Dynamics of User Experience

Michael Minge
Technische Universität Berlin, GRK prometei
Franklinstraße 28-29, Sekretariat FR 2-6
D-10587 Berlin
michael.minge@zmms.tu-berlin.de

ABSTRACT
This paper addresses the consideration of changes in user experience over time. We define user experience as a multidimensional phenomenon, that consists both of the perception of different product qualities and related emotions. First we will introduce the reader to our conceptualization of user experience and report empirical findings with respect to dynamic interactions. Secondly, we describe an experiment that we have conducted to investigate systematic dynamics in the perception of a system's attractiveness, its apparent usability and emotional user reactions. For that, 60 participants were exposed to one simulation of a portable digital audio player that was manipulated with respect to the two factors attractiveness (low, high) and usability (low, high). The dependant variables, attractiveness, usability, and subjective emotional feelings were collected three times: before usage, after free interaction and after task conducted interaction. Finally, we summarize our results and thereby emphasize the practical relevance of dynamics of user experience.

Author Keywords
User experience, dynamics, aesthetics, usability, emotions.

ACM Classification Keywords
H5.2. [User Interfaces]: Evaluation/methodology

INTRODUCTION

In competitive markets that offer consumer products with a nearly interchangeable range of functionality, technical devices have to fulfill a variety of requirements to appeal to consumers. Evidently, it is not only goal conduciveness, i.e. effectivity and efficiency, that is sufficient to ensure market success. The aim of measuring the user experience is to consider a more holistic user centered approach, that takes also aspects into account which go beyond usability.

One approach for defining the concept user experience is to characterize specific dimensions that are important aspects in the experience of technology. For this purpose, Hassenzahl [2] distinguishes two dimensions of product qualities, namely the perception of instrumental (or: 'pragmatic") and non-instrumental (or: 'hedonic") qualities. Whereas the first refer to performance and pure usability aspects, the latter summarize system properties which refer to beauty, visual aesthetics, identification and stimulation. The importance of those aspects is motivated by their immediate perceptibility: While usability evaluation depends basically on interaction with the product, the attributes that enable hedonic judgments are immanent in the product appearance itself.

A third important aspect of user experience are emotional user reactions. For example, in his hierarchic model, Jordan [3] distinguishes several types of pleasure with a product, whereby he insists on high functionality and high usability as necessary preconditions.

An empirical based junction of those dimensions, has realized Mahlke [8] by developing a component model of user experience (the so-called CUE-Model, see Figure 1). Interaction characteristics are summarized by system properties, user characteristics and variables referring to the situational context. Those characteristics are motivators for the process of user experience. According to Mahlke [8] it is only the perception of product qualities (both instrumental and non-instrumental) that is directly influenced by the interaction characteristics. Whereas, emotional user reactions are supposed to be emerged by those perceptions. All three components finally lead to an appraisal of the system. This appraisal determines overall judgements and the following usage behavior.

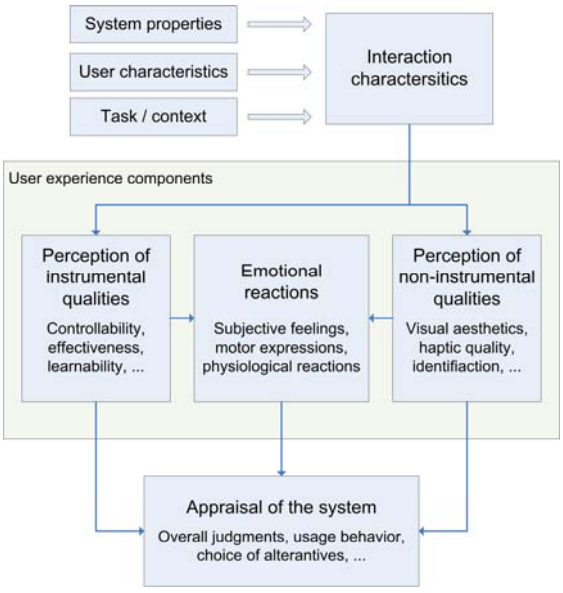

Figure 1. Components of user experience (CUE-Model) [8]

Less considered aspects of this model are the dynamics of these three components: How does instrumental qualities and non-instrumental qualities go together? Which role play the emotions? How does the perceptions change over time?

One typical question in Human-Computer Interaction (HCI) is how attractiveness as a hedonic product quality is related to perceived usability. Although Mahlke [8] postulates the pragmatic and hedonic product qualities to be independent, most findings indicate, that judgments of attractiveness and usability are strongly positive correlated [11]. Basically, it is assumed, that the underlying psychological mechanism is the so-called Halo-Effect: People tend to use apparent qualities of a stimulus, like hedonic qualities, to estimate non evident attributes, like pragmatic aspects. Nevertheless, there are studies that do not support evidence for those strong positive correlations [10], actually a few authors even found negative ones [9].

Other research focus on dynamics in the judgment of aesthetics. Carbon & Leder [1] for example point out the lack of standardized methods for measuring attractiveness. Thereafter, the perceived innovativeness of a product has a dynamic influence on the judgment of attractiveness: A high innovative design first leads to a lower judgment of attractiveness, but after familiarization to a higher one. The low innovative design is more attractive on the first sight, but less attractive after repeated evaluation [1]. Therefore, typical single shot tests would lead to results that do not refer to experiences with the product in everyday life. Anything is known about dynamics in the judgment of innovativeness over time with respect to different levels of attractiveness.

The less considered aspect in HCI are relationships between emotional user reactions and the judgment of product qualities. Only Lindgard and Dudek presume usability and affective aspects of user satisfaction to be independent [7].

A common methodological problem is the lack of experimental design investigating relationships between different aspects of user experience. Correlation studies do not reveal causal relationships between the components of user experience. Moreover, we assume that only an experimental design that additionally focuses on dynamic changes in judgments and emotional reactions is the most promising way to investigate the interplay between the components. Such approach would also be appropriate to measure consistencies in judgments over time.

From the practitioner's perspective, those systematic interactions between judgments and emotions are of the greatest import. Measuring the user experience has to regard those changes in order to reflect typical user experiences with the product in everyday life.

According to the famous study of Tractinsky, et al. [12] we conducted and experiment to investigate time-related dynamics between attractiveness, usability and emotions.

METHOD

The aim of our between-subject experimental design is to investigate dynamic relationships in apparent usability (as 'pragmatic' quality), perceived attractiveness (as 'hedonic' quality), and emotional user reactions in HCI. Therefore we collected data at three points in time and combined single-item with multiple-item measurement scales.

Stimuli and apparatus

In our experiment we used four versions of a computer-based simulation of portable digital audio players, which were introduced and validated by Mahlke [8]. These simulations varied systematically in 'attractiveness' and 'usability'. Both factors included two levels: 'low' and 'high'. They were fully combined cross over.

The attractiveness factor was manipulated by the layout's visual aesthetics. The high attractive version consisted of a curved symmetric body design, which had a color combination of blue and grey (low color differences). The low attractive version was angular shaped and colored blue and green (high color differences). Its display and its key panel were positioned asymmetrically (see Figure 2).

Figure 2: Screenshots of the high attractive (left) and the low attractive (right) simulation of digital audio players.

The usability factor was manipulated by introducing difficulties to the interaction between the audio player and the participants in the low usability condition. These difficulties emerged from three system features: the number of simultaneously discernible menu lines (five versus two), a scrollbar as indicator for available but hidden menu items (given or not), and a cue about the actual position in the menu hierarchy (given or not, see Figure 3).

Figure 3: Screenshots of the high usable (left) and the low usable simulation (right) of digital audio players.

The audio-players were presented as prototypes on a 7" TFT with touch screen functionality, which participants used for providing input. The display was connected to a computer, which ran the simulation of the audio player.

Participants

We recruited 60 volunteers (24 women, 36 men) from the local university. The mean age was 27.4 years (*SD* = 7.21). While 54 participants owned a portable digital audio player, the remaining 6 ones had no experience with such devices.

Design

By the two independent variables 'attractiveness' and 'usability', which consisted of two levels on their part ('high' and 'low'), four conditions of portable digital audio players could be tested: (a) 'high usability' and 'high attractiveness', (b) 'high usability' and 'low attractiveness', (c) 'low usability' and 'high attractiveness', (d) 'low usability' and 'low attractiveness'. Each participant was randomly assigned to one of these conditions.

The experimental design is complemented by a third factor, namely the point in time when participants had to rate the product qualities of the portable digital audio player and their subjective feelings. Since we created three repeated measurement points, this within subject factor has three levels: (a) T1, (b) T2, and (c) T3.

Dependant Variables

As single-item scales we measured perceived 'usability' perceived 'attractiveness', and perceived 'innovativeness' at all three levels of the repeated time factor (T1, T2, and T3). Scales were 7-point Likert scaled. Moreover, participants reported their subjective feelings at all three measurement points on the dimensions 'arousal' and 'valence' arising from the Self-Assessment-Manikin (SAM) by Lang [4] (each 9-point scaled).

At T2 and T3 we employed questionnaires to assess the user's perception of attractiveness and usability in more detail. Selected sub-dimensions of the Subjective Usability Measurement Inventory (SUMI) [3] served to rate usability aspects (controllability, effectiveness, helpfulness, learnability). One dimension of a questionnaire developed by Lavie and Tractinsky [5] was used to measure attractiveness and visual aesthetics. In order to minimize response effects in repeated measurements we randomized the order of the items within the two questionnaires.

Procedure

Participants were first explained the procedure and the applied rating scales. The experiment started with the presentation of the randomly assigned simulation on the TFT. Participants were instructed to rate the portable digital audio player on the three single-item scales: (a) 'attractiveness', (b) 'usability', and (c) 'innovativeness'. Moreover, participants had to assess their subjective feelings by filling in the SAM scales. This procedure constitutes measurement point T1, which served as a pure pre-experimental evaluation.

Afterwards, participants were explained the handling with the audio player. They then practiced it within two minutes without a specific given task. Again they had to rate the single items and the SAM scales. Furthermore they filled in the SUMI and the attractiveness questionnaire (T2).

Finally, participants had exactly 15 minutes to solve as much given tasks as possible with the audio player. For this task conducted sequence we developed a set of 28 tasks that are typical of interacting with digital audio players, e.g. searching a track, organizing contact information, and regulating volume. The content and the order of the tasks were identical for all participants. Lastly, as post-experimental rating T3, participants rated the same items and questionnaires as in T2.

The experiment lasted half an hour on average and participants were paid 5 Euros.

RESULTS

Manipulation check

Pre-experimental means of the attractiveness' judgments indicate that the high attractive designed shape versions were rated as more attractive [F (1,58) = 28.03, $p < .001$].

The success of the usability manipulation was evaluated by comparing the amount of completed tasks within 15 minutes of task conducted interaction (between T2 and T3). Compared to the system of low usability, the high usable system led to a greater percentage of correct solutions [F (1,58) = 69.19, $p < .001$]. Mean rates of the perceived usability were compared in the post-experimental evaluation phase (T3), indicating that the usability factor has a significant effect on the perception of usability as well [F (1,58) = 33.9, $p < .001$].

Judgments of attractiveness and usability

As predicted by the results of Tractinsky et al. [11], judgments of attractiveness and usability are highly correlated at each point in time. Moreover, the correlations even increased between T1 (r = 0.37) and T2 (r = 0.53), and still remained significant at T3 (r = 0.46).

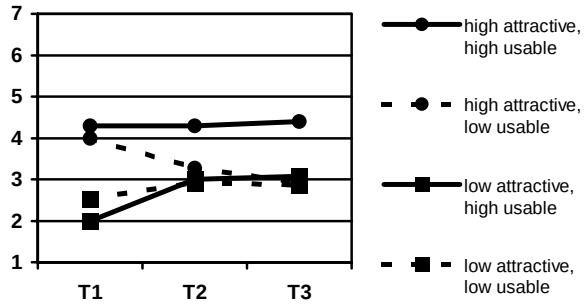

Figure 4. Mean rates of attractiveness (on a 1-7 scale) under three levels of points in time.

Dynamics of the attractiveness' judgments are presented in Figure 4. The overall interaction between the factors attractiveness and measurement point is highly significant [F (6,110) = 3.099, $p < .01$]. As the manipulation check has already indicated, the high attractive versions are both valuated as more attractive before the experiment has

started. But this differences resolve between T1 and T2: The attractive audio player with the lower usability decreased in appreciation. A single independent *t*-test indicates that this decrease is nearly significant [$t(29) = 1.884$, $p = .07$]. Although we found an increase in appreciation for the low attractive devices, the main effect of point in time is not significant with respect to the judgment of attractiveness [$F(2,54) = 0.019$, $p = .981$, n.s.].

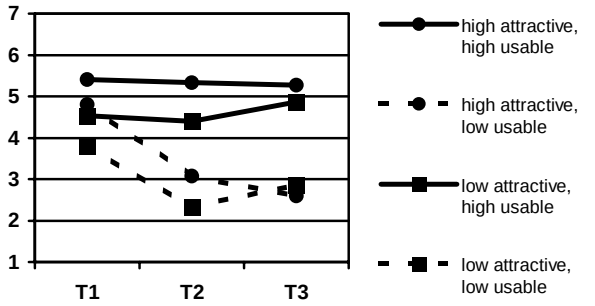

Figure 5. Mean rates of usability (on a 1-7 scale) under three levels of points in time.

Dynamics in the perception of usability displays Figure 5. The overall interaction between the factors usability and measurement point is again highly significant [$F(6,110) = 4.148$, $p < .001$], whereas the repeated factor time in point had no significant general influence on the judgment of usability [$F(2,110) = 1.203$, $p = .304$, n.s.].

We analyzed differences in the usability ratings that are caused by the level of attractiveness in visual aesthetics by three single independent *t*-tests. At T1 the usability ratings were significantly higher for the two attractive versions [$t(58) = 2.547$, $p < .05$]. That means, the high attractiveness of the portable digital audio players overshadowed the perception of usability. But, both at T2 and T3, the mean rates of usability did not longer differ between the two attractiveness levels [$t(58) = 1.844$, $p = .07$, n.s. for T2 and $t(58) = 0.132$, $p = .895$, n.s. for T3]. This indicates that the usability ratings become independent from the attractiveness judgments over time.

Judgments of innovativeness

In sum, there were no significant changes in the judgments of innovativeness over time [$F(2,54) = 1.056$, $p = .355$, n.s.]. But we found the dynamic ratings significantly influenced by the version of the player [$F(6,110) = 2.337$, $p < .05$]. Interestingly, the ratings of innovativeness are only influenced by the attractiveness factor [$F(2,110) = 6.175$, $p < .01$] and not by the usability factor [$F(2,54) = 0.821$, $p = .433$, n.s.]. Figure 6 presents the judgments of innovativeness over time under the levels of attractiveness.

At pre-experimental T1 it is the more attractive version that evokes a higher mean rate of innovativeness. However, the participants who were exposed to the attractive version valued the system as less and less innovative over time. On the other hand for the two lower attractive versions there is an increase in perceived innovativeness. According to Carbon & Leder [1] these results are complied with a differential mere-exposure effect, but with respect to the innovativeness as independent variable.

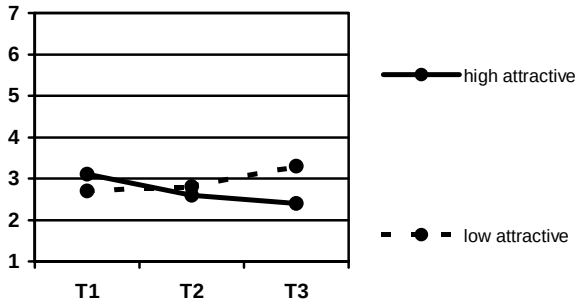

Figure 6. Mean rates of innovativeness (on a 1-7 scale) under three levels of points in time.

Multiple-item questionnaires

We used the multiple-item questionnaires at T2 and T3 to analyse the variance of the single items in more detail. Therefore we calculated a regression analyses to predict the judgments of usability by the abridged SUMI questionnaire [3] and to predict the judgments of attractiveness by the items of the questionnaire measuring visual aesthetics [5].

Both at T2 and T3 the abridged SUMI explained an essential part of variance in the usability ratings ($R^2 = .706$ at T2 and $R^2 = .832$ at T3). Thereby at T2, items that have a significant impact on usability are typical task related items, like 'Tasks can be performed in a straightforward manner using this system' ($p < .05$) and 'It is easy to make the system do exactly what you want' ($p < .05$). Whereas at T3 the only items that have a significant impact on usability are 'There have been times in using this system when I have felt quite tense' ($p < .05$) and 'I think this system is a good product' ($p < .05$), which already reflect a more summarized judgement of the interaction.

Also for the judgments of attractiveness we collect more data at T2 and T3 for specifying the single items. In sum, the determination was again high ($R^2 = .823$ at T2 and $R^2 = .770$ at T3). At both points in time the item 'aesthetic' has a significant impact on the judgment of attractiveness. Moreover at T2 it is also the item 'novel' that significantly determine the judgments of attractiveness ($p < .01$), whereas at T3 it is the item 'fascinating' ($p < .01$) and 'pure' ($p < .05$) that have an impact on attractiveness in addition to the item 'aesthetic'.

Subjective feelings

During interaction with all portable digital audio players we found a decrease of perceived 'arousal' [$F(2,56) = 9.827$, $p < .01$]. Accordingly, the mean rates of 'valence' became more positive regardless of the specific version the participants were exposed to [$F(2,56) = 5.769$, $p < .01$].

With respect to the two independent variables we identified both the 'valence' and the 'arousal' influenced by the intended usability [$F(2,114) = 5.170$, $p < .01$ for 'valence'

and F (2,114) = 4.520, $p < .05$ for 'arousal']. The higher the usability is, the more positive is the 'valence' and the lower is the 'arousal', whereas flaws in usability are related to a more negative 'valence' and higher values of 'arousal'.

There is neither a significant interaction between the system's attractiveness and the ratings of 'valence' [F (2,114) = 0.920, p = .401] nor between attractiveness and 'arousal' [F (2,114) = 0.258, p = .773].

DISCUSSION

First, our results support findings of overshadowing effects of 'hedonic' aspects, i.e. attractiveness, over 'pragmatic' aspects, i.e. usability. We interpret this as an Halo-effect: People use apparent qualities of a product to estimate non evident attributes, i.e. usability. Furthermore, we could show that this overshadowing decreases over time, so that participants value the usability quite independently after a relative short interval of interaction.

With respect to the ratings of attractiveness, we obtained a tendency of increasing appreciation over time, especially for the low attractive simulations. We explain this result by mere-exposure, or familiarization effect: People tend to like those objects more with which they are interacting. Presumably, because of the short duration of the experimental session, these differences become not significant. Interestingly, the judgments of attractiveness were strongly negative influenced by the perceived usability in the case of low usability. This result reflects something like an overshadowing in a quite different way: Participants 'punish' the attractiveness of a technical system because of perceived flaws in usability.

Secondly, we showed that judgments of innovativeness are mainly influenced by the system's design and not by its usability. Furthermore the judgements of innovativeness and attractiveness interact over time with respect to the innovativeness as dependent variable.

Thirdly, we investigated relationships between emotional user reactions and system properties. In our study emotions were strongly influenced by usability aspects and in no sense by the level of attractiveness. In that case we have to admit that although the manipulation check indicates a successful induction of a specific attractiveness level, the absolute amount of the higher attractive version is rather average. Therefore the induction of a real positive emotion, e.g. joy, could have been less likely.

Finally, we showed that not only judgments change over time, but also underlying motivations. At the beginning, usability ratings are focused on goal conduciveness and later on more general using aspects. Attractiveness judgments are first related to novelty and later to aspects that consider on the previous interaction, e.g. fascination.

CONCLUSION

We presented an approach to the experimental study of user experience that focuses on systematic changes of different aspects of user experience. By the use of our stimulus material we could show the relevance of those dynamics both for designing and evaluating technical systems.

We used stimuli material that systematically manipulated factors in electronic consumer devices. Therefore the results are limited to this domain. Further studies should vary other product characteristics and identify user parameters as potential mediators of dynamic effects. Furthermore the interaction intervals could be more elongated than it is done in this study. But up to now we cherish:

The way we experience technology develops over time!

ACKNOWLEDGMENTS

This research was supported by the German Research Foundation (DFG) as part of the Research Training Group 'Prospective Engineering of Human-Technology Interaction' (no. 1013).

REFERENCES

1. Carbon, C.-C. & Leder, H. (2007). Design Evaluation: From Typical Problems to State-of-the-Art Solutions. In *Thexis – Journal of Marketing Research (2*), 33-37.
2. Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer-Interaction (19)*, 319-349.
3. Jordan, P.W. (2000). *Designing Pleasurable Products.* London: Taylor & Francis.
4. Kirakowski, J. The software usability measurement inventory: Background and usage. In P. W. Jordan et al. (Eds.), *Usability Evaluation in Industry*, Taylor & Francis (1996), 169-178.
5. Lang, P. J. Behavioral treatment and bio-behavioral asessment: Computer applications. In J. B. Sidowski, H. Johnson & T. A. Williams (Eds.), *Technology in Mental Health Care Delivery Systems*, Ablex (1980), 119-137.
6. Lavie, T. & Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *Int. Journal of Human-Computer Studies*, *60* (2004), 269-298.
7. Lindgard, G. & Dudek, C. (2003). What is this evasive beast we call user satisfaction? *Interacting with Computers*, *15*, 429-452.
8. Mahlke, S. (2008). *User Experience of Interaction with Technical Systems.* VDM-Verlag.
9. Roast, C., Evriviades, M., Purcell, M., & Steele, B. (2002). *Interaction media – using IT and liking IT.* Paper presented at the Pan Hellenic Conference on Human Computer Interaction. Patras, Greece.
10. Thüring, M. & Mahlke, S. (2007). Usability, aesthetics, and emotion in human-technology interaction. *International Journal of Psychology*, *42*, 253-264.
11. Tractinsky, N., Katz, A.S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*, 127-145.