

User Experience Evaluation Methods: Lessons learned from an interactive TV Case-Study

Dimitri Drouet * and Regina Bernhaupt **

(*) IRIT, ICS Group, 118, Route de Narbonne, 31062 Toulouse, France

Dimitri.Drouet@irit.fr

(**) ruwido, Köstendorferstr. 8, 5202 Neumarkt, Austria

{Regina.Bernhaupt}@ruwido.com

Abstract. Evaluating user experience (UX) is a complicated endeavour due to the multitude of existing factors, dimensions and concepts that all contribute to UX. We report lessons learned from conducting a user study that was adapted to not only evaluate usability but also several aspects of the user experience. In this study we evaluated some of the most important factors of user experience including aesthetics, emotions, meaning and value as well as naturalness. Based on these experiences we propose a set of possible improvements to enhance existing user study approaches. These improvements aim at incorporating a variety of methods to support the various aspects of user experience including all experiences before, during and after interaction with a product.

Keywords: user experience · evaluation methods · aesthetics · emotion · user-centered development process.

1 Introduction

User Experience (UX) is defined as "a person's perception and the responses resulting from the use or anticipated use of a product, system, or service." following the ISO standard [11]. McCarthy et al. [16] argue that UX is a holistic term, as the sum of a set of factors or concepts can be more than just the individual parts. Using a more industry oriented approach, user experience has to be evaluated somehow by enabling some kind of measurement or feedback, to be able to improve the experience. One way is to focus on a set of (well defined) factors or dimensions that are known to be contributing to the overall user experience. In the domain of interactive TV the following UX dimensions have been mentioned to be of importance [5]:aesthetics, emotion, meaning and value, identification/stimulation and (if the interactive TV systems support such functionality) social connectedness. Depending on what the specific interactive TV system offers in terms of interaction technique, functionality or content, these dimensions are complemented by factors like perceived quality of service

(smoothness), naturalness of the interaction technique (e.g. naturalness, eyes-free usage) or engagement.

Evaluation of user experience is still a challenging task. There is a summary of methods available at allaboutUX [1], describing methods like experiential contextual inquiry [1] that is a variation of contextual inquiry focusing on emotional aspects when performing the method instead of focusing on usability problems. Other methods like UX expert evaluation also have their origin in the evaluation of usability and have been adapted to support user experience evaluation. Other methods including questionnaires, like the AttrakDiff,[2] are applicable once a first prototype or system is available, enabling the user to interact and experience the product. The main problem of all these methods originally developed for usability evaluation is that they have to be adapted. What is important for such adaptation is the fact that user experience is not just the experience during usage but can be divided in momentary, episodic, cumulative user experience [1]. In our case we are working on the evaluation of user experience in the field of interactive TV. The usage context thus is in people's homes, especially in the living room. Thus different dimensions of user experience are evaluated for this specific context.

Our goal was to identify how standard usability studies can be adapted to include factors or (sub-) dimensions of user experience. We focused on aesthetics, emotions, meaning/value and naturalness of the interaction in this standard laboratory based user study comparing a standard remote control with a remote control providing a kind of haptic feedback with continuous input. Based on the case-study we show if and how our adaptations were helpful for the evaluation of user experience, before, during and after interacting with the interactive system. We conclude with a description of specific challenges we faced and present some lessons learned.

2 State of the Art

User experience has been defined in several ways. While the ISO definition focuses on the users perceptions it is as important to take the context into account. As Hassenzahl et al. [9] defines it, user experience is "A consequence of a user's internal state, the characteristics of the designed system and the context within which the interaction occurs.". What is important for user experience is that an experience is mainly made out of the actual experience of usage, but also includes the encounter with the system (before usage) and experience that are after the usage of the product. Figure 1 shows how UX is changing over time with periods of use and non-use, and describing that a user experience is a combination of the experience before, during and after interacting with the product, and that the cumulative UX is formed based on a series of momentary and episodic experiences.

For the evaluation of UX there are various methods available [3]. Using the classification on who is involved, we can distinguish methods that are expert-oriented (one expert, group of experts), user-oriented (one person, pairs of users, several users) and automatic methods [2]. Classifying methods by development stage or phase we can distinguish methods for the conceptual and design phase like anticipated experience

evaluation [7] or co-discovery [12]. Such methods support to design for specific experiences and enable early insights on people's experiences with such a concept. For the implementation and development phase when partly functional or functional prototypes are available user experience can be evaluated performing user studies that are combined with methods that enable the measurement of the user experience. Most of these user experience evaluation methods have in common that they stem from standard usability evaluation methods and have been adapted to incorporate user experience.

In the area of interactive TV (iTV) the overall user experience has become a distinguishing factor for the choice of the TV system or service [15]. The evaluation of interaction techniques and systems is performed in the majority of studies using combinations of interviews, questionnaires and observation. The development of specific UX evaluation methods for interactive TV systems has been sparse [5].

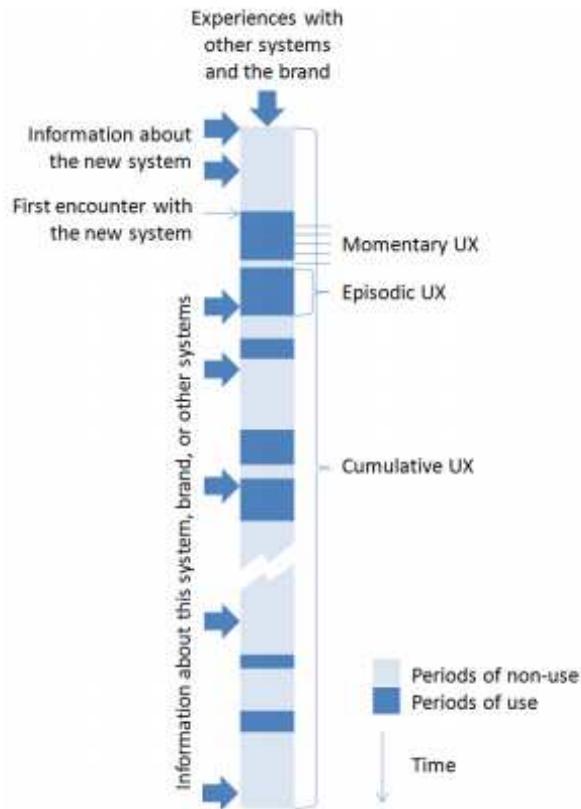


Fig. 1. The various types of user experience ranging from the first encounter with the system to long term experiences that form up the overall cumulative user experience from [1] with permission of the authors.

3 Problem Description, Method Selection and Adaptation

Focus of this work was to investigate how to enhance or adapt a standard usability study with UX measurements to be able to evaluate the UX of a newly developed interactive TV system. This iTV system supports 360 degree videos with a novel type of remote control with haptic feedback and a kind of continuous input (not a simple button press). Main focus for the set up of the method and the adaptation of the method was the need to understand to which degree such an interaction technique would enhance the user experience, compared to a standard remote. For us important was if the interaction would be perceived as natural and usage of the remote would be possible without looking at the remote (this is called eyes-free usage).

An experimental usability study in its standard form typically involves users that are performing a set of tasks with a (prototypical) system in a usability lab. Activities of users are logged using video recordings and recording events within the interactive system. Such studies typically measure in terms of usability the effectiveness (e.g. number of errors, usability problems and task success), efficiency (e.g. time necessary for performing a task) and the perceived satisfaction (e.g. interview questions). These measures are combined with usability questionnaires like the SUS questionnaire or interview questions at the end of the study.

In terms of user experience we adapted the method to include the following:

For aesthetics: taking a part of the IPTV-UX questionnaire [5] that was filled out after performing the tasks and investigating hedonic quality as dimension provided by the AttrakDiff. To evaluate emotion: Emocards after each task were used and video observation of facial expressions was conducted. To understand identification/stimulation: we used the sub-dimensions of the AttrakDiff questionnaire. To evaluate of meaning and value: interview questions. Interaction technique (naturalness, eyes-free usage): naturalness of interaction and eyes-free usage was evaluated using rating scale question. Given that the system did not provide any social communication features and was just a prototype we did not include social connectedness and service quality as UX dimensions.

For the experimental design we counterbalanced remote control order (standard remote called r97 vs new remote called r197). The evaluation was based on a fully functional user interface prototype for interactive TV and a high-fidelity remote control prototype that is close to mass-production.

Figure 2 shows how we have been adjusting the experimental usability study to also cover the various time ranges of the UX. To understand the *first encounter* with the system we video-recorded the user. The video can be used to classify user reaction when first seeing, touching and interacting with the product. During the tasks users are video recorded and eye-gaze is recorded using an eye-tracker. This allows to analyze emotional reactions and to measure objectively the level of eye-gaze towards the remote control. For the *momentary user experience* we asked each study participant after performing a task some rating-questions on the subjective experience (eyes-free perception, naturalness, emotion). The cumulative user experience is measured using the AttrakDiff [2] questionnaire. And the after usage user experience is evaluated using interview questions. With this adaptation not all UX dimensions are explored

for all types of user experience (before, during, after, momentary, episodic, cumulative). The decision to incorporate these measurements was informed by several factors: maximal duration of each session should not exceed 1,5 hours, availability of validated measures and of course the goal of the evaluation to understand UX of the newly developed interaction technique.

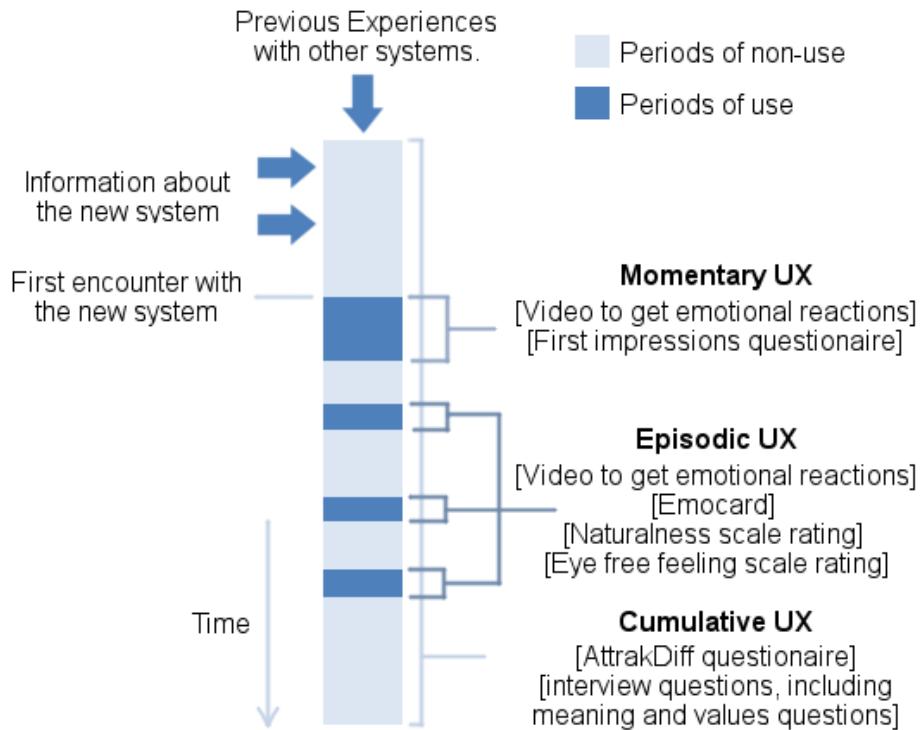


Fig. 2. Depiction of the assembly of various methods to investigate user's experiences before, during and after usage

4 Procedure and Results

The experimental user study was performed in June 2016 in an office of IRIT that was equipped with a 55" television screen. The user was seated on a sofa with about 3 meters of distance. Each session lasted around 1.5 hours. Experimentation involved two different systems, from which we only report the variation of the interaction technique when controlling 360 degree video. Ten participants (age 19 to 23; mean 21.5, SD 1.27) took part in the study and were awarded 20 € for their participation. The procedure followed closely the steps described in Figure 2.



Fig. 3. Results for the AttrakDiff questionnaire showing the traditional remote control r 97 and the new type of remote control r 197.

For the **momentary UX** the participants description included a wide range of comments that were analyzed qualitative in a word cloud, showing the difference in experience the participants had when interacting with the two different remote controls. The **episodic UX** ranged from surprising to feeling in control. The perception of naturalness was 1.65 (for the r 197) and 1.55 (for the r 97), on a scale from 1 (natural) to 5 (not natural). **Cumulative UX:** Results in terms of user experience showed that the new type of remote control r 197 was in terms of cumulative user experience perceived as desired while the traditional remote r 97 was perceived as task oriented (Figure 3). Due to limited space we are not able to report all types of data.

What is the important aspect and contribution is the understanding that the different types of UX can be contradictory and need interpretation. For example the short usage ratings for naturalness differ from the overall evaluation of the user experience. For naturalness the standard remote control was preferred, while for the overall experience measured with the AttrakDiff the r 197 was rated as more desired.

A possible interpretation is that users are facing an unpleasant situation in a user test and thus on a short time evaluation feel more comfortable with a technology there are used to (in this case the r 97). Thus in the short time they got the remote in their hand (7.15 min on average) during the test, it is hard for them to get a real feeling about a new kind of remote control (r 197). This could explain why they are considering the traditional remote control as more natural than the new type of remote despite that they are putting the new type of remote in the desire category and the traditional in the task oriented category. This demonstrates that the combination of various methods can be helpful to understand how the overall user experience develops over time.

5 Lessons Learned

To have a general understanding of the overall user experience it is important to combine methods and methodological approaches that enable the measurement before, during and after interacting with the product in a user study. The combination and combined analysis is key to get a more holistic understanding of UX.

Using a user study is a feasible method to get a reasonably fast first understanding on the overall user experience and allows to evaluate user's first impressions and first time or early usages.

Analysis of the multitude of data and their integration into a bigger picture is currently difficult to achieve. There are no standards for how to integrate differing user experience descriptions and how to conclude from them in a quick and easy way. Using textual analysis or grounded theory to interpret text could be a possibility but will be complicated facing the need to also include quantitative data.

One key limitation at the moment is the missing data on later stages of user experience and how user experience is changing over long term use. Figure 1 clearly shows longer term usage experiences, while the user study in Figure 2 only evaluates very early stages. There is work on these areas [13] on how time affects user experience, but how to integrate longer term evaluation in user studies is currently not solved.

Performing a user study per se incorporates a variety of artifacts due to the methodology [1]. Participants can feel uncomfortable in the testing situation and this might influence the feedback on the UX. Possible counter-steps can be the triangulation of methods and the incorporation of methods that can be applied at later stages (e.g. Field studies or long-term Diaries) to balance limitations of individual methods.

Conclusion

The inclusion of user experience as a central driver for software development is a difficult endeavor. This paper reports on first results on how to adapt an experimental user study to include user experience measurements for before, during and after usage of a system and discusses briefly lessons learned for such an approach. Our research goal is to start to establish a framework that allows the comparison of adaptations and combinations of UX evaluation methods, e.g. by expanding current work on the notation of UX evaluation results in a task-modeling tool [3].

A special focus of our future work will be on measuring UX after usage e.g. by performing post-usage interviews or using creative forms of reminders to prompt memory of the user to describe these after usage experiences. One possible way would be send a video about the interactive system or product to the user, combined with a set of questions to gather post-usage feedback. We feel that the phase of after usage is currently not reasonably addressed by the HCI research community, but it would be central to the understanding how users form opinions on a product due to the experience they made and how such a post usage experience leads to the establishment of a connection with a product or brand.

References

1. All About UX: All UX Evaluation Methods. <http://www.allaboutux.org/all-methods>.
2. AttrakDiff. www.attrakdiff.de
3. Bernhaupt, R. ed: Game User Experience Evaluation. Springer International Publishing, London (2015).
4. Bernhaupt, R., Palanque, P., Manciet, F. and Martinie, C.: User-Test Results Injection into Task-Based Design Process for the Assessment and Improvement of both Usability and User Experience, In proceedings of 6th International Working Conference on Human-Centred Software Engineering and 8th International Working Conference on Human Error, Safety, and System Development, August 29th-31th, 2016, Stockholm, Sweden. Springer LNCS 9856.
5. Bernhaupt, R., Pirker, M.: Evaluating User Experience for Interactive Television: Towards the Development of a Domain-Specific User Experience Questionnaire. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., and Winckler, M. (eds.) Human-Computer Interaction – INTERACT 2013. pp. 642–659. Springer Berlin Heidelberg (2013).
6. Ergonomics of Human-System Interaction Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems (2010).
7. Gegner, L. and Runonen, M. 2012. For What it is Worth: Anticipated eXperience Evaluation. 8th International Conference on Design and Emotion (London, UK, 2012).
8. Hagen, U.: Designing for player experience: How professional game developers communicate design visions. *Journal of Gaming & Virtual Worlds*. 3, 259–275 (2011).
9. Hassenzahl, M and Tractinsky, J.: User experience - a research agenda. *Behaviour & Information Technology* - 25, 2, 91-97 (2006).
10. Hassenzahl, M.: The Thing and I: Understanding the Relationship Between User and Product. In: Blythe, M.A., Overbeeke, K., Monk, A.F., and Wright, P.C. (eds.) Funology. pp. 31–42. Springer Netherlands (2003).
11. ISO 9241-210 Ergonomics of Human-System Interaction - Part 210: Human-centred design for interactive systems (2010).
12. Jordan, P. Designing Pleasurable Products. Taylor and Francis: London (2000).
13. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B.: Measuring the dynamics of remembered experience over time. *Interacting with Computers*. 22, 328–335 (2010).
14. Lazar, J., Feng, H. J. and Hочеiser, H. Research methods in human-computer interaction. John Wiley & Sons, (2010).
15. Lehtonen, T. K. (2003). The domestication of new technologies as a set of trials. *Journal of consumer culture*, 3(3), 363-385. ISO 9241-210
16. McCarthy, J., Wright, P.: Technology As Experience. *Interactions*. 11, 42–43 (2004)