Valence Method for Formative Evaluation of User Experience

Michael Burmester, Marcus Mast, Kilian Jäger, & Hendrik Homans
Stuttgart Media University
Wolframstr. 32, 70191 Stuttgart, Germany
+49 711 25706 101
burmester | mast | jaegerk | homans@hdm-stuttgart.de

ABSTRACT
This paper describes a method for formative evaluation of the user experience based on the user experience model of Hassenzahl [11]. It captures positive and negative feelings during the exploration of an interactive product. In a subsequent retrospective interview phase users indicate for each instance of a positive or negative feeling the product design aspects inducing it. This phase further employs the laddering interview technique [24] to reveal the meaning of product design aspects to the user and the underlying fulfilled or frustrated needs. The generated information helps designers to understand and optimize the user experience potential of a product.

ACM Classification Keywords
H.5.1 Multimedia Information Systems: Evaluation / methodology; H.5.2 User Interfaces: User-centered design

Keywords
User experience; Formative evaluation

INTRODUCTION

User experience research focuses on subjective factors like perceptions and emotional responses of users [12,15,18]. To assess and optimize the potential of products to cause a good user experience, the industry demands user experience evaluation methods [19,25,29]. As emotions play a central role in user experience [12,15,22] several methods are concerned with the measurement of emotions. There are non-verbal self-report instruments like the Product Emotion Measurement Instrument (PrEmo) [7], the Layered Emotion Measurement Tool (LEMtool) [17] or Emotion Sampling Device (ESD) [16]. Objective emotion measurement can be achieved by automated recognition of facial expressions [21]. Other instruments focus on factors influencing user experience, e.g. the perceived hedonic and pragmatic product qualities [AttrakDiff2 questionnaire, 13].

However, the HCI community also has a strong need to understand the effects of design during iterative development [2,29]. The above mentioned user experience measurement instruments are not suited for providing such information. Some approaches have focused on reflection of the design based e.g. on cultural commentators like journalists [9] or involving users into a reflective process through interviews and logbooks [26]. The repertory grid technique is often used to obtain insights into the dimensions of product meaning [e.g. 8]. It requires design alternatives which are not always available under industrial conditions. The latter methods collect feedback from users or other commentators on a broad basis. User experience is one quality among others. The user experience is not captured during the interaction with the product and the focus is not on specific product design aspects. Product design aspects can be classified as: presentation (e.g. color, form), interaction (e.g. controls, navigation bar), functionality (e.g. search possibilities, storing links), and content (e.g. text, video, photo, illustrations) [3].

To optimize design for better user experience it is important for designers to understand which design aspects have caused a positive or negative experience and why. We subsequently describe a method termed "valence method" which has been developed (1) to capture the user experience during the interaction in order not to lose any details, (2) to elicit product design aspects causing good and bad user experience, and (3) to obtain information on the underlying psychological reasons.

THEORETICAL BACKGROUND

The main theoretical foundation of the valence method is Marc Hassenzahl's user experience model [11,12]. Hassenzahl distinguishes between be-goals, do-goals, and motor-goals. Be-goals are closely related to the self and represent peoples' motifs, e.g. feeling bored and having a need for stimulation. As soon as a motif is present, do-goals are generated which are instrumental for fulfilling the motif, e.g. to get stimulation users might have a look at their favorite news portal to see if interesting news are available. Motor-goals are necessary to achieve the do-goals, e.g. open a web browser and select the news portal from the bookmark list. The be-goals can be described as human needs. They are the ends to the question about the reason behind behavior. Several researchers have provided lists of basic human needs [e.g. 23,27]. Sheldon et al. [27] list the following needs: self-esteem, autonomy, competence, relatedness, pleasure-stimulation, physical striving, self-actualizing-meaning, security, popularity-influence and money-luxury. Users have need-induced experiences, e.g. a relatedness experience. If a product design aspect satisfies the need, people feel good and if the needs are frustrated, they feel bad [11,12]. This notion is in conformance with the means-ends theory of Gutman [10] proposing that product design aspects can have consequences related to an underlying value. People are able to report the quality of their evaluative feeling (feeling "good" or "bad") at every moment [11,12] and they can link it to a product design aspect [10]. Reporting the personal meaning and the underlying needs is much more difficult. Reynolds and Gutman [24] developed the laddering technique for this purpose. It reveals the personal meaning or consequence of product design aspects for the user as well as the underlying value or need. The central question of the laddering technique is to persistently ask "why" a certain product design aspect is important for the person until the level of needs has been reached. Abeele and Zaman [1] applied the laddering method to human-computer interaction. After system usage they asked users to recall important product design aspects and subsequently started the "why" questions.

VALENCE METHOD

Overview

The method consists of two phases: In an exploration phase, users experience positive or negative feelings while using a product. They monitor these feelings by setting positive or negative valence markers. The activity itself is similar to the plus-minus method [4,5] but it serves a different purpose and is used with a different instruction. The plus-minus method evaluates the comprehensibility of text. The second phase is a retrospective interview where each valence marker is discussed in order to identify the product design aspect causing the experience, its meaning for the user, and the underlying needs.

Exploration phase

In the exploration phase users get the open instruction to explore the product to be tested in whichever way suits them best. No tasks need to be accomplished. If users were given concrete tasks they would be set to a goal mode [14], i.e. they would strive to reach an extrinsic goal. Instead, users should be in an activity mode [14] where they follow their needs and develop their own intrinsic goals. As a secondary instruction, users are asked to pay attention to their feelings during the exploration. They are asked to continuously monitor all instances of feeling good and bad by pressing a dedicated button for positive (e.g. a green plus sign) and negative (e.g. a red minus sign) feelings on a remote control. In the instruction, emphasis is placed on monitoring even the slightest changes of feeling and to rather press once too often than to be reluctant. The button presses are recorded as valence markers with time stamps in a video recording of the exploration phase. The exploration phase should be kept short because otherwise memory will decrease in the retrospective phase. In our tests, 6 to 8 minutes yielded good results.

In case not all product design aspects relevant to designers were explored by a participant, the participant can be explicitly asked to explore these at the end of the exploration phase.

Retrospective interview phase

In the second phase two aspects have to be investigated for each valence marker: (1) Which product design aspect caused setting the valence marker? (2) What are the related needs? Users watch a video recording of their exploration session augmented by the valence markers presented on a timeline of the video recording. The instruction is to watch the recording, comment on what they were experiencing, and pause at each valence marker. The interviewer then asks what product design aspects caused setting the current valence marker. Here it is first necessary to identify all relevant design aspects. For example, a user commenting a positive valence marker might say "This picture here is so nice. I like it." At this point the design element has been identified. The next step would be to determine the product design aspects of the element. The interviewer could ask: "What is so nice about that picture?" Participant: "It is the wide ocean." Now, the specific aspect of the picture has been identified.

After determining the specific product design aspect, the interviewer determines the meaning and the underlying needs by applying the laddering technique [24]. The central question of laddering is: "Why is this attribute positive or negative?". This type of question is repeated until the need related to the valence marker has been identified. Following the previous example, the question would be "Why is the wide ocean so positive for you?" The answer might be "It looks like a departure to unknown shores" which reveals the meaning of the product design aspect. The next question could be "Why is it so positive to depart to new shores?" "It makes me curious; it is about exploring new things". At this point it becomes clear that the need related to the positive feeling caused by the ocean in the picture was stimulation. As soon as the need has been identified, the user or interviewer resumes playback of the recording until the next valence marker.

Data analysis

The outcome of the retrospective phase is a video recording with synchronized valence markers. From the video recording, a verbal protocol is produced. For each marker, the following attributes are coded: participant number, marker sequence number, time stamp, positive or negative valence, mentioned product design aspects, meaning for the user, and underlying needs. It is convenient to use a single table containing all participants with one row per valence marker and a column for each of the above mentioned marker attributes. In the next step, the terminology used for product design aspects and needs is harmonized throughout the dataset. This enables sorting and filtering. For example, the top ten product design aspects can be extracted. These can be assumed to have a major influence on the user experience. When optimizing design, the meaning of a valence marker shows how people perceive the product design aspect. The meaning in the previous example was that the picture of the ocean landscape stands for "departure to unknown shores" and that it makes the person curious to explore new things. By interpreting the meaning, the underlying need for stimulation can be identified.

A user experience metric can be generated from valence marker frequencies. For each user the number of negative markers is subtracted from the number of positive markers. The result is divided by the sum of all markers. The result is an indicator between -1 (bad experience) and +1 (good experience) for the overall user experience caused by the product. For calculation of the metric a sufficient number of markers is needed. Validity of the metric is subject to the accuracy of the assumption that the overall user experience is a function of frequency, not intensity of feeling [cf. 6].

EXEMPLARY VALENCE METHOD STUDY

Study design

During the development of the valence method several pilot studies with a total of 24 users were carried out to test method candidates and to improve and fine-tune method components and procedures. After those pilot studies we carried out a full valence method study with 10 participants, which is reported here. The test object was a promotion website for an IPTV provider. The exploration phase with duration of 6 to 8 minutes was recorded with the software Morae [28] and a Wiimote [20] was used for setting valence markers.

RESULTS

Descriptive statistics

In this study the 10 participants set 236 valence markers overall (119 positive and 117 negative). Typical product design aspects marked were pictures, animations, text passages, videos, media controls, navigation, color concepts, functions of IPTV like time-shifted recording, or media content like sports. In the 236 valence markers, 80 product design aspects were stated on 27 different webpages. Overall 197 instances of 9 needs were determined to underlie the 80 product design aspects. Stimulation (63) and competence (62) were assigned most often, followed by relatedness (29), self-esteem (19), and autonomy (11). Other needs (popularity, keeping the meaningful, competition, security) were determined to underlie product design aspects less than 5 times each. The mean value of the UX metric was .01 (SD = .26). This indicates a neutral user experience throughout participants.

Qualitative analysis

For all product design aspects plenty of qualitative data about the meaning was generated. In combination with the underlying needs this information is well suited for understanding the perception of design and for generating ideas for design optimization. To give an impression of the qualitative data, we will describe two product design aspects.

The first example is a picture of a couple lying arm in arm on a sofa watching TV. The picture was an element on a webpage explaining the functions of the IPTV product. The picture was the cause for 6 participants to set a valence marker (5 positive and 1 negative). Participants described the situation as “nice to see”, “calm and cozy”, or “nice cuddly”. A representative statement about meaning for participants was “I can identify with them and feel well cared of, so they must have good services too”. This person described a self-esteem experience. She felt that the company cares about her and that she is important. Another mentioned meaning was “They are watching TV together and look happy. They found something to watch together (...). This is possible because of the huge variety of stations.” The large number of stations provides the opportunity to find a joint TV show, so they can have a relatedness experience. The person anticipates this.

The second example is a rather simple illustration of TV stations with logos presented in a table-like grid. The illustration is employed on several webpages as “station overview” and has an animated intro. It was the cause for setting 4 positive and 2 negative valence markers by 5 participants. The positive valence markers were commented for example as “I can get a quick overview and it is easy to decide whether this is the right product” or “I can see my favorite stations that I already know”. These and similar comments can be interpreted as a competence experience because the person feels capable and effective in his decisions. Another type of personal meaning is the following: “This looks like a huge offer. I’m sure that I can find something for me. I like to have the possibility to decide what I really like. I can decide on my own. It is a feeling of freedom.” Here the interpretation would be that the person has an autonomy experience because he can choose programs exactly the way he wants, independent of external influences.

One of the two negative valence markers was described as “I search for quality in their offer. This is too much. It is not possible that all the stations are interesting. If they list 30 stations as highlights something is not right.”. We interpret this as a frustrated stimulation need because of his fear that the stations may not be interesting but boring.

CONCLUSION

The exploration phase of the valence method provides insight into the product design aspects causing positive or negative feelings. The retrospective interview phase generates a wealth of qualitative data on meaning related to the product design aspects and the underlying human needs. The subjective meanings in particular reveal many facets of the user experience. They enable a better understanding of the design and can serve as the starting point for design improvements.

A current limitation of the method is that participants should use the product or prototype for the first time during the evaluation. In our tests, extensive previous usage resulted in a strong decrease of valence markers set. A related method may be required which employs a retrospective interview on experiences over longer usage periods. The valence method has so far been used in lab settings only. To increase ecological validity, it should be enhanced so that the valence markers can also be collected in authentic usage situations in the field. It is further planned to develop guidelines for interpreting needs and then to study the reliability of the qualitative analysis. Another step will be to study the usefulness of the results for designers.

REFERENCES

1. Abeele, V.V. & Zaman, B.: Laddering the User Experience! workshop paper (*UXEM09 at Interact2009*), publ. online, available at wiki.research. nokia.com/index.php/UXEM09/Papers, 2009.
2. Ardito, C., Buono, P., Costabile, M.F., De Angeli, A. & Lanzilotti, R.: Combining Quantitative and Qualitative Data for Measuring User Experience of an Educational Game, in *E.L.-C. Law, N. Bevan, G. Christou, M. Springett & M. Lárusdóttir (eds.) Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, Institute of Research in Informatics of Toulouse, 27-31, 2008.
3. Burmester, M. & Dufner, A.: Designing the stimulation aspect of hedonic quality - an exploratory study. In M. Pivec (Ed.), *Affective and Emotional Aspects of Human-Computer Interaction: Emphasis on Game-Based and Innovative Learning Approaches,* IOS Press, 2006.
4. De Jong, M.: *Reader Feedback in Text Design: Validity of the Plus-minus Method for the Pretesting of Public Information Brochures,* Rodopi B.V. Editions, 1998.
5. De Jong, M.: Dynamics of Iterative Reader Feedback, *Journal of Business and Technical Communication*, Vol. 20, No. 2, 159-176, 2006.
6. Diener, E., Sandvik, E. & Pavot, W.: Happiness is the Frequency, Not the Intensity, of Positive Versus Negative Affect, in E. Diener (ed.), Assessing Well-Being. Springer, 213-231, 2009.
7. Desmet, P. M. A.: Measuring Emotions, in *Blythe, M., Overbeeke, C. J., Monk, A.F. & Wright, P.C. (Eds.), Funology. From Usability to Enjoyment,* Human-Computer Interaction Series, 3, Kluwer Academic Publishers, 111-123, 2003.
8. Fallman, D. & Waterworth, J.: Dealing with User Experience and Affective Evaluation in HCI Design: A Repertory Grid Approach, in *Proceedings of CHI 2005, Conference on Human Factors in Computing Systems*, April 2-7, Portland, Oregon. 2005.
9. Gaver, W.: Cultural commentators: Non-native interpretations as resources for polyphonic assessment, *International Journal of Human-Computer Studies archive, Volume 65, Issue 4*, 2007, 292-305.
10. Gutman, J.: A Means-End Chain Model Based on Consumer Categorization Processes. *Journal of Marketing 46 (2),* 1982, 60-72.
11. Hassenzahl, M.: User Experience (UX): Towards an experiential perspective on product quality, in Proc. of IHM'08, 2-5 Sept. 2008, Metz, France, 11-15, 2008.
12. Hassenzahl, M.: *Experience Design – Technology for all the right reasons.* Morgan & Claypool Publ., 2010.
13. Hassenzahl, M., Burmester,M. & Koller, F.: Der User Experience auf der Spur: Zum Einsatz von www.attrakdiff.de, in *H. Brau, S. Diefenbach, M. Hassenzahl, F. Koller, M. Peissner & K. Röse (eds.), Usability Professionals 2008,* IRB, 78-82, 2008.
14. Hassenzahl, M., Kekez, R. & Burmester, M.: The importance of a software's pragmatic quality depends on usage modes, in *H. Luczak, A. E. Cakir, & G. Cakir (eds.), Proceedings of WWDU 2002,* 275-276, 2002.
15. Hassenzahl, M. & Tractinsky, N.: User Experience – a research agenda. *Behaviour & Information Technology, Vol. 25, No. 2, March-April*, 2006, 91-97.
16. Hole, L. & Williams, O.M.: The Emotion Sampling Device (ESD), in *D. Ramduny-Ellis & D. Rachovides (Eds.), Proceedings of the 21st BCS HCI Group Conference, HCI 2007, 3-7 September 2007,* British Computer Society, 177-178.
17. Huisman, G. & Van Hout, M.: The Development of a Graphical Emotion Measurement Instrument Using Caricatured Expressions: the LEMtool, in *C. Peter, E. Crane, H. Agius & L. Axelrod (eds), Emotions in HCI – Designing for People. Proceedings of the 2008 International Workshop.* Fraunhofer, 2008.
18. ISO 9241-210: Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. Beuth, 2010.
19. Kaye, J., Boehner, K., Laaksolahti, J. & Ståhl, A.: Evaluating Experience-Focused HCI, in *Proceedings of CHI 2007*, April 28-May 3, ACM, 2117-2120.
20. Nintendo. *Wii Remote*. Available at www.nintendo.com/ wii/what/controllers, 2010.
21. Noldus *FaceReader*. Available at www.noldus.com/ human-behavior-research/products/facereader, 2010.
22. Norman, D.: *Emotional Design*. Basic Books, 2004.
23. Reiss, S. & Haverkamp, S.M.: Toward a comprehensive assessment of fundamental motivation: Factor structure of the Reiss Profiles. *Psychol. Assess.,10,* 97-106, 1998.
24. Reynolds, T.J. & Gutman, J.: Laddering Theory, Method, Analysis, and Interpretation. *Journal of Advertising Research, 28,* 11-31, 1988.
25. Roto, V., Ketola, P. & Huotari, S.: User Experience Evaluation in Nokia, *in CHI'08 Workshops, 3961-3964.*
26. Sengers, P., Boehner, K., Shay, D. & Kaye, J.: Refective Design, in *Proceedings of Århus'05* 8/21-8/25/05 Århus, Denmark, ACM, 49-58.
27. Sheldon, K.M., Elliot, A.J., Kim, Y & Kasser, T.: What is satisfying about satisfying events? Testing 10 candidate psychological needs, *Journal of Personality and Social Psychology 80[2]*, 2001, 325-339.
28. TechSmith. *Morae*. Available at www.techsmith.com/ morae.asp, 2010.
29. Väänänen-Vainio-Mattila, K., Roto, V. & Hassenzahl, M.: Towards Practical User Experience Evaluation Methods, in *E.L.-C. Law, N. Bevan, G. Christou, M. Springett & M. Lárusdóttir (eds.) Meaningful Measures: Valid Useful User Experience Measurement (VUUM),* Institute of Research in Informatics of Toulouse, 19-22, 2008.