts. in *Participatory Design Conference*, (2004), 83-92.
7. Jung, Y., Persson, P., et al. DeDe: design and evaluation of a context-enhanced mobile messaging system. In Proceedings of CHI, Portland, OR, (2005), 351 - 360.
8. Kamvar, M. and Baluja, S. A Large Scale Study of Wireless Search Behavior: Google Mobile Search. In Proceedings of CHI, Montréal, Quebec, (2006), 701-709.
9. Kane, S.K., Karlson, A.K., et al. Exploring Cross-Device Web Use on PCs and Mobile Devices. In Proceedings of Interact, Uppsala, Sweden, (2009), 722-735.
10. Karlson, A.K., Iqbal, S.T., et al. Mobile taskflow in context: a screenshot study of smartphone usage. In Proceedings of CHI, Atlanta, GA, (2010), 2009-2018
11. Lee, I., Kim, J., et al. Use Context for the Mobile Internet: A Longitudinal Study Monitoring Actual Use of Mobile Internet Services. *International Journal of Human-Computer Interaction*, *18* (3). 269-292.
12. MacMillan, D., Morrison, A., et al. Further into the Wild: Running Worldwide Trials of Mobile Systems. In Proceedings of Pervasive, Helsinki, Finland, (2010).
13. Nylander, S., Lundquist, T., et al. "It's Just Easier with the Phone" - A Diary Study of Internet Access from Cell Phones. In Proceedings of Pervasive, Nara, Japan, (2009), 354-371.
14. O'Hara, K., Slayden Mitchell, A., et al. Consuming Video on Mobile Devices. In Proceedings of CHI, San Jose, CA, (2007), 857-866.
15. Oulasvirta, A. and Sumari, L. Mobile Kits and Laptop Trays: Managing Multiple Devices in Mobile Information Work. In Proceedings of CHI, San Jose, CA, (2007), 1127-1136.
16. Palen, L. and Salzman, M., Voice-mail diary studies for naturalistic data capture under mobile conditions. in *CSCW 2002*, (New Orleans, Louisiana, 2002), 87-95.
17. Palen, L., Salzman, M., et al. Going wireless: behavior & practice of new mobile phone users. In Proceedings of CSCW, (2000), ACM Press, 201-210.
18. Ståhl, A., Höök, K., et al. Experiencing the Affective Diary. *Personal and Ubiquitous Computing*, *13* (5). 365-378.
19. Sundström, P., Ståhl, A., et al. In Situ Informants Exploring an Emotional Mobile Messaging System in Their Everyday Practice. *International Journal of Human-Computer Studies*, *65* (4). 388--403.

Unsupervised User Observation in the App Store: Experiences with the Sensor-based Evaluation of a Mobile Pedestrian Navigation Application

Benjamin Poppinga, Martin Pielot
OFFIS - Institute for Information Technology
Oldenburg, Germany
{poppinga, pielot}@offis.de

Niels Henze, Susanne Boll
University of Oldenburg, Germany
{niels.henze, susanne.boll}@informatik.uni-oldenburg.de

ABSTRACT

Traditional methods to observe a participant during a field study are often not very scalable and obtrusive. Given the facts of more and more available smart phones and mobile distribution channels, e.g. Apple App Store, the emerging logging observation method gains an increasing attention. In this paper we report on our experiences of conducting a user study in the Android Market by relying on the logging methodology, and thus on sensors of a common mobile smart phone. Based on our preliminary findings we identify the major challenges a researcher needs to face, when an in-market study should be conducted.

1. INTRODUCTION

To observe the mobile user experience various observation techniques exist. For field studies often ethnographic observation techniques, like shadowing, are used. In shadowing an experimenter follows a participant and takes notes on the observed behaviour. Shadowing is known to be highly situated [3, 5]. However, this technique doesn't scale very well. Additionally, because of its obtrusiveness, it might change the observed participant's behaviour.

To overcome the disadvantages of low scalability and high obtrusiveness, new observation methods are developed. In theory, passive automated logging through sensors seems to reach almost the same situatedness, while being scalable and unobtrusive [3, 5]. In practice logging has been rarely applied for mobile observation during the last years. One reason for this might be that suitable data sources, e.g. sensors, were not available on a common mobile device and needed to be self-build [1]. While these self-build sensor systems reduce scalability, they are able to infer users' everyday situations [2].

Nowadays a commercial off-the-shelf mobile smart phone, like the iPhone, has a variety of sensors integrated. Thus, principles were earlier specialized hardware was required, can now be ported to the phone (e.g. a pedometer). McMillan et. al. [4] successfully applied logging in the large scale in a mobile game which they submitted into the App Store. Given all these sensors makes logging more and more interesting as scalable, unobtrusive, and situated observation technique.

However, while there are some well-known concepts, like e.g. a pedometer algorithm, available and ready for instant application, a holistic view on how to use, combine, and apply sensors to log a specific user action is missing. In this paper we present our approach towards unsupervised in-market studies and identify three major challenges based on our preliminary findings.

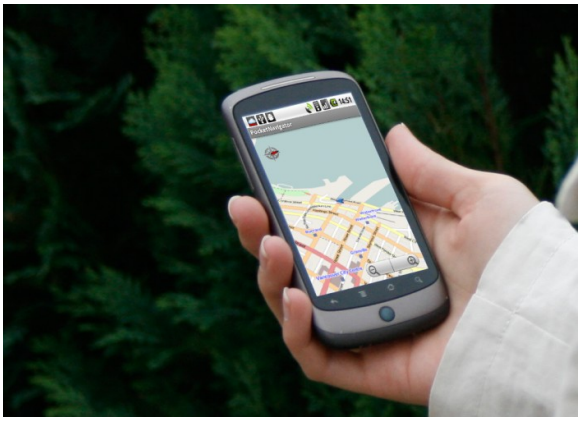


Figure 1: The PocketNavigator is a mobile pedestrian navigation application. Our integrated sensor-based observation technique is invisible for the user. However, the participation within the user study is defined as opt-in to maintain ethical correctness.

2. EXPERIMENT DESIGN

Originating from the interest to provide tactile feedback as additional navigation aid, we developed the PocketNavigator[1]. The PocketNavigator is a personal navigation application, available for free in the Android[2] market (see Figure 1). Designed as traditional map-based application, a map surface, the user's location, and a waypoint-based route towards an arbitrary destination can be provided [6].

However, in addition the application is complemented by a concept that encodes the direction towards the next waypoint in vibration patterns. If the waypoint is straight ahead of the user, two vibration pulses of equal length are shown.

[1] `http://www.pocketnavigator.org/`, last visited August 31, 2010.
[2] `http://www.android.com/`, last visited August 31, 2010.

If the next waypoint is on the right, the duration of the second pulse increases. The same happens to the first pulse, if the waypoint is on the left. If the waypoint is behind a user, three pulses are shown.

The additional values we assumed for the tactile feedback are that a user will need to watch on the display less often, will do less navigation errors, and will be less often disoriented. These three assumptions serve as hypothesis for an experiment we decided to conduct remotely and unsupervised in the Android Market. If a concrete research question should be answered, it is recommended to define the hypothesis right before any sensor data is gathered.

Then, for each hypothesis the observable values need to be identified. Therefore one should think about what are observable events, supporting or not supporting the hypothesis. The own imagination or personal, field-related experience are a good entry point for these definitions. However, often comparable studies in literature already propose a definition how a specific parameter can be observed. In case of the PocketNavigator, we decided to measure e.g. if the user looks at the display be using the roll and pitch angle, as there is no eye tracking available.

In the last step the to be measured values will be assigned and represented through available sensors. In the exemplary case if the user is watching the display we decided to use the accelerometer, which is able to provide the required values roll and pitch. As one can imagine, every matching of an hypothesis to an observable behaviour and then to a set of sensors induces some noise and inaccuracy. Thus it is necessary to design and validate the sufficient representation of a to be observed behaviour iteratively. At some time if the selected representations are reasonable accurate, the experiment can be released to the market.

3. IDENTIFIED CHALLENGES

The PocketNavigator is still available and the study (i.e., the logging) is still ongoing. Until now we can report of 500 people who participated in the study. In this section we transfer our experiences into general challenges which need to be approached to further establish sensor-based observation in mobile applications. We identified three challenges: recruiting, analysis, and the question on internal validity.

3.1 Recruiting

In the participant recruitment process, the very first aspect is that a good application title and description needs to be provided in the market to attract participants. Further, a nice application icon and some screenshots can also attract users. Without question the application should provide the advertised functionality and should be robust and reliable.

To fulfil the ethical requirements of the society or the projects requirements, where the application is developed in, the study needs to be announced to the user in a sufficient and apparent way. Thus, the mentioning of the study in the application's general terms and conditions is ineligible. More, a separate menu entry should clarify the purpose and frame of the study, as a traditional informed consent does. Obviously the participation in the user study should be an opt-in instead of an opt-out. Like in a traditional field study, a user should be able to withdraw at every time.

Early releases of the PocketNavigator presented the study in a separate info view, selectable through the application's menu. If interested in participation, the user must explicitly check a checkbox. However, under this condition the acquisition of participants proceeded quite slow. In an updated version, we proactively announce the study through a simple and short pop up dialog. If the user disagrees to participate in the study, a more detailed info screen on the study is shown, trying to convince the user. This approach leads to a participation rate of about 5 to 10%.

3.2 Data Analysis

The recording of sensor values within the application is one thing. However, the gathered data of each client must be available to do analysis. Therefore we used a custom made server, to which each client connects via sockets and transmits the gathered data in chunks. Alternatively a script, running on an existing server can be used, like e.g. PHP. This can also be easily combined with encryption algorithms, like SSL. To avoid loss of any data, a backup and watchdog is recommended.

Once the application is in the market and the participants are sending their data, it's possible to do some analysis. From our personal experience we recommend to do the analysis on a regular basis, to identify overlooked aspects or strange application behaviours, which can be solved by adapting the logging algorithms. With every adoption it is important to monitor the version a participant is using to not confuse different types of data during analysis.

The actual analysis is done by custom made tools, as universal analysis tools most probably doesn't exist for a specific use case. In case of the PocketNavigator we build one application which does a summary over the data of all participants and prepares an output file, which is readable by e.g. Microsoft Excel, to do some further analysis. Second we build an application which is able to replay the behaviour of an individual user by displaying the values of the sensors in real time. The first tool is more suited for quantitative analysis, while the second tool can give insights in individuals situations, which can be treated as qualitative data.

3.3 Internal Validity

In controlled experiments internal and external validity are two contrasting aims. Internal validity is the validity of the inference of causal relationships, or how confident the observed effects can be attributed to the experimental manipulation. External validity is the validity of the generalisation of experimental findings, or how confident the observed findings can be generalised beyond the experiments setting.

Typically, experiments (especially those conducted in the lab) focus on internal validity. The disadvantage of this approach is that the experimenters often can only carefully generalise their findings to actual usage scenarios. Studying applications in "real" use by making them available to a wide range of users - as we did with the PocketNavigator - stresses external validity at the expense of the internal validity.

In the case of the PocketNavigator we identified two factors that threaten the internal validity: the design as quasi-experiment and the unpredictable usage.

3.3.1 Experiment vs. Quasi-Experiment

In a true experiment, conditions get allocated randomly. As we are studying the effect of the vibro-tactile feedback technique, in a true experiment, half of the participants would be chosen to use the tactile feedback and the other half not.

However, in our actual study design we allowed the participants to choose for themselves if the tactile feedback should be turned on or off. We were afraid that people get annoyed by the tactile feedback, giving the application bad ratings in the Android Market, and in consequence deterring potential future users.

Thus, the experiment is not a true but a quasi experiment. Due to the lack of randomization it is harder to rule out confounding variables and unsystematic variance. In our case, people that decide to use the tactile feedback could have certain traits or be in certain situations which favour or disfavour the usage. For example, if only people with lots of experience use the tactile feedback, because they are more open to new innovations, their navigation performance could be disproportionally better than average because of either their experience or the tactile feedback.

3.3.2 Unpredictable Usage

Another problem that turned up is the unpredictable usage of the application. In a typical experiment the task is well-defined and well-known to the person analysing the data. In the case of the PocketNavigator we neither have a way to dictate a certain usage pattern to the users nor can we completely understand the usage at a certain time. In the following we give a few examples of unpredicted usage patterns that could have threatened the internal validity if we had not identified them:

Example 1: Lying on table. In the first stream of data we received from our participants we had many situations where no navigation at all took place. Having a close look at the data, the accelerometer indicated that the device was oriented parallel to the surface and the GPS signal showed no walking speed. From these data we inferred that many users might be testing the application indoors first, leaving the device on the table and probably keep running the application in the background.

Example 2: Car Driving. At a later stage we were investigating the effects of the tactile feedback on the average walking speed. However, we were surprised by the huge variance in the walking speed averages. Taking a closer look at the individual data we found that some walking speeds were unnaturally high (e.g. $>$ 70km/h in average) for pedestrians, so we inferred that people had used it in their cars or any other vehicle.

Example 3: Background idling. Android offers parallel and background executing. As the PocketNavigator is expected to run in the pocket we designed it to continue running when the screen saver is activated or another application is pushed to the front. The problem is that the Android OS does not really terminate applications but only pushes them into the background until the resources are needed otherwise. Thus, in a few cases the application kept running in the background producing nonsense data.

4. CONCLUSION

In this paper we report on our experiences on applying a sensor-based virtual observer to the Android Market. We identify three major issues, which need to be considered and approached in future developments: recruitment, data analysis, and internal validity.

In our future work we want to extend and apply the in-market observation methodology for true experiments, as well as for more open research questions, which can not be answered within an experiment. Additionally we want to apply logging as observation method in a traditional field study to prove the validity of the method. Finally we are interested in the advantages, disadvantages, and limitations of the virtual observer in different settings.

5. ACKNOWLEDGEMENTS

The authors are grateful to the European Commission which co-funds the IP HaptiMap (FP7-ICT-224675). We like to thank our colleagues for sharing their ideas with us.

6. REFERENCES

[1] T. Choudhury, G. Borriello, S. Consolvo, D. Haehnel, B. Harrison, B. Hemingway, J. Hightower, P. . Klasnja, K. Koscher, A. LaMarca, J. A. Landay, L. LeGrand, J. Lester, A. Rahimi, A. Rea, and D. Wyatt. The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing*, 7(2):32–41, 2008.

[2] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay. Activity sensing in the wild: a field trial of ubifit garden. In *Proc. of CHI*, 2008.

[3] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proc. of MobiSys*, pages 57–70, New York, NY, USA, 2007. ACM.

[4] D. McMillan, A. Morrison, O. Brown, M. Hall, and M. Chalmers. Further into the wild: Running worldwide trials of mobile systems. In *Proc. of Pervasive*, volume 6030 of *LCNS*, pages 210–227, Helsinki, Finland, May 2010. Springer.

[5] I. Mulder, H. ter Hofte, and J. Kort. Socioxensor: Measuring user behaviour and user experience in context with mobile devices. In *International Conference on Methods and Techniques in Behavioural Research*, 2005.

[6] M. Pielot, B. Poppinga, and S. Boll. Pocketnavigator: Vibro-tactile waypoint navigation for everyday mobile devices. In *Proc. of MobileHCI*, 2010.

Understanding Mobile Social Behaviour Using Smartphones

Fehmi Ben Abdesslem and Tristan Henderson
School of Computer Science, University of St Andrews
St Andrews, Fife KY16 9SX, UK
{fehmi,tristan}@cs.st-andrews.ac.uk

ABSTRACT

Understanding the behaviour of users as they share information with mobile social applications is important for enhancing their experiences and improving the services provided. In this paper, we present an approach to studying users' behaviour with the Experience Sampling Method, using a single mobile device to ask questions of users and simultaneously monitor their activities and contexts. While our approach presents benefits compared to traditional questionnaires, we also present the challenges faced, and the problems still to be explored.

1. INTRODUCTION

More and more mobile social applications have become available to smartphone users, allowing them to share personal information with their social networks anywhere at any time. Designing such applications must not only provide users with the ability to share information, but also take into account their concerns regarding disturbance, intrusiveness, and social implications of sharing personal information in their everyday lives. Failure to do so may lead to public outcry or expensive redesigns of services after they have been launched, as has occurred recently with Facebook's privacy controls[1], or Google Buzz.[2]

Studying users' behaviour is paramount for understanding these concerns. Formal interviews and questionnaires allow us to collect self-reported information about users' behaviours when using mobile social applications, but users may forget some details about their experiences or report inaccurate information when answering questionnaires. The behaviour of mobile social application users can also be studied by analysing the information shared on social network sites (SNSes), but this only allows the examination of those information that have been shared, rather than the information that have not been shared, or the contexts in which users do not wish to share. A third way to study users' behaviour, that addresses some of these drawbacks, is the Experience Sampling Method (ESM) [6]. ESM is a diary method that consists of asking participants to stop at certain times, either on a pre-determined basis (signal-contingent) or when a particular event happens (event-contingent), and report about their experiences in real time.

In this position paper, we advocate the use of ESM, possibly in addition to questionnaires and analyses of SNS accounts, for capturing information about mobile users' behaviour *in situ*, when the mobile social application is actually used. We share our experiences in using a mobile phone for asking questions of participants about their self-reported experiences, and for collecting data about their actual, rather than self-reported, behaviour.

The remainder of this paper is organised as follows. In the next section we describe our testbed using ESM with smartphones to study the behaviour of mobile social application users. We then share our experiences in deploying and using such testbeds by presenting their benefits compared to surveys and SNS anaysis in Section 3 and the challenges raised in Section 4. We finally conclude the paper in Section 5.

2. EXPERIENCE SAMPLING WITH SMARTPHONES

ESM has already been widely used to study users' behaviour by polling participants in real-time during their everyday lives, particularly studying how they share their location. Consolvo et al. [4] use PDAs to ask signal-contingent questions to participants at random times about location disclosure to their social relations. Disclosure to their social network was hypothetical and questions were both asked and answered through the same device. Anthony et al. [1] study how privacy preferences vary with place and social context by sending basic signals to participants using pagers, for them to fill in questionnaires in a notebook. Disclosure was also hypothetical, and since questions were too numerous to be easily answered on an electronic device, they were both asked and answered through the notebook.

Our research is interested in how, when, where and to whom people share their locations with their social network, to better understand their privacy concerns. We go a step further than previous experiments by actually disclosing location to the participants' social network. Moreover, we use a single device to detect location, ask ESM questions, and then collect both ESM answers and detected locations. We believe that carrying only one device is much less intrusive than carrying a notebook to answer the questions, a pager for the signals that an ESM question must be answered, and a sensing device to collect automatic data such as location.

Our first experiment [2] involved 40 participants sharing their location to their social network with a smartphone over the course of one week. Each participant was given a Nokia N95 8GB smartphone, constantly running a custom application that detects their location using GPS and Wi-Fi scanning. Locations were regularly uploaded to our server through the cellular network, and published on their Facebook SNS account according to their disclosure choices. To this end, participants were asked during a pre-briefing session to set up friend groups on Facebook if these did not already

[1] http://mashable.com/2010/05/23/facebook-ceo-mistakes/
[2] http://news.bbc.co.uk/1/hi/technology/8517613.stm

Figure 1: Using a smartphone to ask a participant whether he/she would share a photograph with his/her social network friends.

exist (e.g., family, classmates) and default disclosure choices.

Six types of signal- or event-contingent ESM questions were sent to the participants through an SMS handled and displayed by the application:

- **Signal-contingent.** Ten signal-contingent questions were sent each day, at random times of the day.
  1. *"We might publish your current location to Facebook just now. How do you feel about this?"*
     We asked the participant about his/her actual feeling by reminding that his/her location can be published without any consent. The participant could answer this question on a Likert scale from 1 to 5.
  2. *"Take a picture of your current location or activity!"*
     The participant could accept or decline to answer this question. If the participant answered positively, the phone's camera was activated and the participant was asked to take a photograph. .
- **Event-contingent.** Up to 10 questions per day were sent whenever the system detected that the participant had stopped at particular locations.
  1. *"Would you disclose your current location to:* [friends list]*?"*
     We asked the participant for the friends lists to whom he/she wanted to share his/her location. We first asked if the location could be shared with 'everyone'. If the participant answered 'Yes', then the question was over and the participant's location was shared to everyone on Facebook. Otherwise, if the participant answered 'No', the phone asked if the participant's location could be shared with 'all friends'. If so then the question was over, and the location was shared with all of the participant's Facebook friends. Otherwise we iterated through all of the friend lists that had been set up by the participant. Finally, sharing with 'nobody' implied answering 'No' to all the questions.
  2. *"You are around* [location]*. Would you disclose this to:* [friends list]*?"*
     This question mentions the detected place. This is to determine whether feedback from the system makes a participant share more.
  3. *"Are you around* [location]*? Would you disclose this to:* [friends list]*?"*
     This is the same question as above, but we asked the participant to confirm the location. If the participant confirmed the location, then we asked the second part of the question. Otherwise, we asked the participant to define his/her location by typing a short description before asking the second part of the question. This was to determine the accuracy of our location/place-detection.
  4. *"You are around* [location]*. We might publish this to Facebook just now. How do you feel about this?"*
     This question was intended to examine preferences towards automated location-sharing services, e.g., Google Latitude [5]. Locations were explicitly mentioned to determine whether the participants felt happier when the location being disclosed was mentioned. Note that this question does not ask to whom the participant wants the location to be shared: default settings given in the pre-briefing were used instead.

Figure 1 shows how we ask participants for their sharing preferences when they take a picture of their location or activity.

3. BENEFITS

Analysing the data available on users' SNS accounts is an attractive method for collecting large quantities of data. Paterson and Siek [10] studied information disclosure and awareness of disclosure implications on Couchsurfing.com, an online social networking site where users connect with others interested in traveling and staying at each other's homes. Nosko et al. [8] examined disclosure in online social networking profiles of Facebook users. Patchin and Hinduja [9] determined the extent to which adolescent information disclosure on MySpace.com has changed between 2006 and 2009 by analysing their personal content made publicly available. Lewis et al. collected and studied Facebook profiles and friendship networks of 1,710 college students from 2007 to 2009. But such studies can only focus on the information shared by participants. Nevertheless, the information that is *not* shared is also important, especially if we are to understand the concerns that lead to information not being shared. In our ESM studies, we encourage participants to share their location when a new location is detected and to share pictures of their activity. When the participants decide to keep their location or picture private by not sharing it with anyone, we know that this information does not appear in the user SNS account. Hence, our method also allows us to study what information is not shared by participants. To illustrate that private locations (i.e., not shared to anybody) can be detected, Figure 2 shows the proportion of private locations for each location type. Participants of our experiment kept their location private when at the Library, much often than when they were at a Leisure or Academic place. When participants are at the Library, only 64.7% of locations appear on their

Table 1: Location-sharing choices of participants.

Group	*Number of participants*	*Responses to location-sharing requests*	*Locations that were shared*
Never share location on Facebook	31	431	77.5%
Share location on Facebook	9	95	78.9%

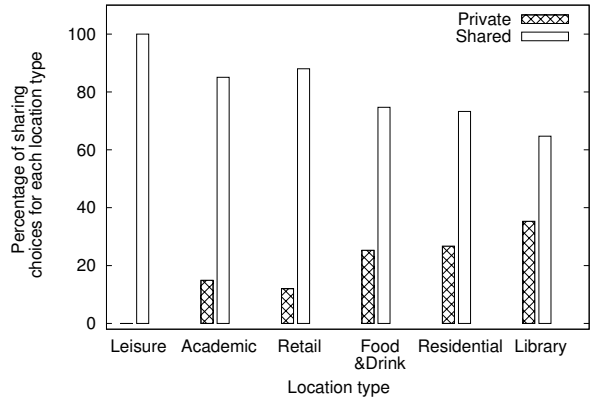


Figure 2: Proportion of sharing choices at different types of locations. Leisure locations were always shared with someone.

SNS accounts, and analysing only this shared information would disregard the important fact that the participants decided to keep their location private when at the Library 35.3% of the time.

Compared to surveys, our method collects answers with the device when participants are actually using the mobile social application, during their everyday lives. This provides us with more accurate answers than when they are asked through a survey where they may forget about the context and their actual behaviour. Moreover, asking the participant several times during the one whole week at random times and locations provides richer data for analysis: we received 2,054 *in situ* answers to the ESM questions and our system detected 2,011 locations. Participants expressed their sharing preferences for 988 of these locations, and took 730 photos, always with sharing preferences. Another benefit of our method is that collected data can be compared to self-reported information provided by questionnaires. Before our experiment, we asked participants to fill in a questionnaire where they were asked whether they shared (at least once) their location on Facebook (e.g., by mentioning their location in their status updates). Out of 40 participants, 31 of them reported that they never share their location on their Facebook accounts. During our experiment (cf. Table 1), those participants who self-reported to never share their location on Facebook actually shared 77.5% of their locations, while participants who self-reported to share their locations on Facebook shared 78.9% of their locations. In other words, while their self-reported behaviours were very different, the actual behaviour of these two groups was very similar, and this behaviour would have been missed by a questionnaire alone.

4. CHALLENGES

Compared to SNS analysis or traditional surveys, implementing the Experience Sampling Method to study the behaviour of mobile social application users is more complicated and time consuming. Our method requires designing, implementing and deploying an appropriate testbed composed of smartphones to collect data and a server to monitor and store these data. But while it would be difficult for our method to be as simple as a traditional survey or an analysis of participants' SNS accounts, there are a few main challenges we can address to improve the method and avoid its potential shortcomings.

A first challenge is to reduce the energy consumed by the smartphones. Using a single device to collect data, ask questions and collect answers necessitates the use of more energy than the normal use of such a device to answer calls. In particular, monitoring users' behaviour continuously may involve multiple sensors to be triggered frequently, which may quickly deplete the battery. Hence, managing efficiently the sensors to save energy is an important challenge to collect data on participants' behaviour in their everyday lives. For instance, in our system, we use the accelerometer embedded in most smartphones to detect motion, and switch off the GPS when the participant is not moving to save energy [3].

Another challenge is to avoid the experiment being too intrusive. Polling participants in their everyday lives may disturb them and answering ESM questions may be sometimes inappropriate. A partial solution is to ask participants for the times they do not want to receive ESM questions. Answering the questions may also take time, especially when they are received frequently. Instead of a notebook, using an electronic device may be easier to use when replying questions, if they are appropriately designed to be quickly replied, by pressing a few keys. But avoiding to ask some questions is even better: detecting an activity or a context instead of asking the participant not only provide other data than self-reported information, but also helps understanding the ESM answers given by the participants. For instance, the location can be detected instead of asking the participant.

Remotely managing the devices while they are used by the participants is also challenging. Participants can move anywhere during the experiment and so monitoring malfunction and and misusage of the device is difficult to achieve. Using smartphones is helpful here, as commercial cellular networks can be used to communicate with the device, rebooting it or for downloading an updated version of the experimental mobile social application.

As for every experiment involving human beings as participants, ethical considerations must be carefully taken into account, especially when the experiment is running during their everyday lives, as personal information may be collected. In particular, privacy issues may be experienced by the participants, and, although unlikely, potential psychological harm, discomfort, or stress. For the latter, the risk is difficult to quantify or anticipate in full prior to the start of the experiment, but the participants always have the option to withdraw from the experiment at any time, without any justification. As for privacy issues, what, how, and when data is collected must be made clear to the participant before they provide any consent to participate, as well as where information is stored and who has access to it. Anonymisation of personal data allowing participants' identification must be guaranteed.

5. CONCLUSION

In this position paper, we advocate using ESM to get better data on the behaviour of users sharing information with mobile social application. ESM allows collecting experiences *in situ*, which we believe is more accurate than when collected later through a survey.

To implement ESM, we suggest using a single device to ask questions and collect the answers, but also to monitor data that is not self-reported to better understand the user's behaviour.

Our use of the ESM methodology has multiple benefits compared to questionnaires, and can provide additional data in the information that is not shared by the user. Nevertheless, there are a number of challenges that we addressed, and solutions that still need further exploration. To this end, we are in the process of designing and running further studies.

Acknowledgements

This work is funded by the EPSRC/TSB *Privacy Value Networks* project (EP/G002606/1).

6. REFERENCES

[1] D. Anthony, T. Henderson, and D. Kotz. Privacy in location-aware computing environments. *IEEE Pervasive Computing*, 6(4):64–72, Oct. 2007. DOI 10.1109/MPRV.2007.83.

[2] F. Ben Abdesslem, I. Parris, and T. Henderson. Mobile experience sampling: Reaching the parts of Facebook other methods cannot reach. In *Proceedings of the Privacy and Usability Methods Pow-Wow (PUMP)*, Dundee, UK, Sept. 2010. Online at http://scone.cs.st-andrews.ac.uk/pump2010/papers/benabdesslem.pdf.

[3] F. Ben Abdesslem, A. Phillips, and T. Henderson. Less is more: energy-efficient mobile sensing with SenseLess. In *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds (MobiHeld)*, pages 61–62, Barcelona, Spain, Aug. 2009. DOI 10.1145/1592606.1592621.

[4] S. Consolvo, I. E. Smith, T. Matthews, A. Lamarca, J. Tabert, and P. Powledge. Location disclosure to social relations: why, when, & what people want to share. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 81–90, Portland, OR, USA, Apr. 2005. DOI 10.1145/1054972.1054985.

[5] Google Latitude. http://www.google.com/latitude/.

[6] R. Larson and M. Csikszentmihalyi. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15:41–56, 1983.

[7] K. Lewis, J. Kaufman, and N. Christakis. The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, 14(1):79–100, Oct. 2008. DOI 10.1111/j.1083-6101.2008.01432.x.

[8] A. Nosko, E. Wood, and S. Molema. All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, 26(3):406–418, May 2010. DOI 10.1016/j.chb.2009.11.012.

[9] J. W. Patchin and S. Hinduja. Changes in adolescent online social networking behaviors from 2006 to 2009. *Computers in Human Behavior*, August 2010. DOI 10.1016/j.chb.2010.07.009.

[10] K. Peterson and K. A. Siek. Analysis of information disclosure on a social networking site. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, A. A. Ozok, and P. Zaphiris, editors, *Online Communities and Social Computing*, volume 5621, chapter 28, pages 256–264. Springer Berlin Heidelberg, Berlin, Heidelberg, July 2009. DOI 10.1007/978-3-642-02774-1_28.

Experiences from the Use of an Eye-Tracking System in the Wild

Kuparinen Liisa
Department of Computer Science and Information Systems
University of Jyväskylä
Finland
+358 (0)400 248079
liisa.kuparinen@jyu.fi

Irvankoski Katja
Department of Cognitive Science, Institute of Behavioural Sciences
University of Helsinki
Finland
+358 (0)40 5962861
katja.irvankoski@helsinki.fi

ABSTRACT
Eye-tracking systems have been widely used as a data collection method in the human–computer interaction research field. Eye-tracking has typically been applied in stationary environments to evaluate the usability of desktop applications. In the mobile context, user studies with eye-tracking are far more infrequent. In this paper, we report our findings from user tests performed with an eye-tracking system in a forest environment. We present some of the most relevant issues that should be considered when planning a mobile study in the wild using eye-tracking as a data collection method. One of the most challenging finding was the difficulty in identifying where the user actually looked in the three-dimensional environment from the two-dimensional scene video. In a concrete matter that means it is difficult to assure whether the gaze is directed to an object short of the user or to a distant object that is partly occluded by the closer one.

1. INTRODUCTION
According to Renshaw and Webb [10], the benefits of eye-tracking include the independence of data from user memory, the eliciting indication of problem solving strategies and a large amount of quantitative data. Examples of situations where the use of an eye-tracking system would be useful are when there is a need to get information about the most important objects used in navigation or to identify which objects in traffic a driver of a car notices and misses. In addition to eye-tracking, other methods such as interviews, observation and performance accuracy are applied to validate or to complete the findings observed in the eye-tracking data.

Another issue is the need to research mobile user experience in the field instead of the laboratory. For example, Nielsen et al. [8] stated that the field setting elicits a significantly increased amount of usability problems, as well as problems with interaction style and cognitive load that are not identified in the laboratory setting. If the research target is to investigate wider user experience in a natural context as well as to identify usability problems, the importance of a field study is even more evident.

The use of eye-tracking systems has been very sparse in the research of mobile user experience. Along with stationary environments, they have been used for example in the research of shopping behaviour, infants' natural interactions, and various everyday tasks [2][4][5]. To our knowledge, the research of mobile user experience in a forest environment is virtually non-existent.

In this paper, we focus on using an eye-tracking camera in a typical Finnish rural environment – a forest. The emphasis of the experiments is more in the validity testing of the eye-tracking method in user tests than in the use of mobile devices in order to discover the issues that must be considered when planning eye-tracking tests in the wild.

2. TESTS IN THE WILD
We executed multiple pilot eye-tracking tests in a forest environment with different tasks in different conditions. The eye-tracking system we used was iView X™ HED from SensoMotoric Instruments. This monocular system consists of an eye camera and a scene video camera which are attached to a bicycle helmet. The first tests were executed without a mobile phone. In that phase, the goal was to assess the feasibility of using an eye-tracking system in a forest environment and to pilot test task settings for future studies. During the tests, we took the users to the forest area to do simple navigation tasks. The tasks included, for example, walking through a certain route with a little guidance (no maps, paper or mobile applications were used), describing what he or she saw, describing how he or she located him/herself and describing the route in such a way that another person could follow it.

After completing the first experiments, a test with a mobile map service was executed. In this single experiment, the user walked a route according to given instructions and located herself on the map. The user was also asked to navigate on foot to a certain position pointed on the map. The composition of the test is presented in Figure 1.

In addition to recording eye-tracking data and interviewing the user during the test situation, the users were interviewed after the tests as well. These post-experiment interviews were conducted to validate and complete the eye-tracking data and observations made in both of the field test cases.

Figure 1. The goals of the test tasks were to resolve the current location on the mobile map and to navigate to a predefined position. The eye-tracking camera was attached on the bicycle helmet and the laptop used for data recording was carried in the backpack.

3. CHALLENGES

In this section, we present the main findings of using an eye-tracking system in a mobile context.

Some problems concerning the use of eye-tracking systems are commonly recognised in stationery environments. Those issues include, for example, the difficulties of tracking a person's eye movements if he or she wears glasses, if his or her pupil size is very small (e.g. when tired), the colour of iris is tepid or if the person has very long, downward or made-up eyelashes [3].

Along with these problems, we also discovered some special issues that should be considered when conducting eye-tracking research in a mobile context.

3.1 Data Quality

There are some issues in using an eye-tracking system in the wild that may risk the quality of data. Perhaps the most challenging issue in executing an eye-tracking test in a field setting is that the off-the-shelf eye-tracking systems are unable to provide definite information about distance of focused gaze in three-dimensional environment [9]. The monocular system we used provides data consisting only of gaze cursor on the recorded scene video, that is gaze position relative to the head (and video frame) [7]. Therefore, we faced situations where we could not be sure whether the user focused his or her gaze on a tree three meters ahead or to the lake that could be seen between the branches of the tree.

Few commercial binocular eye-tracking systems are available such as NAC Image Technology's EMR-9, which has some parallax error compensation. In addition to these, different labs using eye-tracking methodology have been developing eye-tracking systems that resolve the parallax problem and head movement both in natural environment and virtual reality [9][11]. One solution to this problem is the use of thinking-aloud. In addition to the lack of head tracking and depth information, the features of a forest environment make it difficult to define explicit areas-of-interests on recorded scene video data.

Calibration of an eye-tracking camera is much more difficult in the mobile context than in stationary conditions. In a mobile context, especially when investigating mobile device use, the gaze distance varies from couples of dozen centimetres to hundreds of metres. However, the gaze data is the most accurate at the calibration distance due to parallax errors [7]. We handled the calibration by using a large rectangular area, wall or a large paperboard several metres away from the user in the same environment that the test was going to occur. The calibration was then tested by comparing the equivalence of what the video showed and what the user said he or she was looking at. Generally, the calibration needed to be corrected several times. We discovered that calibration should be repeated during the test because it quite easily weakened in motion even though the helmet with the eye-tracking camera was strapped very tight.

Due to the unreliability of the calibration and parallax errors the eye-tracking system may not be trustworthy enough to examine eye movements in the mobile device's small screen. However, the eye-tracking system is very suitable for tracking when, in which situations and for how long a user takes the mobile device in hand and checks it for location or direction.

3.2 Experimental Conditions

Regarding the experimental conditions, the most obvious ones concern weather conditions, which differ from the stable environment of a research laboratory. It is important to take into account that, for example, rain may prevent executing the tests at the planned time. The use of eye-tracking cameras also requires adequate light, thus, it is typically also impossible to execute tests early in morning or late in the night – at least in the winter time. Moreover, the lighting conditions may vary during one single experiment session.

Wearing a helmet or other attachment object with an eye-tracking camera, which has multiple hanging wires, and carrying a laptop in a backpack or a shoulder-case handicaps the movements of the user and influences his or her behaviour, at least until he or she gets used to the equipment. For that reason, it is recommended that the actual test is not performed until the user has had some time to become familiar with the equipment. Improvements to the mobility of eye-tracking systems are being made, but to the best of our knowledge, the current solutions are not yet unobtrusive to the user. For example, in 2008, a research executed with a new kind of eye-tracking solution, light-weighted EOG goggles, was reported by Bulling et al. [1], but also in that solution the user has to carry a laptop with him or her. On the other hand, Tobii Technology has recently introduced Glasses Eye Tracker, which uses smaller recording unit instead of a laptop.

One limiting factor in eye-tracking tests in the mobile context is the low battery capacity that applies to many eye-tracking systems. Keeping that in mind, it is impossible to plan a user test that would last for hours. With our test equipment, the maximum duration for test recordings was about half an hour. The weather conditions (e.g. cold or hot) as well as the bag for the recording laptop also influence this factor.

Finally, it is essential to pay attention to the careful design and definition of test tasks in order to be aware of the user's goals and to interpret the gaze data [5].

3.3 Underlying Cognitive Processes

One should be aware that eye-tracking data does not give all-encompassing data of the allocation of the user's attention. Eye movements can be an indication of a shift in attention (overt attention); on the other hand, a user may shift his or her attention to another target without moving his or her eyes (covert attention) [6]. In our study, the dissociation between where user looked and what she paid attention to was evident in the picture recognition test as well. After the user had walked the route in the forest, she was asked about what she saw and was then shown pictures and asked to decide whether they were taken of the route. The user was shown 16 pictures, of which five were from the route (see example in the Figure 2) and nine were from other forest scenes. The recognition rate was very low; only a couple of the pictures were recognized properly. The results of our recognition test cannot be completely trusted though because they are based on a very small amount of data.

Figure 2. One of the pictures used in the recognition test. The task given to the user after walking a certain route in the forest was to identify whether the shown pictures were taken on the route.

4. CONCLUSIONS

Despite the many challenges of using eye-tracking systems in a mobile context, they provide a valuable method for gathering data that could not be reached by any other method; for example, behavioural methods such as think-aloud verbal reports and reaction-time-based methods lack the kind of data that can be gathered by eye-tracking solutions. The problematic issues presented should be considered when preparing a test with an eye-tracking system in the wild. Some of the issues, such as the weather and light conditions, are easy to take into account. Instead, some of the problems identified in this study, such as the difficulties of defining area of interests in three-dimensional data, should be reacted by the eye-tracking systems' manufacturers.

This paper is in a state of a position paper and many of the presented findings still require validation.

5. ACKNOWLEDGMENTS

This work was supported by the Graduate School in User-Centered Information Technology (UCIT), the Nokia Foundation and Academy of Finland (project 1129346). Great thanks also go to Antti Nurminen, Mikko Berg, Ville Lehtinen and Tuomo Nyyssönen for helping with the research.

6. REFERENCES

[1] Bulling, A., Roggen, D., and Tröster, G. 2008. It's in your eyes: towards context-awareness and mobile HCI using wearable EOG goggles. In *Proceedings of the 10th International Conference on Ubiquitous Computing* (Seoul, Korea, September 21-24, 2008).

[2] Castagnos, S., Jones, N., and Pu, P. 2009. Recommenders' influence on buyers' decision process. In *Proceedings of the Third ACM Conference on Recommender Systems* (New York, USA, October 23-25, 2009).

[3] Duchowski, A. T. 2007. *Eye Tracking Methodology: Theory and Practice.* London: Springer-Verlag.

[4] Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., and Adolph, K. E. 2010. Head-Mounted eye-tracking of infants' natural interactions: A new method. In *ETRA '10: Proceedings of the 2010 symposium on eye-tracking research & applications.* ACM.

[5] Hayhoe, M. and Ballard, D. 2005. Eye movements in natural behavior. *Trends in Cognitive Sciences*. 9, 4, 188-94.

[6] Henderson, J. M. 2003. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences.* 7, 11, 498-504.

[7] iView X System Manual. Version 2.2. SensoMotoric Instruments.

[8] Nielsen, C. M., Overgaard, M., Pedersen, M. B., Stage, J., and Stenild, S. 2006. It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In *NordiCHI '06: Proceedings of the 4th Nordic Conference on Human-Computer Interaction* (Oslo, Norway, October 14-18, 2006).

[9] Pfeiffer, T., Latoschik, M. E., Wachsmuth, I., and Herder, J. 2008. Evaluation of binocular eye trackers and algorithms for 3d gaze interaction in virtual reality environments. *Journal of Virtual Reality and Broadcasting*, 5, 16.

[10] Renshaw, J. A. and Webb, N. 2007. Eye tracking in practice. In *Proceedings of the 21st BCS HCI Group Conference HCI 2007* (Lancaster University, UK, September 03-07, 2007).

[11] Wagner, P., Bartl, K., Günthner, W., Schneider, E., Brandt, T., and Ulbrich, H. 2006. A pivotable head mounted camera system that is aligned by three-dimensional eye movements. In *Proceedings of the 2006 symposium on eye tracking research & applications.*

imprint

Observing the Mobile User Experience, OMUE 10
Volume 1 - October 17th, 2010
ISSN 2190-8427 (Printed edition)
ISSN 2191-2181 (Online edition)

Editor
Benjamin Poppinga
OFFIS - Institute for Information Technology
Escherweg 2, 26121 Oldenburg, Germany
Phone: +49 441 9722 - 138
E-Mail: poppinga@offis.de

organisers

Benjamin Poppinga
OFFIS - Institute for Information Technology
Charlotte Magnusson
Lund University
Wilko Heuten
OFFIS - Institute for Information Technology
David McGookin
University of Glasgow
Niels Henze
University of Oldenburg
Ginger B. Claasen
C-LAB, Siemens AG
Martin Pielot
OFFIS - Institute for Information Technology
Håkan Eftring
Lund University
Jörn Peters
Soest District

acknowledgements

The organisers are grateful to the European Commission which co-funds the IP HaptiMap (FP7-ICT-224675). Special thanks go out to Tobias Hesselmann and Amna Asif for their engagement in the reviewing process. We would also like to thank Alexander Stoer for chairing a session during the workshop.

copyright

HaptiMap