

Evaluating User Experience in Technology Pilots

Minna Isomursu

VTT, Finland, minna.isomursu@vtt.fi

Abstract: Computing devices and digital services have been moving rapidly from professional environments into the everyday life. This means that technology will influence the evolution of our everyday environment, including our physical surroundings, social encounters, and development of society. It becomes more and more important to evaluate the effects of technology in the realistic complex setting of everyday life in addition to controlled laboratory environments. In this paper, findings related to evaluating user experience in real-life trial conditions are summarized. The user experience evaluation methods are classified into four groups: (1) methods used before the pilot, (2) methods used during the pilot, (3) methods used immediately after the pilot and (4) follow-up studies. Each class bears their unique goals, possibilities and limitations for collecting user experience related data, and understanding it.

Keywords: user experience, piloting, trials

1. Introduction

This paper discusses the methods and problems related to collecting information about user experience evoked by ubiquitous applications and services in the context of technology pilots. Ubiquitous computing integrates technology with our everyday life and environment. The complexity and constraints of an everyday environment are difficult, if not impossible, to simulate in a laboratory environment. In situ evaluations are needed for evaluating new technological solution or technology based service in a setting with real users in real use environment (Consolvo, 2007). User experience evaluation can be done in this context for predicting or estimating what kind of user experience can be expected in real-world usage setting.

Capturing information about user experience is challenging, and real-life settings make it even more challenging as the everyday life context is complex and cannot be fully controlled. In an evaluation situation, challenges in capturing user experience occur on several levels and phases. Firstly, as human experience is

Please use the following format when citing this chapter:

Isomursu, M., 2008, in IFIP International Federation for Information Processing, Volume 272; *Human-Computer Interaction Symposium*; Peter Forbrig, Fabio Paternò, Annelise Mark Pejtersen; (Boston: Springer), pp. 47–52.

always subjective, the evaluation method should capture relevant parameters describing the user experience, which can then be recorded for analysis. Secondly, as user experience is dynamic (Forlizzi, 2000), it can change and evolve during the process of interaction. For example, at first the user can be happy and excited about the new product, but later become disappointed, sad or even angry if and when problems occur. Thus, user experience must be sampled several times during the use of the product, which often means organising long-term experiments. Thirdly, interpretation of captured data about user experience is difficult. For example, interpreting emotions from facial expressions captured on video has been an active and debated area of research for decades.

This paper discusses issues that arise when user experience is evaluated within the context of a technology pilot. Arranging a technology pilot requires that the technology under evaluation is mature enough that it can be used by real users in a realistic usage environment. However, the technology can be part of infrastructure or process that does not exist yet.

User experience is a subjective state. It does not have an objective reference, and therefore it cannot be objectively measured. An experience of one person cannot be experienced as such by another person. Therefore, systematic capture and analysis of user experience is very difficult. Furthermore, it is extremely difficult for humans to compare even their own experiences when they are separated by time. Human memory about experiences is utterly unreliable thus rendering our ability to recall past experiences so that we could compare them with other experiences, or describe them reliably after time has passed (Robinson, 2002). Also, our ability to predict our own experiences in a hypothetical or future setting is very limited (e.g. Gilbert, 1990). Therefore, the hypothesis of the research reported here is that the most reliable understanding of user experience can be achieved by: (a) evaluating user experience in a situation as close to actual realistic usage situation as possible to avoid the need for users to imagine or predict their experiences in a hypothetical situation, (b) collecting information and description of the experience at the time it happens to avoid the need to rely on the memories of the user in describing the experience, and (c) using the direct subjective information given by the person having the experience for defining and measuring the experience.

2. Research Setting

The work reported here has been done within the context of the SmartTouch (<http://www.smarttouch.org>) project, where ubiquitous mobile applications and services have been implemented and piloted within several application domains in field settings (e.g. Häikiö, 2007). The goal has been to evaluate new technological solutions with real users in real use environment. The hypothesis was that this can provide information about problems and issues that can be expected to occur in

large-scale use in the real-world usage setting. Technology pilots aim at exposing the technology to real use under circumstances that can be observed and followed. In the research described here, user experience evaluation data collected through technology trials were used for evaluating the feasibility of technical construction, usability, value created for the end user, and the ethical issues related to adopting new technology.

3. Experiences From User Experience Evaluation In Technology Pilots

In this chapter, the user experience evaluation of technology pilot are divided into four phases: (1) before use evaluation, (2) during use evaluation, (3) after use evaluation, and (4) follow-up evaluation. Each evaluation phase has its specific goals, evaluation focus, and sets its own requirements for the evaluation methods. The user experience evaluation method should aim at not disturbing or changing the actual usage situation so that the actual user experience will change. However, the fact is that this is extremely difficult, as research has shown that mere measurement of a phenomenon has effects on the phenomenon itself (Morowitz, 1993 ; Heisenberg, 1927).

3.1 Before Use Evaluation

Before-use evaluations proved to be valuable especially for the following: (1) getting information about attitudes and expectations that are relevant for interpreting the results of the pilot, and (2) setting the baseline for the evaluation by describing and measuring the starting point so that improvements and changes introduced by the technology can be identified and measured.

Before pilot use it is the best time to evaluate the attitudes and expectations the users have towards issues that may be relevant for evaluating the results and impacts of the pilot. For example, in the elderly meal-ordering pilot, the attitudes towards the use of mobile phone proved to be a strong impacting factor towards the perceived usefulness of the service. Also, the pilot may result in attitude changes that can be identified only if the attitudes before and after the pilot can be measured and compared. In the SmartTouch project, information about attitudes and expectations were collected with contextual interviews when the sample size was small (under 20 users), and questionnaires when the experiment involved larger amounts of users. In some cases, questionnaires were printed on paper forms, and on others, they were implemented through a web-based survey tool. However, the experiences show that it is difficult to predict before the experiment what could be relevant parameters related to expectations and attitudes that would be needed for interpreting the results, as the values and attitudes of users often unfold only during the piloting. This could be solved by deeper user study

concentrating on the values and attitudes of the users already before the pilot. Also, better models and methods for describing and modeling expectations and attitudes towards ubiquitous and persuasive technology would be needed.

The experiment often aims at improving or supporting the life of the pilot users in some way. For evaluating if improvement happens, the situation before the trial needs to be evaluated. For example, in the elderly meal-order case, one of the goals was to improve the satisfaction of the meal-ordering clients towards the meals offered. This could be done only by first evaluating the satisfaction towards meals before piloting, and then again after piloting. For evaluating improvement, it is crucial to identify right value creation parameters that are used for evaluation.

3.2 During Use

At the beginning of the piloting period, the pilot users are often introduced with the new technology under experimentation, and perhaps trained for using it. Observing the introduction and training situations allows a good opportunity for exploring the issues related to the adoption of the technology in question.

Collecting information about user experiences at the time they happen require in situ data collection methods (Consolvo, 2007) that can be applied during the use of technology. This means that the tools and methods used for collecting user experience data need to be integrated into the everyday practices of the pilot users, just as the technology under evaluation. Our experiences show that when the technology under evaluation is well integrated into the everyday practices of the user and therefore quite invisible, the user experience evaluation method may actually “steal the show” (Isomursu, 2007), if it is more visible and needs more attention and cognitive processing from the user, than the actual technology under evaluation. Humans are not very good at analyzing what actually caused an experience (Dutton, 1974), so it can be difficult for users to identify if the experience was caused by the technology under evaluation, or the user experience evaluation method (or any other event in the life of the pilot user).

During-use evaluation can focus not only on evaluating the user experience evoked by the technology under evaluation, but also how the technology affects the lives of its users. With ubiquitous technology, the technology is often invisible and therefore it is not designed to evoke experiences, but to help in daily tasks.

As the usage situations, including the physical and social environment, usage tasks, etc., may be very different between pilots, it can be necessary to integrate the experience collection method case by case into the pilot experiment. Also, automated compilation of activity logs make it possible to follow the actual usage patterns that have emerged during use.

3.3 After Use

At the end of the pilot use the users usually discontinue using the technology under piloting. This is a point where typically a feedback survey is performed. At this point, the users can report about their user experiences in the form of storytelling, and reflect on their experiences. However, as humans are naturally not very good in memorizing experiences, the limitations of after-use methods must be acknowledged.

After-use evaluation provides an opportunity to evaluate possible changes in attitudes of the users by comparing situations before and after use, and hearing the explanations of users for the possible attitude changes. Experiences indicate that the reply rates for questionnaires made after the pilot are higher than the questionnaires made before the pilot. One explanation might be that the pilot users feel they have more to contribute after the pilot as they are able to tell about their experiences by sharing stories. At this point of pilot experiments, users are familiar with the technology, its limitations and possibilities, and feel they better share the language and concepts used by technology developers and researchers. This can be exploited by combining after-use evaluation with brainstorming or other methods suitable for participatory design. Brainstorming sessions can be used for collecting user experience data, as improvement ideas and new scenario proposals often are loaded with user experience knowledge.

3.4 Follow-up Study

Follow-up studies are valuable in estimating the long-term effects of the experiment. The attitudes of trial users can fluctuate with time, and this does not necessarily end right after piloting. As the pilot use often provides the users with new possibilities to control their lives, depriving them from this feeling of control may have negative and even tragic effects (Schultz, 1978) that can be observed only after some time has elapsed after the experiment. The effects of the pilot experiment should be analyzed not only for collecting information about user experience, but also for evaluating the ethical issues related to the experiment. From the research point of view, it can be problematic to balance between avoiding negative impacts of the pilot and creating high-impact concepts. If the concept is found extremely valuable by the users and it is able to considerably contribute towards a higher quality of life for the pilot users, losing the possibility to use technology after the pilot may have strong negative effect on the well-being of the pilot users. The negative psychological effect of loss can be much stronger than the positive effect achieved through pilot.

However, the goal of concept design is to create high-impact concepts that would be appreciated and valued by the users. Methods and examples for balancing between these two contradictory goals would be welcome.

4. Summary

This paper summarizes experiences about evaluating user experience with experimental pilots. The experiences reported in the paper have been collected in various experiments where the use of new technology has been evaluated in the everyday life of pilot users. In each experiment, a different set and combination of methods were used. For the purpose of this paper, the methods were classified into four classes depending on the point of time they have been used in the pilot process.

As it is very difficult, if not impossible, to fully understand and analyze the human experience, using several different methods in different phases of user experience evaluation can provide the designers and researchers with data that can be used to reveal details and characteristics of an experience from different viewpoints. Therefore, combining methods seems natural. Furthermore, as user experience is tightly context dependent, tailoring methods for each experimental context is probably also necessary. This means, that developing user experience capturing and evaluation methods can be as challenging as developing the technology under evaluation, as they both need to integrate and merge with the everyday lives of the pilot users.

References

- Consolvo, S., Harrison, B., Smith, I., Chen, M., Everitt, K., Froehlich, J., Landay, J. Conducting In Situ Evaluations for and With Ubiquitous Computing Technologies. *International Journal of Human-Computer Interaction*. 22(1&2), 103-118 (2007)
- Dutton, D., Aron A.: Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of personality and social psychology*, 30 (1974).
- Forlizzi, J., Ford, S.: The building blocks of experience: an early framework for interaction designers. in *Proceedings of the DIS 2000 Seminar, Communications of the ACM, New York*, 419–423 (2000).
- Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Zeitschrift für Physik*, 43 (1927)
- Häikiö, J., Isomursu, M., Matinmikko, T., Wallin, A., Ailisto, H., Huomo, T.: Touch-based user interface for elderly users. in *Proceedings of MobileCHI, ACM Press* (2007)
- Isomursu, M., Tähti, M., Väinämö, S. and Kuutti, K.: Experimental Evaluation of Five Methods for Collecting Emotions in Field Settings with Mobile Applications. *International Journal of Human Computer Studies*. Elsevier. Volume 65 (Issue 4), 404 – 418 (2007)
- Morwitz, V., Johnson, E., Schmittlein, D.: Does measuring intent change behavior. *Journal of consumer research*, 20, 1 453-469 (1993)
- Robinson, M. Clore, G.: Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological bulletin*, 128 (6), 934-960 (2002)
- Schultz, R., Hanusa, B.: Long-term effects of control and predictability-enhancing interventions: Findings and ethical issues. *Journal of personality and social psychology*, 36 (1978)