
User experience evaluation with experimental pilots

Minna Isomursu
VTT
Kaitoväylä 1
90571 OULU, FINLAND
minna.isomursu@vtt.fi

Abstract

This position paper analyzes learnings from several experimental pilots, where the aim has been to evaluate user experience related to adopting new, innovative technology into the everyday lives of pilot users. The goal is to examine how experimental pilots can be used in evaluating user experience. 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 analyzing and understanding it.

Keywords

user experience, piloting

ACM Classification Keywords

H5.2. User interfaces

Introduction

This paper discusses evaluating user experience in the context of experimental pilots. An experimental pilot is used for evaluating new technological solution in a pilot setting with real users in real use environment. User experience evaluation can be done in this context for

predicting what kind of user experience can be expected in real-world usage setting. Experimental pilots aim at exposing the technology to real use under circumstances that can be observed and followed. Development and adoption of technology has always had profound effects into our society and lives. These effects have mostly been random, and even accidental [12]. Experimental pilots aim at making these effects visible before technology is taken into broad use. Arranging an experimental pilot requires that the technology under evaluation is mature enough that it can be used by real users in real usage environments. However, the technology can be part of infrastructure or process that does not exist yet. The non-existing parts can be simulated during the pilot. These can be, for example, payment systems or legislation.

User experience evaluation

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. 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 (e.g. [13]), or describe them reliably after time has passed. Also, our ability to predict our own experiences in a hypothetical or future setting is very limited [10][6]. Therefore, here the hypothesis is that the most reliable understanding of user experience can be achieved by: (1) evaluating user experience in a situation as close to actual usage situation as possible for avoiding the need for users to imagine or predict their experiences in a hypothetical situation, (2)

collecting information and description of the experience at the time it happens for avoiding the need to rely on the memories of the user in describing the experience, and (3) using the direct subjective information given by the person having the experience for defining and measuring the experience. Experimental pilots aim at fulfilling these conditions. However, the facts that experimental pilots are limited in time, and may include simulated parts, may and probably will have effects on the user experience. As users are aware that the evaluated technology is part of their lives only for a limited period of time, their commitment to adopt the technology as an integral part of their lives may be weak. If the users would think that they would be stuck with the technology, they would need to create strategies to successfully integrate technology into their everyday life. If problems would arise, knowing that they would need to use the technology despite of problems would trigger a process for reducing cognitive dissonance [9]. Irrevocability of decision has been shown to be an important contributor to user experience [7][8]. Simulating parts of the infrastructure or process can be managed by two optional ways. The first option is that the users are asked to imagine that simulation would be real, and predict how their experiences would be in a real situation. The second option is that the researchers can make an assumption that the simulation is so close to real that user experiences in simulated environment are similar to those that would take place in non-simulated environment. For example, our research group has performed user experience evaluations that require mobile payments not currently available. Our observations indicate that the absence of real payment has profound effects on user experiences related to reliability and security of the technology.

Here, user experience evaluation during an experimental pilot is 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. In the following sections, examples of real pilot experiments are used to highlight methods that have been found to be suitable in certain situations. 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 [1][2].

Before-use evaluation

Before pilot use it is the best time to evaluate the attitudes and expectations the users have. 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. Before use, 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. For example, in the SmartTouch project our research group has observed how touch-based user interface can be adopted to support meal ordering by elderly home care customers [3]. Observation of in-home training situation before actual use provided an excellent opportunity to get understanding on e.g. how the physical and cognitive constraints of elderly users affect adoption of touch-based user interface, how the pilot

users placed the physical interfaces in their homes, and what were their first emotional responses towards seeing, touching and using the application. For large scale use, questionnaires can be used for collecting information about expectations and attitudes.

During use

Collecting information about user experiences at the time they happen require data collection methods that can be applied during the use of technology. This sets high requirements for not only integrating the technology under evaluation into the everyday lives of pilot users, but also the user evaluation methods. 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” [5]. Humans are not very good at analyzing what actually caused an experience [4], 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. For example, in the elderly home-care pilot [3] we planned to use a diary method for collecting user experiences evoked by new technology. However, the users used the diary to report how satisfied they were with the meals. This was quite understandable, as from the viewpoint of the user, the ultimate goal of the application was not to provide a pleasant user experience with technology, but to provide the home care customers with better meal service, and thus increase their satisfaction with their meals. Integrating user experience collection methods

into the everyday lives of pilot users often requires mobility from the experience collection method [5]. If the technology can be used in versatile mobile usage environments, so should the evaluation method. 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. For example, the Experience clip method [11] was developed for collecting user experience data utilizing the social environment of pilot users.

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. We have collected this kind of data with interviews or questionnaires, or a combination of both. Our 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. At this point of pilot experiments, users are familiar with the technology, its limitations and possibilities. This can be exploited by combining after-use evaluation with brainstorming. The goal of

brainstorming can be to create new ideas for using the same technology for other purposes, or creating improvement ideas. Brainstorming sessions can be used for collecting user experience data, too, as improvement ideas and new scenario proposals often are loaded with user experience knowledge. For example, improvement ideas might indicate which technology features have contributed towards undesired user experiences, and new application ideas can help in understanding what kinds of skills and contexts the users feel technology serves best.

Follow-up studies

Follow-up studies are excellent 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 [13] that can be observed only after some time has elapsed after the experiment. The effects of the pilot experiment should be known not only for collecting user experience, but also for evaluating the ethical issues related to the experiment. We have mainly used interviews to collect user experience data in follow-up studies. The focus has been in observing changes in attitudes, and effects of the pilot experiment in the lives of the pilot users. Also, the follow-up study has involved participatory elements, as the users are asked if any ideas for improved or new uses of technology have emerged after the experiment.

Summary and discussion

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. Each experiment has used a different set of methods, which are here 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 with the everyday lives of the pilot users.

Acknowledgements

This work was done in SmartTouch project (ITEA 05024) and has been partly funded by Tekes.

Citations

- [1] Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Zeitschrift für Physik*, 43 (1927)
- [2] Morwitz, V., Johnson, E., Schmittlein, D. Does measuring intent change behavior. *Journal of consumer research*, 20, 1 (1993), 453-469.
- [3] Häikiö, J., Isomursu, M., Matinmikko, T. Wallin, A. Ailisto, H. Huomo, T. Touch-based user interface for elderly users. In *Proceedings of MobileCHI*. (2007)
- [4] Dutton, D., Aron A. Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of personality and social psychology*, 30 (1974).
- [5] Isomursu, M., Tähti, M., Väinämö, S., Kuutti, K. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *International Journal of Human-Computer Studies*, 65, 4, (2007)
- [6] Gilbert, D., Wilson, T. Miswanting: Some problems in the forecasting of future affective states. In *Feeling and Thinking: The role of affect in social cognition*. Cambridge university press. (1990).
- [7] Knox, R., Inkster, J. Postdecision dissonance at post time. *Journal of Personality and Social Psychology*, 8, (1968).
- [8] Frenkel, O., Doob, A. Post-decision dissonance at the polling booth. *Canadian journal of behavioural science*, 8, (1976).
- [9] Festinger, L. *A theory of cognitive dissonance*. Stanford university press, Stanford, CA, USA, 1957.
- [10] Wilson, T., Wheatley, T., Meyers, J., Gilbert, D., Axsom, D. Focalism: A source of durability bias in affective forecasting. *Journal of personality and social psychology*, 28 (2000)
- [11] Isomursu, M. Kuutti, K., Väinämö, S. Kuutti K. Experience clip: method for user participation an evaluation of mobile concepts. *Proceedings of Participatory Design Conference*. ACM. (2004)
- [12] Berdichevsky, D., Neunshwander, E. Towards an ethics of persuasive technology. *Communications of the ACM*. Vol. 42, No 5. (1999)
- [13] 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).
- [14] Schooler, J. Engstler-Schooler, T. Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive psychology*, 22, (1990)