The Challenge of Evaluating the Mobile and Ubiquitous User Experience

Jensen, Kasper Løvborg; Larsen, Lars Bo

Published in:
Proceedings of Second International Workshop on Improved Mobile User Experience

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

The Challenges of Evaluating the Mobile and Ubiquitous User Experience

Kasper Løvborg Jensen and Lars Bo Larsen

Section for Multimedia Information and Signal Processing
Department of Electronic Systems
Aalborg University, Denmark,
{klj,lbl}@es.aau.dk

Abstract. New methods are needed to face the challenges of evaluating the mobile and ubiquitous user experience. By utilizing the sensing and processing capabilities of today's mobile devices it is possible to capture rich quantitative data about the usage and context of mobile and ubiquitous applications in the field. This can be used to conduct large-scale user experiments in the field based on such data. This paper discusses how the capture and analysis of this can be automated and put into a framework to facilitate such studies and presents a prototype implementation.

1 Introduction

The world has never been more mobile than today. Riding an avalanche of technology a multitude of mobile and embedded devices are becoming wirelessly interconnected in networks ranging from local to global, enabling anywhere, anytime access to services and information. Emerging paradigms such as ubiquitous and pervasive computing are changing the way people think about and use computers. With these paradigms comes a wave of novel application build on e.g. context-awareness and adaptive interfaces. A great challenge for HCI researchers and practitioners is to evaluate the user experience of such applications to ensure that they are and will be both useful and usable.

This paper aims to address the way in which researchers and practitioners of mobile HCI can go about evaluating these systems. We suggest that it makes sense to talk specifically about the mobile and ubiquitous user experience rather than the broader UX (User eXperience). Introducing a shorthand notion for mobile and ubiquitous user experience we propose μX $/mju : \epsilon ks/$. This is not to be confused with the abbreviation MUX, which is already used widely for Multi-User eXperience.

1.1 Defining μX

The point of defining μX is to be able to broadly refer to experience of using the diverse and growing class of systems, applications and services, which are inherently mobile and ubiquitous in their nature. They constitute a complex

user experience due to their mobility, ubiquity of use and sensitivity to context in which they are used.

μX can be defined as:

> *"the user experience arising from systems, services and applications with which the interaction is essentially mobile and ubiquitous"*

This can be further specified:

Mobile: Mobility should be attributed to the user and the device(s) during interaction as defined in [4], and not as static properties of the user or devices themselves. E.g. if the primary use of an application on a mobile phone is always in a static context it is not considered mobile interaction and thus not μX.

Ubiquitous: Meant in the broader sense i.e. that interaction happens anywhere and at any time, which together with mobility makes it possible to occur in any context. Note that this is different than Mark Weiser's definition for ubiquitous computing [20] since is can refer to a single personal device and transparency of the technology/computer(s) is not required.

A μX application is an application on a mobile device which gives rise to the mobile and ubiquitous user experience. It is more easily explained by contrasting it to it's opposite, which would be the archetypal desktop application. Desktop applications are run on relatively powerful desktop computers with plenty of resources and a reliable network connection. The environment in which the interaction takes place is controlled and static. Often the user will be sitting down, calm in front of a big screen and focused on all that is going on both visually and audibly.

By contrast, μX applications are used in the wild. They are used anytime, anywhere and in any context - often while the user is on the move, and often as a secondary task to some primary activity done in parallel. The cognitive load in most usage situations is significantly higher than average for desktop applications and the user experience is sensitive to contextual parameters such as environmental (noise, lighting, etc.), social (setting, presence of people, etc.) and network (availability, bandwidth, price, etc.) conditions.

Having defined μX applications, we can now address the challenge of evaluating them, since the only way to ensure that present and future μX applications will indeed be both useful and usable is to build and evaluate prototypes with real users. The question is then how to do this most effectively and efficiently.

1.2 Paper Outline

Section 2 will discuss existing methods and why they are not ideally fit for the characteristics of μX. Section 3 and section 4 will present and discuss a new method and framework from a theoretical and a practical point of view respectively, thus relating it to existing work and discussing implementation

issues. Section 5 presents a prototype implementation of the proposed framework which was used in a longitudinal field study. Section 6 will discuss the pros and cons of the proposed method and point to some research perspectives.

2 A Look in the Toolbox

By taking a look in the current toolbox of methods for user evaluation it can be seen that few methods are actually fit for μX . The characteristics of μX applications as defined above suggest that they should optimally be tested under the conditions in which they are expected to run, i.e. in the field. However, research has shown that existing methods for doing so are very cumbersome and costly to perform and the gain of such field evaluations has been questioned [13].

A survey of mobile HCI research in the years 2000-2002 concluded that even though most researchers thought field testing was the best approach, very few studies were actually based on field experiments [11]. They also showed that the predominant approach is to do experiments in usability laboratories where users are asked to interact with a prototype system in a controlled environment and context. The reasons for this are discussed in the following.

2.1 Laboratory vs. Field Evaluation

The issue at its core is whether or not field experiments are worth the hassle [13]. High experimental control and easy data collection are virtues of the laboratory, while quite the contrary is true for field experiments. The general belief is also that field experiments are more costly with regard to time and resources. The gain is realism when evaluating the services in their natural environment and context of use.

Whether or not to evaluate in the field is widely debated within mobile HCI with some questioning the gain, [13, 10], and others claiming it is indeed worth the hassle [16]. It should be noted that most of the comparative studies are performed based on usability metrics such as number and severity of the found usability problems and not how well the methods investigate the broader UX. If the aim is to look at more than just usability these comparative studies can not be regarded as conclusive.

Often field studies have tried to bring the lab to the field. However, bringing methods like the think-aloud verbal protocol, video capture and human observers out into the field takes a lot of resources [13]. Issues such as logistics, interfering with the context and noisy data cause problems. Besides, in many contexts it would be inappropriate for the test users to go about their business with an entourage of researchers.

Others have tried to bring the field to the lab e.g. in [12] combining field studies (in situ) with lab studies (in vitro) to what they call in sitro - simulating the mobility and context in the lab. The limitations of this method are how real a context can actually be created, that only problems foreseen by the experimenters can be tested and that it is essentially scripted scenarios of use just like a regular lab test. Also, it can be costly to set up.

Both are trying to get some of the best of lab and field testing, but both are essentially augmented versions of existing methods. Our belief is that radically new methods must be added to the toolbox to truly investigate μX.

2.2 Other Aspects

Evaluating some aspects of UX (and thus μX) requires studies over longer periods of time, since the user's initial response to an application might not reflect the way he or she feels about it after using it for a longer period. E.g., learnability and memorability, which are part of the usability goals [17], are difficult to assess from short duration experiments.

Existing methods are very good for finding usability errors in the interface based on fine grained analysis of observational data and user feed-back. Usage of methods such as the think-aloud protocol to capture what the users' is thinking is a good way to get insights into their mental models. However, this potentially influences the experiment e.g. the controversial Hawthorne effect [14].

Further, studies of this type of systems will some times require exploratory experiments since the dynamic nature of μX entail that the situations and problems which will arise cannot be predicted.

In summary, the current trend is to evaluate μX applications with methods adapted from regular desktop applications. To truly uncover the essence of the mobile and ubiquitous user experience and thus to learn how to properly design μX applications new methods must be developed that focus on the unique characteristics of μX as discussed in section 1.1. We believe that it is very important to explore how such studies can be conducted in more efficient and effective ways, thus making it more attractive for both researchers and practitioners to explore μX in the field.

3 A new Approach

The following will describe our ideas for a new method for investigating μX in the field. First an overview will be given and related to existing knowledge and state-of-the-art. After this section 4 will discuss key issues of implementing this method into a framework.

The basic ideas behind the approach are:

- The study of μX should be taking to the field and be unobtrusive to the user's interaction in the natural context.
- Utilize the sensing and processing powers of mobile devices to capture usage and context information.
- Large scale studies over longer duration and an increased number of users.
- Focus on quantitative and objective data and results.

The solution is automating the time and resource consuming parts in doing field experiments, thus enabling them to be run more or less autonomously. In

a large survey of methods for automating usability evaluation in general, [7] defines the activity of doing such evaluations into three main parts which can be automated: capture, analysis and critique. The main idea is to fit the most effective and efficient subset of suitable methods into one coherent framework, which can (at least partially) automate the resource consuming capture and analysis parts in field studies and let experimenters focus on the critique part, i.e., interpreting and using the results.

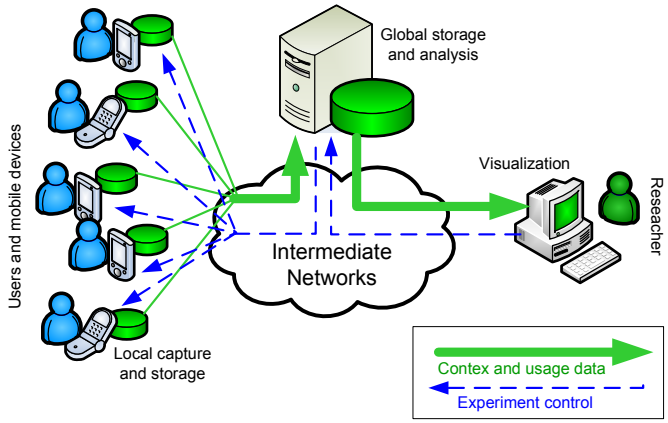

Fig. 1. Conceptual framework showing the flow of data and control

Figure 1 illustrates the main concept in a framework. Usage and context data is automatically captured on the users' devices and transmitted to a central server, where it is synthesized and partly analysed automatically for easy access by the researcher. Visualization is an essential final step in this process. From the researchers point of view the data collection runs autonomously and can be monitored and controlled remotely as needed.

Table 1 roughly summarizes the differences in potential of doing experiments using the proposed method and framework compared to existing practise.

Aspect	Current methods	Proposed method
Number of users	Few	Many
Duration	Short (hours)	Long (weeks)
Data type	Qualitative	Quantitative
Setting	Artificial (lab)	Realistic
Tasks	Scripted	Realistic
Context	Artificial	Realistic

Table 1. Properties of current methods compared to the proposed method

3.1 Automatic Capture and Analysis

Table 2 compares some state-of-the-art tools created for automatically capturing data in field experiments. Larger companies such as Nokia have developed in-house tools which may be even more advanced than these, but they are not openly available.

The data can roughly be categorized as relating to usage, context or subjective user feedback (qualitative). All three types of information have been captured with success in proof-of-concept studies. In this paper the focus is on usage and context. Important questions are what types of data to capture, when to capture it and how to use it afterwards to best investigate the μX?

Tool	Context	Usage	User attitude	Platform
ContextPhone [18]	∨			Symbian S60
MyExperience [3]	∨		∨	Windows Mobile
EDEM [5]		∨	∨	Java
SocioXensor [19]	∨	∨	∨	Windows Mobile
RECON [15]	∨	∨	∨	Windows Mobile

Table 2. Existing tools for capturing data

Both usage and contextual information which are objective and quantitative can be captured at many levels of abstraction and at various granularities of detail. The "easy" approach is to simply capture everything - all the time just to be sure. This is sort of an analogy to video filming usability experiments in the laboratory test. It leaves the researcher with a huge amount of post-experiment data to analyze which is a very time consuming activity. The amount of data produced and the effort needed to analyse it is the main reason for also, at least partially, automating this phase.

Rich data can be sampled with regard to usage and context by using automatic capture. Studies such as [1], [2] and [8] give nice glimpses of what can be gained by mining such data from field studies. However, generally there is a significant lack of proven methods for exploiting these vast amounts of data in order to get insights into the user experience; specially how to include contextual data.

The process of processing the data into user experience information is far from trivial. Every study seems to have its own ad hoc data analysis, and few generic methods exits for addressing this. In [6], Hilbert and Redmiles survey and classify some existing methods and tools for extracting usability information from user interface events. This is further discussed in section 5.

4 Implementing the Framework

As stated, our goal is to combine the best of the methods and tools for capture, analysis and visualization into one coherent framework for conducting large scale

user experiments in the field. From an implementation point of view, the acronym SIERRA summarizes what we believe to be a range if essential qualities for such a framework to have if it is to gain widespread use for evaluating μX systems.

Safe: Ensure privacy and security of the user's data. This also implies robustness.

Invisible: Be transparent to the users and not interfere with the user experience. This also implies efficiency and autonomy.

Efficient: Have a low footprint on the mobile device with regard to CPU and memory usage. Use of network resources and battery should be minimal.

Robust: Should be fault tolerant and be able to recover from system crashes and take care of unexpected errors and events. Autonomy is implied by this.

Remotely controllable: Have some remote experiment control so that e.g. the logging policy can be changed by the experimenters during a longitudinal study.

Autonomous: Some degree of "intelligence" is needed e.g. for doing opportunistic data reporting when network conditions are good.

The above mentioned aspects all dictate how the framework should operate at run-time i.e. during the experiment. Of equal importance are requirements for how the framework should be set up, deployed and used before and after the experiment. It should be sufficiently generic to apply to a large range of platforms and scalable enough to apply in all sizes of experiments. Also it must be time and resource efficient to set up.

5 Prototype Framework

Figure 2 shows a prototype implementation of a framework, which was used for a field evaluation of a mobile diabetes management service called DiasNet Mobile. Detail of the service, the study and the results can be found in [8] and [9].

In figure 2 interpretation refers to what [6] calls *transformation* i.e. the abstraction of low level log events into higher level concepts of interest. Part of this process is also *selection*, which is separating interesting data from "noise". Such transformations have been done to transform the raw data of low level interaction and events into a logical structure that sorts the user's interaction into sessions, activities and actions. The translated log will form the basis for the main analysis, which is based on a combination of *sequence detection* and *calculating counts and summary statistics*. The output from the automated analysis is an annotated log file and a set of tables which can be imported directly into any spreadsheet program, R, Matlab etc. A set of scripts are used for automatically generating the desired graphs and diagrams representing *visualization* which is applied as a last analysis step to draw on the human brain's ability to visually recognize patterns and trends.

The study of DiasNet Mobile did however not include contextual information. Another study has been done resulting in a specific tool for capturing context and usage information called RECON (REmote CONtext) [15]. This tool has

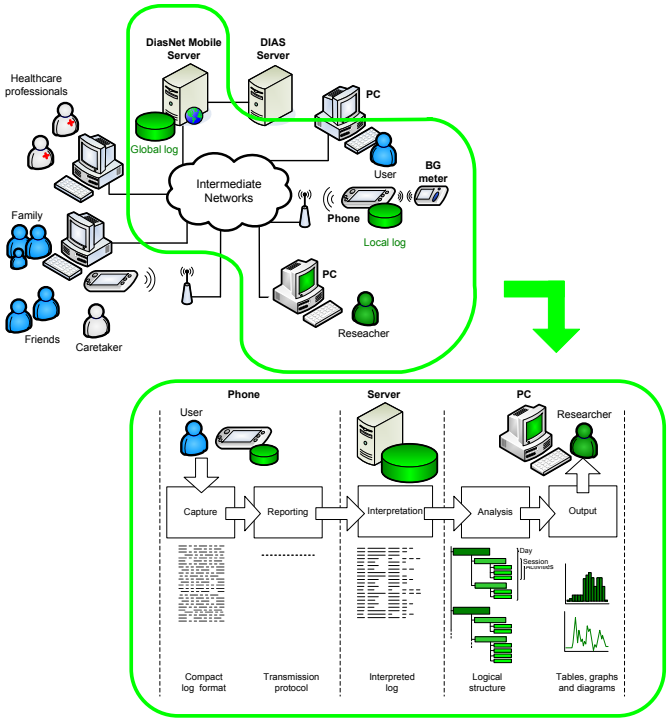

Fig. 2. The implemented prototype framework and how it was used in the DiasNet Mobile field study

so far been tested with a few users, but has not yet used in a real large scale field evaluation. Briefly described, RECON is able to monitor interaction with applications, recording of context e.g. localization using WiFi. It can also prompt users for subjective input triggered by context events or specified usage patterns modelled as Finite State Machines. The logging policy can be controlled remotely during the experiment.

6 Discussion

The approach presented here is purely quantitatively oriented. The need for qualitative and subjective measures is fully recognized as being essential for uncovering the true user experience. The focus on quantitative data may require grounding by qualitative input and feed-back from the users. Such knowledge can be gained from using existing and proven methods or through some of the tools presented in section 3.1 which support user feed-back through various experience sampling methods. This is a very interesting combination indeed.

Doing this type of studies has several major drawbacks which must be addressed. The single most critical challenge in our view is dealing with the uncer-

tainties inherent in the proposed type of studies. These uncertainties are rooted in the limited control of the experiment and incomplete knowledge due to lack of data. An interesting question is whether they can be properly compensated for, e.g. if the sheer amount of captured data can be used to filter out this noise, and good data can be uncovered. Sometimes the noise might actually be the most interesting, since it can represent special events. Another limitation is the demand for a functional high fidelity prototype which makes it mostly applicable in the later phases of development.

The method is highly scalable and in theory any number of simultaneous users could be active in the experiment for any period of time. Depending on how many simultaneous users it can be costly to buy equipment for a large scale test, but this should be weighed against the price of human resources and e.g. a usability lab.

6.1 Some Research Perspectives

The following summarizes what we believe to be some of the most interesting research directions related to this kind of framework:

- how to combine this methods with existing (qualitative) methods into a complete UCD methodology for developing μX applications
- how to handle the uncertainties in the data
- how qualitative/subjective user experience sampling can be combined with the quantitative/objective context and usage data.
- how to use the data to build and manage dynamic user models e.g. user preferences in context-aware applications

6.2 Conclusions

We have defined μX and discussed the challenges of evaluating it compared to the currently most used methods inherited from evaluating desktop applications. There is a definite need for new methods to address this and we have presented one such method, namely to do automatic capture and analysis of quantitative usage and contextual data in large scale field studies. We have presented and discussed how this type of studies can be catalyzed through a generic framework incorporating the best tools and methods for capture, analysis and visualization and a prototype implementation of such a framework has also been presented.

7 Acknowledgments

The authors wish to thank Tais Holland Mogensen and Christian Ølholm for their work on RECON as part of their Master's thesis [15].

References

1. A.Oulasvirta, R. Petit, M. Raento, and S. Tiita. Interpreting and acting on mobile awareness cues. *Human-computer Interaction*, 21:97–135, 2007.
2. N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4), 2006.
3. J. Froehlich, M. Chen, S. Consolvo, B. Harrison, and J. Landay. Myexperience: A system for in situ tracing and capturing of user feedback on mobile phones. In *MobiSys*, 2007.
4. L. Gorlenko and R. Merrick. No wires attached: Usability challenges in the connected mobile world. *IBM Syst. J.*, 42(4):639–651, 2003.
5. D. M. Hilbert and D. F. Redmiles. An approach to large-scale collection of application usage data over the internet. In *ICSE '98: Proceedings of the 20th international conference on Software engineering*, pages 136–145, 1998.
6. D. M. Hilbert and D. F. Redmiles. Extracting usability information from user interface events. *ACM Comput. Surv.*, 32(4):384–421, 2000.
7. M. Y. Ivory and M. A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Compututing Surveys*, 33(4):470–516, 2001.
8. K. L. Jensen and L. B. Larsen. Evaluating the usefulness of mobile services based on captured usage data from longitudinal field trials. In *Mobility Conference*, 2007.
9. K. L. Jensen, C. F. Pedersen, and L. B. Larsen. Diasnet mobile: A personalized mobile diabetes management and advisory service. In *Second Workshop on Personalization for E-Health*, 2007.
10. A. Kaikkonen, A. Keklinen, M. Cankar, T. Kallio, and A. Kankainen. Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies*, 1(1):4–17, 2005.
11. J. Kjeldskov and C. Graham. A review of mobilehci research methods. In *Proceedings of the 5th International Mobile HCI 2003 conference*, pages 317–335, 2003.
12. J. Kjeldskov and M. B. Skov. Studying usability in sitro: Simulating real world phenomena in controlled environments. *International Journal of Human-Computer Studies*, 22:7–37, 2007.
13. J. Kjeldskov, M. B. Skov, B. S. Als, and R. T. Høegh. Is it worth the hassle? exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *Proceedings of the 6th International Mobile HCI 2004 conference*, pages 61–73, 2004.
14. R. Macefield. Usability studies and the hawthorne effect. *Journal of Usability Studies*, 2(3):145–154, 2007.
15. T. H. Mogensen and C. Ølholm. Recon: A framework for remote automated field evaluation of mobile systems. Master's thesis, Aalborg University, 2007.
16. C. M. Nielsen, M. Overgaard, M. B. Pedersen, J. Stage, and S. Stenild. Its worth the hassle! the added value of evaluating the usability of mobile systems in the field. In *NordiCHI*, 2006.
17. J. Preece, Y. Rogers, and H. Sharp. *Interaction Design: beyond human-computer interaction*. John Wiley & Sons, 2002.
18. M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen. Contextphone - a prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing*, 4(2), 2005.
19. G. Ter Hofte. What's that hot thing in my pocket? socioxensor, a smartphone data collector. In *Third International Conference on e-Social Science*, 2007.
20. M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, 1991.