Dynamics of User Experience.

Judgments of Attractiveness, Usability, and Emotions Over Time.

Michael Minge[1] & Manfred Thüring[2]

[1] Technische Universität Berlin
Department of Cognitive Psychology and Cognitive Ergonomics, MAR 3-2
Marchstrasse 23, 10587 Berlin, Germany
E-Mail : michael.minge@tu-berlin.de

[2] Technische Universität Berlin
Department of Cognitive Psychology and Cognitive Ergonomics, MAR 3-2
Marchstrasse 23, 10587 Berlin, Germany
E-Mail : manfred.thuering@tu-berlin.de

- Technical Report -

Technische Universität Berlin, 2011-12-15

Keywords: User experience, temporal dynamics, hedonic features, pragmatic features, aesthetics, usability, halo-effect, emotions.

ABSTRACT

User experience has emerged as an extension of the traditional concept of usability providing a holistic perspective on the user's interaction with technology. Recently, the experiential view of the user has become an essential part of the human-centered-design approach [18]. The concept of user experience can be characterized as a multidimensional phenomenon that comprises both, the perception of different product qualities as well as emotions that arise while a user interacts with a product. The interrelations of these components are described in the 'Component Model of User Experience' (CUE-Model [30]) which serves as the theoretical basis for our experiments. To examine user experience and its alterations over time, two laboratory experiments were conducted. In both studies, participants employed different versions of mobile digital players which were systematically varied with respect to their hedonic, non-instrumental qualities, i.e. visual aesthetics, and their instrumental qualities, i.e. the inherent usability. Essential aspects of user experience, i.e. judgments of visual attractiveness and usability, and emotional responses were measured repetitively at different stages of usage. The analyses show that influences of visual aesthetics and of inherent usability on quality judgments and emotions change over time. Moreover, evidence for two halo-effects was also found: On the one hand, visual aesthetics influenced perceived usability according to the notion "beautiful is usable" by Tractinsky et al. [31], on the other hand, the inherent usability of the systems impacted the perceived aesthetics and emotional responses. While we labelled the first one as 'hedonic halo-effect', we called the latter one 'pragmatic halo-effect'. Based on our results, we propose substantial changes of the CUE-Model and draw conclusions for future research.

INTRODUCTION

Nowadays, technology is not only about usability and user performance. Additional aspects from the user's perspective, such as expectations, perceptions, judgments, feelings, and motivations, become an essential part of any user-centred design activity [18]. Against this background, user experience has emerged as a comprehensive approach for designing and evaluating interactive systems. A variety of attempts have been made in order to define what user experience actually is, and how it could be measured and categorized [e.g. 1, 9, 16, 21, 23]. According to Blythe, Reid, Wright, and Geelhoed [4] two main perspectives can be distinguished: While holistic approaches lay emphasis on subjectivity and on the uniqueness of an experience, reductionist or analytic approaches characterize specific dimensions representing important aspects of human-technology interaction. In line with a number of authors, it is the reductionist view that enables a systematic investigation of experiences and therefore offers the opportunity to quantify and to distinguish diverse experiences [e.g. 14, 23].

As a common analytic framework, the 'Component Model of User Experience' (CUE-Model) by Thüring and Mahlke [30] addresses central issues of different reductionist theories and thereby inspired both, scientific research as well as practical application (see Figure 1). According to the model, each interaction between a system and its user is determined by user characteristics, system properties and contextual components, such as the task that the user aims to accomplish. In the course of his interaction, the user experiences the behavior of the system and perceives the quality of two different product features which Mahlke [23] calls 'instrumental' and 'non-instrumental'. This classification traces back to Hassenzahl [13] who called the two dimensions 'pragmatic' and 'hedonic'. While the first refers to technical features and usability aspects, the latter summarizes system properties that refer to beauty, visual aesthetics, identification, and stimulation. Incidentally, in the CUE-Model, the term 'perception' is used in a broad sense and is not restricted to sensory processes and perceptual organization. It also comprises 'higher' cognitive activities, such as identification, categorization and judgment. In their experiments, Mahlke and Thüring [25] addressed both, 'instrumental' as well as 'non-instrumental' system qualities by varying the usability and the visual appeal of an interactive device. The authors showed that both aspects, independently, determine the overall appraisal of the system. This appraisal affects the user's general opinion as well as future behavior and usage.

Apart from judgments, emotions, by nature, are a determinant of any experience. Since emotional responses to an object are likely to influence behaviour and future acceptance immediately and sometimes unconsciously [10], well-known approaches, such as ‘emotional design’ [e.g. 8, 26], have started to stress their influence on human-technology interaction. In line with cognitive theories of emotion [e.g. 29], affective responses are related to specific cognitive processes. As portrayed by the CUE-Model, such cognitions are part of the two perceptual components and are capable to trigger a synchronized emotional reaction within the user. According to Scherer [29] such a reaction consists of subjective feelings, physiological activation, and motor expressions. While physiological activation can be recorded through autonomic body responses, such as electrodermal activity and heart rate, motor expressions are measurable by the activation of specific facial muscles [6]. Subjective feelings can be characterized in terms of valence and arousal [28], both of which depend on the character and relevance of the perceived product qualities. Empirical results confirm the importance of considering emotions as indicators of user experience. Mahlke and Minge [24] for instance showed that a system with usability flaws led to negative emotions (valence) of high intensity (arousal) while the interaction with a flawless version of that system was accompanied by positive feelings combined with a less arousal level. Both versions were also accompanied by different levels of physiological activation, and they were appraised differently, i.e., the flawless version was rated as more capable, pleasant and convenient.

----

insert Figure 1 about here, please

The CUE-Model inspired a number of experiments in addressing the interplay of instrumental and non-instrumental system qualities, emotional responses and user judgments [23]. One research issue that has a longer history in human-computer interaction concerns the relationship between the perception of instrumental and non-instrumental qualities of a product. Tractinsky et al. [31] replicated the study of Kurosu and Kashimura [22] and re-ported positive correlations between judgments of attractiveness and usability. Those significant correlations were suggested to result from a psychological mechanism called halo-effect: One’s tendency to use apparent features of a stimulus (i.e. hedonic qualities), to estimate non-evident attributes (i.e. pragmatic aspects) [31]. However, empirical studies so far have led to inconsistent results. While Roast et al. [27] surprisingly found both product qualities to be negatively correlated, results by Mahlke [9] and Hassenzahl [13] suggest both product qualities to be independent. For a partially explanation of these inconsistencies we followed the assumption of Karapanos et al. [22] who considered time as an important contextual factor inducing diversity in user experience. Since experiences are “dynamic and temporally-bounded” [16], systematic changes in judgments and emotions may result when a user interacts with a system over a certain period of time. Consequently, we made three additional assumptions with respect to the directionality and the temporal dynamics of the halo-effect as an explanation of the reported results:

1. Many non-instrumental qualities are apparent at first sight. Therefore, positive features, like an attractive design, may influence usability judgments positively, while negative hedonic aspects may have the opposite effect.
2. Perceived usability comes into play when a user interacts with a system. Hence, a different type of the halo-effect could be supposed, that is to say that negative instrumental qualities, such as usability flaws, may influence aesthetic judgments negatively, while positive instrumental qualities may have the opposite effect.
3. The perception of instrumental and non-instrumental qualities may change over time. While, for instance, the visual attractiveness of a device may wash off after a while, the perceived usability may even increase in the course of usage. Hence, both types of the halo-effect may vary over time and reinforce or neutralize each other.

This pattern of potential influences calls for studies that systematically vary pragmatic and hedonic features of an interactive product and capture judgments of its perceived qualities repetitively. A similar attempt was already conducted in the experiments reported by Thüring and Mahlke [30], but without any temporal aspects being addressed. This paper portrays two experiments investigating the temporal dynamics of perceived usability, perceived attractiveness, and subjective emotional feelings.

STUDY 1

The first experiment aimed to investigate temporal influences of instrumental and non-instrumental product qualities on users’ judgments and emotions. To capture the temporal dynamics of user experience, Study 1 employed procedure and materials similar to the studies by Mahlke and Thüring [25], but simultaneously introduced the repetition of measures over time as an experimental factor (in a similar way as Tractinsky et al. [31]). Participants interacted with a simulated mobile device under laboratory conditions. Judgments of visual attractiveness, usability, and emotions were collected at three different times. In particular, it should be investigated to which extent ratings of usability are influenced by non-instrumental product qualities, such as visual aesthetics ('hedonic halo-effect'), and to which extent ratings of visual attractiveness may be affected by instrumental qualities, such as different degrees of usability ('pragmatic halo-effect'). Moreover, relations between judgments and emotions were regarded when systems of different product features were employed over time.

Participants

Sixty participants (24 female, 36 male) were recruited from the local university. The mean age was 27.4 years (*SD* = 7.21). Fifty-four participants reported to own a portable digital audio player and the remaining six had no prior experience with such devices.

Stimuli

Interactive simulations of portable digital audio which had been introduced by Mahlke and Thüring [25] were used as stimuli. These simulations varied systematically with respect to their impact on perceived instrumental and non-instrumental product qualities. Both factors included two levels each (low and high) and had been validated previously in two pilot studies [25].

Instrumental qualities were manipulated by the product's inherent usability. Therefore, in the low usability condition, difficulties were introduced to the interaction between the audio player and the participants. 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 indicating the actual position in the menu hierarchy (given or not, see Figure 2).

----

insert Figure 2 about here, please

Non-instrumental qualities were manipulated by the product's visual aesthetics. While the less aesthetic version consisted of an angular shaped body design with a color combination of blue and green (high color differences) and an asymmetrical position of display and key panel, the more aesthetic version was curved shaped and colored blue and grey (low color differences, see Figure 3).

----

insert Figure 3 about here, please

The audio-players were presented as prototypes on a 7” hand-held monitor with touch screen functionality which participants used for providing input. The display was connected to a computer which ran the simulation.

Design

A 2 x 2 x 3 factorial design was applied in the experiment with inherent usability and visual aesthetics as between-subject factors, each having two levels (low and high). Participants were randomly assigned to one of the resulting four conditions: (a) high usability and high aesthetics, (b) high usability and low aesthetics, (c) low usability and high aesthetics, (d) low usability and low aesthetics. The experimental design was complemented by a third factor, i.e., the time when participants produced their ratings. This within-subject factor consisted of three levels: T1 (before using the system), T2 (after freely exploring the system), and T3 (after using the system for solving a number of tasks).

Dependent variables

Perceived usability and perceived visual attractiveness on single-item scales were regarded as dependent variables. Both items were judged on a 7-point Likert scale. Although measurement theory advocates the use of multiple item measures, various findings indicate that single-item measures are acceptable given the circumstances which are similar to the current study [compare 31, 32]. To minimize response effects resulting from repeated measurements, we varied the order of the items at T1, T2, and T3. In addition to these ratings, participants reported their emotional feelings using the 9-point scales arousal and valence of the Self-Assessment-Manikin (SAM) by Bradley & Lang [5].

Procedure

After an explanation of the experiment and a short display of the system, participants were asked to fill in the SAM scales. Also, they were required to rate the visual appearance by the single-item perceived visual attractiveness and the impressions of usability by the single-item perceived usability. Thus, individuals’ “visceral reactions” [compare Norman, 26] and their first impressions were captured at a very early stage (T1).

Consequently, the operating of the audio player was explained to the participants and they were also given 2 minutes to practice with the system. Since no task was given to guide this first interaction, they were free to get acquainted with the system as they thought best. After this short exploration, they filled in the SAM scales and rated the visual attractiveness and usability of the system for the second time (T2).

In the third part of the experiment, participants were asked to solve a number of tasks which are typically carried out using an audio player, such as looking for a track or volume regulation. Twenty-eight tasks were available and each person had fifteen minutes to solve as many of them as possible. The content and the order of the tasks were identical for each participant. To conclude the experiment, subjects once again rated their emotional state as well as the visual attractiveness and usability of the system (T3).

Finally, sociodemographic data was acquired. Participants also answered a standardized questionnaire measuring “individual differences in the centrality of visual product aesthetics” (CVPA by Bloch et al. [3]) and questions capturing their usage experience of portable MP3-players. Both variables – the individual preference for aesthetics and the usage experience in years – served as covariates for further analysis. An experimental session lasted half an hour on average. Each participant received a gratification of 5 Euro.

Results

A multivariate analysis of covariance (MANCOVA) was carried out with the independent variables usability and visual aesthetics as between-subject factors and time levels as within-subject factor. As covariates the analysis includes participants’ values of the CVPA questionnaire and the number of years individuals were experienced in using portable MP3-players.

As expected, a significant main effect of our variation of non-instrumental product qualities on perceived visual attractiveness was found [F (1,54) = 14.390, $p < .01$, *part.* $\eta^2 = 0.210$]. The mean scores are displayed in Table 1, indicating that the more aesthetic systems were perceived as more visual attractive. Moreover, the ratings on perceived visual attractiveness were substantially affected by an interaction between the system’s usability and the three levels of repeated measurements [F (1,108) = 5.113, $p < .01$, *part.* $\eta^2 = 0.086$]. According to that, participants regarded the flawless systems as more visual attractive in the course of usage, while they downgraded their ratings on visual attractiveness in case of the flawed versions. A second interaction on perceived attractiveness was caused by the factors visual aesthetics and time level [F (1,108) = 8.953, $p < .01$, *part.* $\eta^2 = 0.142$], indicating that the highly aesthetic version of the audio player was perceived as less visual attractive over time, while the less aesthetic one became more appreciated. No other significant differences were found for the ratings on visual attractiveness.

----

insert Table 1 about here, please

The ratings on perceived usability were influenced by the manipulation of instrumental product qualities [F (1,54) = 41.823, $p < .01$, *part.* $\eta^2 = 0.436$]. Compared to the flawed system, ratings on perceived usability were, in sum, substantially higher in case of the flawless system (see Table 2). A significant interaction effect between the factors usability and time level was also found [F (1,108) = 12.261, $p < .01$, *part.* $\eta^2 = 0.158$]. Results of a Bonferroni post hoc analysis confirmed that at T1, the inherent usability of the system – that had not yet been experienced at this point – did not affect the perceived usability ($p = .13$). Only at T2 and T3, after participants had gained an impression of the system's usability, the ratings became increasingly affected by this factor ($p < .01$). Finally, the analysis revealed a significant interaction effect between the visual aesthetics and time level [F (1,108) = 3.096, $p < .05$, *part.* $\eta^2 = 0.054$]. While participants at the beginning rated the perceived usability higher in case of the more aesthetic system, usability judgments at the end (T3) are not influenced by the visual appearance any longer ($p = .87$). No other significant differences were found for the ratings on usability.

----

insert Table 2 about here, please

Mean ratings on arousal and valence measuring emotional reactions are displayed in Table 3. Arousal was neither influenced by usability nor by visual aesthetics. A significant main effect was merely found in the within-factor time level [F (1,108) = 4.790, $p < .01$; *part.* $\eta^2 = 0.081$], indicating that participants were more aroused at the beginning than at the end. However, valence was marginally affected by both, usability [F (1,54) = 3.976, $p = .051$; *part.* $\eta^2 = 0.069$] and time level [F (1,108) = 3.060, $p = .051$; *part.* $\eta^2 = 0.054$]. According to that, interacting with the more usable version and later stages of usage led to more positive emotions. Finally, there was a significant interaction effect between usability and time level on valence [F (1,108) = 3.507, $p = .033$; *part.* $\eta^2 = 0.061$]. Compared to the less usable system, it was the flawless version that considerably stronger evokes positive emotions in the course of usage.

----

insert Table 3 about here, please

No temporal effect was significantly influenced by the covariates prior experience and centrality of visual product aesthetics. Nevertheless, a general effect of prior usage experience on perceived usability was indicated [F (1,54) = 8.630, $p < .01$, *part.* $\eta^2 = 0.138$], suggesting that highly experienced participants regarded usability as substantially lower.

Discussion

Results indicated that in Study 1 judgments of usability and visual attractiveness are substantially influenced by the variation of instrumental and non-instrumental product qualities. Significant main effects revealed that, in general, the flawless systems were regarded as more usable, while the more aesthetic versions led to more favourable judgments. The nature of repeated measures, allows a more detailed investigation of temporal influences through which three interesting interacting effects over time were observable.

Firstly, usability ratings were affected by an interaction between the system's visual aesthetics and time. Hence, at an early stage, before having experienced the audio player's usability, perceived usability was rated higher in case of higher visual aesthetics. This result is in line with the observation that, at T1, usability did not significantly affect ratings of usability. The latter comes as no surprise since, at that particular time, participants had not yet interacted with the audio players. In accordance with the study by Tractinsky et al. [31], Study 1 nicely replicated the halo-effect of visual aesthetics on perceived usability and hence supported the authors' notion of "beautiful is usable" [31]. As indicated by the significant interactions, the halo-effect has vanished in the course of usage, while participants experienced the player's inherent usability. Interestingly, even short episodes of usage - as in our exploration phase - can be sufficient for the formation of usability impressions. Since the duration of usage went along with the solving of tasks, it is difficult to tell whether this development was caused by merely by time or by the intensity of usage.

Secondly, the results supported our hypothesis, that different degrees of usability may affect judgments of non-instrumental product qualities, e.g. visual attractiveness. Similar to the 'hedonic halo-effect' a 'pragmatic halo-effect' indicated by an interaction between usability and time was also evident through the ratings of visual attractiveness. These findings comply with Jordan's idea of a hierarchical pyramid where functionality and usability are regarded as fundamental user needs ahead of experiencing hedonic qualities and pleasure [19]. Evidence so far supports the assumption, that one might as well reverse Tractinsky's notion to 'usable is beautiful' when products are employed intensively over a longer period of time. Note that two mechanisms may be responsible for this development. On the one hand, visual attractiveness may “wash off” after a while – on the other hand, visual attractiveness may become dominated by the factor ‘usability’ when people get the chance to interact with the system. So far, it is difficult to tell, which of those mechanisms would have stronger influence the occurrence of a ‘pragmatic halo-effect’.

A third interaction was revealed between visual aesthetics and measurement time on ratings of visual attractiveness. In the course of usage, participants rated the visual attractiveness of the less aesthetic system increasingly higher, whereas the more aesthetic system was regarded as less visual attractive. The concept of typicality as predictor of aesthetic appeal may help to explain this result. In the context of product experience, Hekkert et al. [17] found, that one prefer objects with the best combination of typicality and novelty following the design principle “most advanced, yet acceptable” [17, p. 111]. Therefore, the relationship between preference and typicality is an inverted-U function, where the most preferred objects are those with a moderate level of typicality. Nevertheless, Carbon and Leder [7] showed, that this effect of typicality on ratings of visual attractiveness systematically change over time. The authors invented the repeated evaluation technique (RET) to capture those dynamic effects. They found that at first sight participants judged the visual attractiveness higher in case of rather typical, less innovative car interior designs, whereas after an elaborated exposure of different designs, rather not typical, more innovative designs were favoured. The current observation concerning digital audio players is somehow in line with these findings. Asymmetry and color combination of the less aesthetic player gives a considerably more unfamiliar impression of a mobile consumer device than the rounded body shape of the more aesthetic version does. In contrast to the study by Carbon and Leder participants in our experiment had not the chance to evaluate other designs but they employed and experienced the player over a longer period. While the less typical design may have led to a positive familiarization, the more typical design could have induced a sort of boredom in the course of usage. Note that further research is necessary to support this speculative assumption.

With respect to user emotions, participants reported a more positive feeling in case of good usability what comes as no surprise. In contrast, the variation of visual aesthetics seems to have no influence on emotions. This could have been caused by the manipulation itself being not strong enough to produce real joy or by the experimental situation which focused on an instrumental working context. Since emotions were only measured retrospectively by a subjective questionnaire, further experiments should consider additional physiological data to capture emotional reactions continuously and more objectively. Furthermore, one could argue that the manipulation of visual aesthetics did not lead to a truly high aesthetic version.

Although significant differences concerning judgments of product qualities could be observed, mean scores indicate, that participants rated visual attractiveness in case of the more aesthetic version only on an average level. To strengthen our assumptions on ‘hedonic’ and ‘pragmatic halo-effects’, Study 2 employed a reworked manipulation of instrumental and non-instrumental product qualities. In order to study temporal influences of usability flaws on the occurrence of halo-effects in more detail, the position of flaws during the experiment was systematically varied.

STUDY 2

To investigate temporal dynamics of judgments and emotions, Study 2 pursue a similar approach as the first experiment. Again, participants employed a simulated mobile audio-player to solve prototypical usage tasks. With respect to the manipulation of non-instrumental product qualities, new body shapes with a different degree of visual aesthetics were developed and pretested. To vary instrumental qualities, participants experienced usability flaws at specific stages during the experimental session. The aim of this design was to replicate the occurrence of both halo-effects and to explore positional influences of experienced usability flaws on (a) the intensity of halo-effects and (b) overall evaluations which assess user experience retrospectively. As results by Hassenzahl and Sandweg [15] indicated, summative judgments of perceived usability are mostly determined by those events which were experienced immediately before the evaluation. The authors explained their finding with the concept of a ‘recency-effect’: “Individuals construct their summary assessment on the basis of what comes to their mind about the episode they just experienced. The more recent a detail, the more easily it comes into mind.” [15, p. 1285]. Since Study 2 additionally includes both, judgments of non-instrumental qualities and user emotions, ‘recency-effects’ should be investigated in a more comprehensive way. Finally, relations between judgments and emotions should be clarified. As already suggested ahead, subjective feelings were complemented by a multi-component approach including physiological and motor expressive aspects of emotional reactions.

Participants

Ninety-six individuals (half of them female) participated in the study. The mean age was 26.01 years (SD = 4.37). Eighty-seven participants owned a portable digital audio player.

Stimuli

Again, portable digital audio players were chosen as the domain of study. With respect to non-instrumental product features we reworked the manipulation of visual aesthetics, still focusing on shape (rounded or angular), symmetry (high or low), and color combination (grey and silver or red and green). Besides that, the design of the key panel was varied: While buttons of the low aesthetic system were angular-shaped and separately ordered, the highly aesthetic version integrated the same button functionality into a black circle (see Figure 4).

----

insert Figure 4 about here, please

With respect to system features influencing instrumental product qualities, usability aspects were manipulated. In contrast to Study 1, flaws were not implemented as a constant feature of the system, but should occur at different stages of product usage (at the beginning, in the middle, at the end). Therefore, waiting loops were developed, making the system incontrollable for twenty-five seconds. Meanwhile, a verbal feedback and a progress bar were presented. To ensure a clear distinction between the manipulations (visual aesthetics and usability), three pre-tests were conducted. In case of visual aesthetics, results confirmed, that both versions significantly differed from the single-item scale's mean. The same apparatus as in Study 1 was employed for presentation's purpose.

Design

A 2 x 4 factorial design was applied in the experiment with visual aesthetics (low and high) and position of usability flaws (at the beginning, in the middle, at the end, and no flaws) as between-subject factors. Participants were randomly assigned to one of the resulting eight experimental conditions. Ratings were measured several times: Before (pre), during (T1, T2, and T3), and after (post) participants had worked with the digital audio-player. As covariates the analysis again includes participants' values of the CVPA questionnaire [3] and the number of years individuals were experienced in using portable MP3-players.

Dependent variables

The current study similarly employed single-items measures for perceived usability and visual attractiveness. Subjective feelings were recorded through by the Self-Assessment-Manikin (SAM, [5]). In order to realize a multi-component approach for measuring emotions [22] we additionally registered objective data. Heart rate (HR) and electrodermal activity (EDA) were captured as physiological reaction indicators. The muscle activation (EMG) of the corrugator supercilii served as indicator for facial expressions. In addition to the single-items, participants completed the AttrakDiff [13] at the end of the experiment.

Procedure

After an explanation of the experiment and a short display of the system, participants were asked to rate the single-items usability and visual attractiveness and to complete the SAM scales. These data captured participants' first impressions and served as a pre-experimental measurement (pre).

Next, electrodes for measuring physiological reactions and facial expressions were attached, and baseline values were recorded for two minutes. Afterwards, the operating of the audio-player was explained and participants started to solve the first block of four typical tasks for using an audio player, similar to Study 1. While physiological data was registered continuously, judgments on visual attractiveness, usability, and emotional feelings were collected after the four tasks had been solved (T1).

In the second and third part of the experiment, participants worked on further interaction blocks, consisting of four tasks each. Before the experiment starts, the complexity of the tasks was ensured to be comparable between the three blocks. The presentation order of both, the blocks and the four tasks within the blocks, was balanced. At the end, judgments on visual attractiveness, usability, and emotional feelings were collected, respectively (T2 and T3). Finally, electrodes were removed and participants completed the AttrakDiff. Once again, they answered the single-items and the SAM scales (post). At last, sociodemographic data was acquired. An experimental session lasted forty minutes on average. All participants received a gratification of 5 Euro.

Results

Since the independent variables are measured by a different frequency, the results are reported based on the analysis that were calculated. The first analysis included merely the pre and post data. Afterwards, the results of physiological reactions and immediate judgments at the end of each block are reported. Consequently, analysis of AttrakDiff allowed summative evaluation to be deducted.

Pre and post data

In order to compare pre and post data a 2 x 4 x 2 MANCOVA was calculated, with visual aesthetics (low and high) and position of usability flaws (no flaws, at the beginning, in the middle, at the end) as between-subject factors and time (pre and post) as within-subject factor. The analysis confirmed that participants rated the visual attractiveness higher in case of the more aesthetic version than in case of the less aesthetic one [F (1,86) = 45.176, $p < .01$; *part.* $\eta^2 = 0.339$]. Moreover, the analysis revealed a significant main effect of the factor time on the ratings of visual attractiveness [F (1,86) = 9.715, $p < .01$; *part.* $\eta^2 = 0.099$], indicating that participants perceived the visual attractiveness higher after having experienced the system. A Bonferroni post hoc analysis manifested, that this increase is only substantial ($p < .05$) within the control condition in which usability flaws were excluded. Besides that, visual attractiveness was not significantly affected by the independent variables, neither by the factor flaw position nor by any other significant interaction.

Perceived usability, however, was significantly influenced by the position of usability flaws [F (3,86) = 2.910, $p < .05$; *part.* $\eta^2 = 0.090$]. For a pairwise comparison of the flaw positions, we calculated Bonferroni post hoc tests. Nevertheless, there could no significant differences been identified between the control condition and the experimental conditions.

More interestingly, a second main effect for the factor visual aesthetics was found [F (1,86) = 5.853, $p < .05$; *part.* $\eta^2 = 0.062$], indicating that participants regarded the more aesthetic audio-players as more usable. Experiencing the system over time had no substantial effect on the ratings of usability. Also, there were no significant interaction effects found.

With respect to emotional feelings, the factor visual aesthetics affected participants' arousal [F (1,86) = 5.332, $p < .05$; *part.* $\eta^2 = 0.057$]: It is the less aesthetic version that led to higher ratings on the arousal dimension. Neither significant effects on arousal nor on valence were found.

The effects were not influenced by the covariates prior experience and centrality of visual product aesthetics (CVPA). Nonetheless, a general effect of prior usage experience on arousal was observed [F (1,86) = 5.279, $p < .05$, *part.* $\eta^2 = 0.058$], indicating that lower experienced participants reported a higher arousal in general.

Immediate judgments and physiological data

To analyze the repetitively measured data, three experimental conditions of different flaw positions (at the beginning, in the middle, at the end) were respectively contrasted against the corresponding control condition. Consequently, three single 2 x 2 MANOVAS with visual aesthetics (low and high) and usability (flawed and flawless) as between-subject factors were computed. Each condition consisted of $n = 24$ participants.

Over all blocks, single-item ratings on visual attractiveness were significantly higher in case of high aesthetics ($p < .01$), insuring that this manipulation had a persisting effect on the perception of non-instrumental product qualities. However, an expected influence of the usability manipulation on the single-item ratings of usability could marginally only been identified at T1, after participants had solved the first block of tasks [F (1,44) = 3.04, p = .07; *part.* η^2 = 0.072] and at T3, after participants worked on the last block [F (1,44) = 2.91, p = .06; *part.* η^2 = 0.057].

Therefore, at T1 and T3, the ratings on visual attractiveness were significantly influenced by the factor usability [F (1,44) = 7.66, p < .01; *part.* η^2 = 0.151 and F (1,44) = 6.90, p < .05; *part.* η^2 = 0.144], indicating that in both cases the flawed versions were regarded as less visual attractive. At no time the factor visual aesthetics affected perceived usability.

With respect to emotional reactions, significant influences of the factor usability at later stages of the experiments were only found. While the flawed version led to a high EDA values at T2 [F (1,44) = 4.10, p < .05; *part.* η^2 = 0.089], the factor usability affected the muscle activation of the corrugator supercilii [F (1,44) = 4.00 p < .05; *part.* η^2 = 0.077] as well as ratings on arousal [F (1,44) = 4.34, p < .05; *part.* η^2 = 0.090] and on valence [F (1,44) = 4.62, p < .05; *part.* η^2 = 0.102] at T3. Here, flaws led to a higher muscle activation, higher arousal and less positive emotions. The factor visual aesthetics had no influence on the emotional state of the participants.

Summative evaluations

Ratings on the dimensions of the AttrakDiff questionnaire (identification, stimulation, pragmatics, and global attractiveness) were analyzed by a 2 x 4 MANCOVA with visual aesthetics (low and high) and position of usability flaws (no flaws, at the beginning, in the middle, at the end) as between-subject factors and time (pre and post) as within-subject factor. The analysis revealed a significant main effect for the factor visual aesthetics on identification [F (1,86) = 28.162, p < .01; *part.* η^2 = 0.247] and on global attractiveness [F (1,86) = 19.371, p < .01; *part.* η^2 = 0.184]. Mean ratings indicate that the more aesthetic version led to higher judgments on both dimensions. Moreover, there was a significant main effect of visual aesthetics on the ratings of pragmatics [F (1,86) = 5.363, p < .05; *part.* η^2 = 0.059], indicating, that the more aesthetic version was regarded as more usable.

With respect to the factor position of usability flaws we found a significant main effect on perceived pragmatic quality [F (3,86) = 2.742, p < .05; *part.* η^2 = 0.087]. Bonferroni post hoc tests indicated that, compared to the control condition, summative ratings on pragmatics were substantially lower if flaws were experienced at the end (p < .05). The analysis revealed no other significant effects.

Results were not affected by the covariates, albeit participants with a lower prior usage experience reported a higher stimulation in general [F (3,86) = 4.527, p < .05; *part.* η^2 = 0.050].

Discussion

Despites the similar findings in terms of hedonic and pragmatic halo-effect, several differences were also indicated. However, there is an evidence of a persisting impact on judgments of instrumental product qualities ('hedonic halo-effect'). As results by Tractinsky and colleagues [31] would have predicted, we found a strong influence of visual aesthetics on perceived usability, both in pre and post measurements. Beyond that, summative evaluations did not support our hypotheses of a 'pragmatic halo-effect'. Significant influences of instrumental qualities on ratings of visual attractiveness were illustrated in the experiment: At T1 and T3, participants rated visual attractiveness lower given that they had experienced usability flaws immediately before the evaluation. Hence, the 'pragmatic halo-effect' of the present study seems to be much weaker than the one of Study 1, while the 'hedonic halo-effect' appears to be much stronger. This could be explained by our reworked manipulation of usability and visual aesthetics. As usability flaws waiting loops including progress bars and verbal feedback were presented. This manipulation was however, assumed to be insignificant in creating such a persisting level of a low or high usability impression, as suggested by the post hoc comparisons between the levels of usability factor on ratings of usability suggest. Therefore, experiencing the manipulation of usability may have not been influential enough on the ratings of non-instrumental qualities, i.e. visual attractiveness.

In line with Study 1, data support the assumption of a mere-exposure effect for the ratings of visual attractiveness [33]. Both versions of the MP3-player, the more aesthetic and the less aesthetic one, were perceived as being more visually attractive at the end. Interestingly, this effect was merely substantial in case of experiencing no usability flaws at all. Since it is the instrumental quality which affects the perception of a non-instrumental quality, i.e. visual attractiveness, a 'pragmatic halo-effect' may play a major role at this point.

With respect to emotional feelings, results reveal an effect of visual aesthetics on arousal. The high color combination of the less aesthetic version (red and green) should have particularly contributed to this result. The effect of usability on emotional valence was greater than the one for visual aesthetics on valence. Data also show that emotional reactions could only be observed if participants had interacted with the mobile device and experienced the player's usability. EMG data and other physiological measures support this interpretation. Although expected differences in the measured variables at all measurement points were not observed, it can be concluded that the current experimental variation has produced the desired effects in terms of distinct patterns of subjective feelings, motor expressions, and physiological reaction.

To summarize both experiments, we will now discuss the results in a broader sense and thereby take a closer look into the nature of halo-effects.

GENERAL DISCUSSION

Visual attractiveness is a central non-instrumental quality for many technical devices, such as audio players or mobile phones. In the current experiments, participants were found to be particularly sensitive to this quality. Before they had the opportunity to interact with the device, they rated the appeal of the different player versions according to our variation of visual aesthetics. As Norman [26] proposed, hedonic features are processed at the 'visceral level': 'This is where appearance matters and first impressions are formed. Visceral design is about the initial impact of a product, about its appearance, touch and feel.' [26, p. 37]. The current findings concerning the instrumental qualities of the audio players nicely comply with Norman's ideas on 'behavioral design' where 'function comes first and foremost' [26, p. 70] and where the usability of a product – and not its appearance – matters.

As depicted, apparent system properties may influence the perception of non-instrumental qualities in terms of a 'hedonic halo-effect' as well as inherent usability may affect judgments of visual attractiveness in terms of a 'pragmatic halo-effect'. These effects change in the course of system usage, and there is some evidence that each of these factors has dynamic characteristics of its own. In addition, they may influence each other.

It was assumed that two different mental mechanisms are responsible for these mutual influences, one of them being cognitive in nature, the other emotional.

The first mechanism can be described in terms of heuristics. Kahneman, Slovic, and Tversky [20] were among the first to study heuristics and biases in the field of cognitive psychology. Their research revealed that one relies on rules of thumb when they cannot solve a problem by merely retrieving facts from memory. Gigerenzer and colleagues [12] showed that such simple heuristics 'make us smart' by producing sufficiently good solutions at reasonable costs. From this perspective, Tractinsky's notion "beautiful is usable" may actually be a heuristic for judging the usability of a system without having to use it. Such a heuristic may be justified by previous experience showing that most well designed systems are also highly usable.

The second mechanism focuses on the role of emotions. The basic idea is that the perception of instrumental or non-instrumental qualities in the course of system usage is accompanied by corresponding emotions. For instance, enjoying a well designed interface may lead to pleasure, or being delayed by usability flaws when solving a task may cause anger. These emotions might affect the perception of system qualities and bias judgments of perceived attractiveness and usability according to the valence and arousal of the emotion.

How can these two mechanisms be brought in line with the results of the current experiments?

At early stages, it was assumed that the 'hedonic halo-effect' results from a heuristic analog to Tractinsky's "beautiful is usable". As Gigerenzer and Brighton [11] pointed out: 'Heuristics are always second-best … [and we] use heuristics only because of our cognitive limitations.' (p. 109). Since our participants had no chance to experience the systems before rating its usability, there was no other information but visual appearance they could draw on. Hence, using a heuristic, like "beautiful is usable" (instead of just guessing the degree of usability) is a very rational way to cope with that situation.

After having to work with the system one has got an impression of the usability. Consequently, participants could base their judgments on impressions from exploring the system. Results of both studies revealed that only at later stages, usable system versions led to more positive emotions than the versions with usability flaws. Additionally, only at later stages the more usable systems were perceived as being more visual attractive ('pragmatic halo-effect'). Evidence so far supports the assumption that longer periods of emotional experiences correlate with the occurrence of a 'pragmatic halo-effect'. Hence, as the discussion shows, heuristics as well as emotions may both contribute to the halo-effects we found. Of course, these theoretical assumptions must be further refined and tested. We will take a first step in that direction in our conclusions.

CONCLUSIONS

Heuristics are an essential part of the human cognitive architecture, and it seems that one is able to draw on the whole 'toolbox' of such general rules for judgment and problem solving [11]. In the framework of the CUE-Model, it was assumed that – given the right circumstances – such heuristics influence the perception of instrumental and non-instrumental qualities. A halo-effect occurs when the heuristic acquires its input from one type of quality and uses it to produce judgments with respect to the other type. For instance, a heuristic such as "beautiful is usable" may take the degree of visual attractiveness as an input to estimate the degree of usability, thus producing the 'hedonic halo-effect' found in the study.

In order to account for this mechanism, the CUE-Model was modified by introducing relations between the perception of instrumental and non-instrumental qualities, representing the 'hedonic' and the 'pragmatic halo-effect'. The links cover both directions because there might be heuristics which take perceived non-instrumental qualities as input to generate or modify judgments on instrumental qualities and vice versa (see Figure 5).

----

insert Figure 5 about here, please

As pointed in the general discussion, heuristics *alone* are not sufficient to explain the two halo-effects in our experiments. We suppose that emotions resulting from the perception of instrumental and non-instrumental qualities play a major role in that respect. For the ‘pragmatic halo-effect’, favorable emotions resulting from instrumental qualities shift judgments of non-instrumental qualities in the positive direction while negative emotions shift them in the opposite direction. Analogically, positive emotions resulting from non-instrumental qualities may contribute to the ‘hedonic halo-effect’ by biasing judgments of instrumental qualities. To represent potential influences of emotions in the context of halo-effects, we propose to change the uni-directional links between the perception of qualities and emotions by additional pathways, where emotions play the role of mediating variables. At this point, a note of caution is required concerning the status of emotion in our experiments. Although there were significant effects with respect to valence, arousal, and partially even to the pattern of physiological data, the independent variables in both experiments failed to impact all measured variables in the expected way. Therefore, further experiments are still required to investigate under which circumstances heuristics alone are capable to determine the occurrence of halo-effects and what kind of emotional experiences reinforce or diminish them. For example, one could assume that temporal dynamics and halo-effects may intensify when emotions increase in strength.

Additionally, emotions are no longer assumed to directly influence the overall appraisal of the system but rather conveyed and moderated by the cognitive processes inherent to the perception of instrumental and non-instrumental qualities (compare Figure 5). Obviously, this hypothesis requires further investigation too and belongs to important topics for future research. Other issues in need of clarification are the long-term development of perceived instrumental and non-instrumental qualities, the exact characteristics of interaction that influence emotions and judgments, the further deployment of physiological measures to register emotional responses continuously as well as the question of further heuristics that might play a role in judging interactive products.

Beside these experimental investigations, a theoretical challenge lies in the transformation of the theoretical framework of the CUE-Model into a procedural and more formal model. Such a model, together with empirical insights into the topics mentioned above, is required to sharpen the concept of user experience. A precise concept may provide the basis for making the usage of interactive products more enjoyable and the joy more intense and sustained.

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). We would like to thank Jasmin Hannighofer and Andrea Stein for their work on the studies and Adeline Chandra for many clarifying and helpful comments.

REFERENCES

1. Alben, L. 1996, Quality of Experience. *Interactions,3*(3), 11-15.
2. Ariely, D. & Ziv, C. (2003). Summary Assessment of Experiences: The Whole is different from the Sum of its Parts. In G. Loewenstein, D. Read & R. Baumeister (Eds.). Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice. NY: Russell Sage.
3. Bloch, P. H., Brunel, F. F. & Arnold, T. J. (2003). Individual Differences in the Centrality of Visual Product Aesthetics (CVPA): Concept and Measurement. *Journal of Consumer Research, 29,* 551-565.
4. Blythe, M., & Wright, P. (2003). Introduction – From Usability to Enjoyment. In: M. Blythe, K. Overbeeke, A.F. Monk, & P.C. Wright (Eds.), *Funology: From Usability to Enjoyment.*
5. Bradley, M. M. & Lang, P. J. (1994). Measuring emotions: the self-assessment manikin and the semantic differential. Journal of Behavioral Therapy and Experimental Psychiatry, 25(1), 49-59.
6. Cacioppo, J. T. & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. American Psychologist, 45, 16-28.
7. Carbon, C.-C. & Leder, H. (2005). The Repeated Evaluation Technique (RET). A method to capture dynamic effects of innovativeness and attractiveness. *Applied Cognitive Psychology*, *19*(5), 587-601.
8. Desmet, P. (2002). *Designing Emotions*. Doctoral Thesis, Delft University of Technology.
9. Forlizzi, J. & Battarbee, K. (2004). Understanding experience in interactive systems. *In the Proceedings of DIS2004*, 261-267.
10. Frijda, N.H. (1986). *The emotions.* Cambridge, UK Cambridge University Press.
11. Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1,* 107–143.
12. Gigerenzer, G., Todd, P. M., & the ABC Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.
13. Hassenzahl (2002). Beyond usability – Appeal of interactive products. *iCom – Usability and Emotion* (1), 32-40.

14. Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In G. Szwillus & J. Ziegler (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung* (pp. 187-196). Stuttgart: B.G. Teubner.

15. Hassenzahl, M. & Sandweg, N. (2004). From Mental Effort to Perceived Usability: Transforming Experiences into Summary Assessments. *In CHI 2004,* April 24-29, Vienna, Austria. ACM 1-58113-703-6/04/0004.

16. Hassenzahl, M. & Tractinsky, N. (2006). User Experience – a research agenda. *Behavior & Information Technology, 25*(2), 91 -97.

17. Hekkert, P., Snelders, D., & Van Wieringen, P. C. W. (2003). Most advanced, yet acceptable: typicality and novelty as joint predictors of aesthetic preference in industrial design. *British Journal of Psychology, 94,* 111-124.

18. ISO 9241-210 (2010). *Ergonomics of human-system interaction - Part 210: Human-centred design process for interactive systems*. Geneva: International Standardization Organization (ISO).

19. Jordan, P.W. (2000). *Designing Pleasurable Products.* London: Taylor & Francis.

20. Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

21. Karapanos, E., Zimmermann, J., Forlizzi, J., & Martens, J.-B. (2009). User Experience Over Time: An Initial Framework. *Proceedings of CHI 2009* (pp. 729-738). New York: ACM.

22. Kurosu, M. & Kashimura, K. (1995). Apparent usabilty vs. inherent usability. *CHI 1995 conference Proceedings.* New York: ACM Press, 292-293.

23. Mahlke, S. (2008). *User Experience of Interaction with Technical Systems. Theories, Methods, Empirical Results, and Their Application to the Design of Interactive Systems.* Saarbrücken, Germany: VDM Verlag.

24. Mahlke, S. & Minge, M. (2008). Consideration of Multiple Components of Emotions in Human-Technology Interaction. In C. Peter & R. Beale (Eds.), *Affect and Emotion in HCI, LNCS 4868* (51-62). Berlin: Springer.

25. Mahlke, S. & Thüring, M. (2007). *Studying Antecedents of Emotional Experiences in Interactive Contexts.* CHI 2007 Proceedings (pp. 915-918). New York: ACM Press.

26.Norman, D. (2005). *Emotional Design: Why We Love (or Hate) Everyday Things.* New York: Basic Books.

27.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.

28.Russell, J.A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology, 39*(6), 1161-1178.

29.Scherer, K.R. (1984). On the nature and function of emotion: A component process approach. In: K.R. Scherer & P. Ekman (Eds.) *Approaches to emotion* (293-317). Hillsdale: Erlbaum.

30.Thüring, M. & Mahlke, S. (2007). Usability, aesthetics, and emotions in human-technology-interaction. *International Journal of Psychology*, 42, 253-264.

31.Tractinsky, N., Katz, A.S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13,* 127-145.

32.Wanous, J.P., Reichers, A.E., & Hudy, M.J. (1997). Overall Job Satisfaction: How Good Are Single-Item Measures? *American Psychological Association, 82*(2), 247-252.

33.Zajonc, R.B. (1968). *Attitudinal Effects of Mere Exposure. Journal of Personality and Social Psychology, 9*(2), 1-27.

FIGURE CAPTIONS

Figure 1. Components of user experience (CUE-Model) [14]

Figure 2: Screenshots of the more aesthetic (left) and the less aesthetic (right) simulation of digital audio players.

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

Figure 4: Screenshots of the more aesthetic (left) and the less aesthetic (right) simulation of digital audio players.

Figure 5. Modified CUE-Model.

Figure 1.

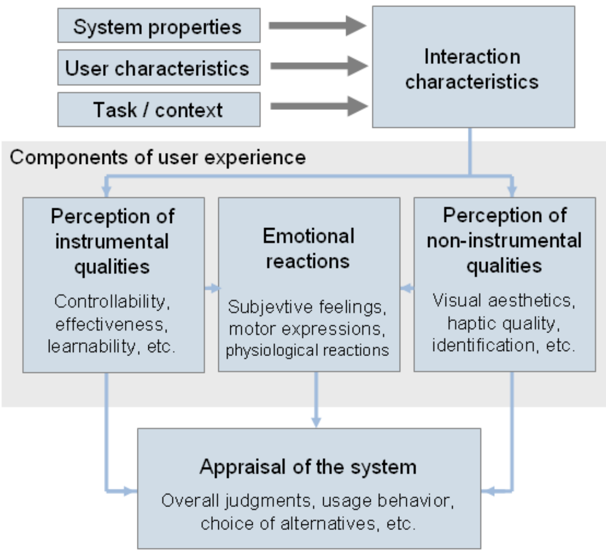

Figure 2.

Figure 3.

Figure 4.

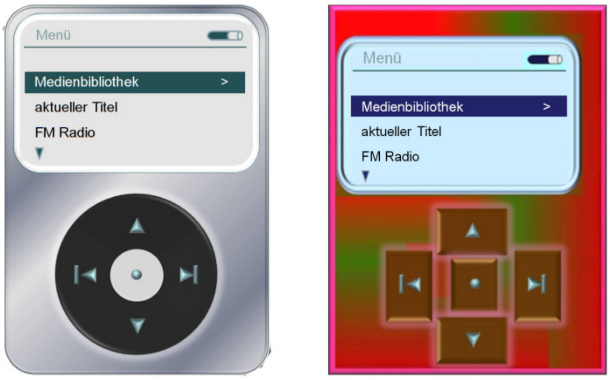

Figure 5.

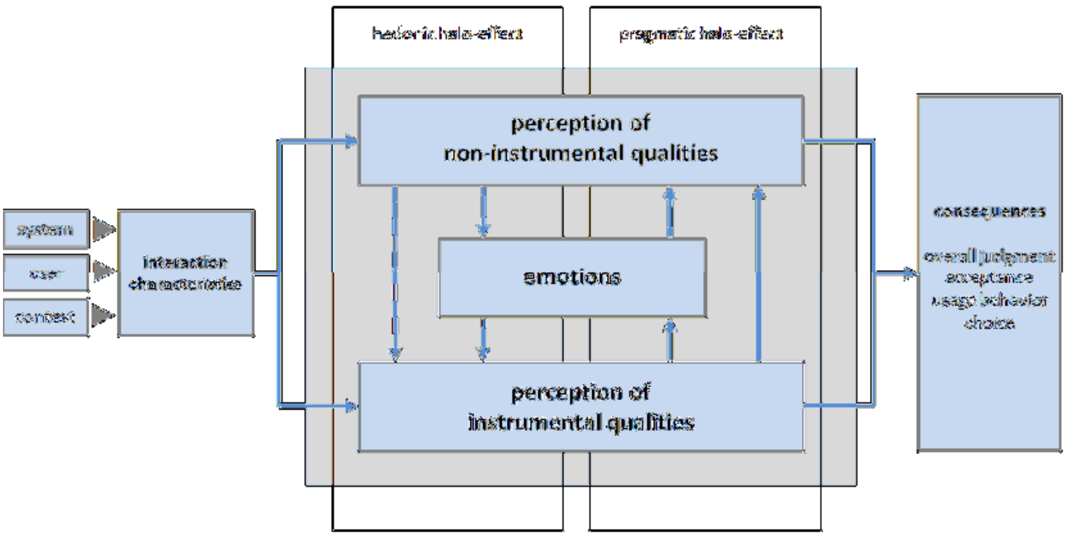

TABLE CAPTIONS

Table 1. Mean scores and standard deviations of perceived visual attractiveness.

Table 2: Mean scores and standard deviations of perceived usability.

Table 3: Mean scores and standard deviations of arousal and valence measuring emotional reactions.

Table 1.

	Aesthetics low				Aesthetics high			
	Usability low		Usability high		Usability low		Usability high	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
T1	2.53	(1.51)	2.00	(1.00)	4.00	(1.41)	4.33	(2.00)
T2	2.93	(1.58)	3.00	(1.56)	3.27	(1.10)	4.33	(1.50)
T3	2.87	(1.77)	3.07	(1.28)	2.93	(1.16)	4.40	(1.50)

Table 2.

	Aesthetics low				Aesthetics high			
	Usability low		Usability high		Usability low		Usability high	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
T1	3.80	(1.70)	4.53	(1.36)	4.80	(1.37)	5.40	(1.12)
T2	2.33	(1.54)	4.40	(1.30)	3.07	(1.16)	5.33	(1.50)
T3	2.87	(1.69)	4.87	(1.64)	2.60	(1.30)	5.27	(1.62)

Table 3.

Arousal	Aesthetics low				Aesthetics high			
	Usability low		Usability high		Usability low		Usability high	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
T1	6.53	(1.36)	6.60	(1.50)	5.67	(1.95)	6.60	(1.92)
T2	6.33	(1.40)	5.73	(1.39)	5.87	(1.19)	5.20	(1.82)
T3	5.47	(1.51)	5.53	(1.36)	5.33	(1.36)	4.20	(2.15)

Valence	Aesthetics low				Aesthetics high			
	Usability low		Usability high		Usability low		Usability high	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
T1	2.80	(1.57)	3.80	(1.47)	3.80	(1.47)	2.93	(1.16)
T2	3.20	(1.70)	4.93	(1.67)	3.87	(1.64)	4.87	(2.20)
T3	3.20	(2.11)	5.00	(1.77)	3.87	(1.96)	4.33	(2.90)