INTERNATIONAL JOURNAL OF PSYCHOLOGY, 2007, 42 (4), 253–264

Psychology Press
Taylor & Francis Group

Usability, aesthetics and emotions in human–technology interaction

Manfred Thüring and Sascha Mahlke

Berlin University of Technology, Berlin, Germany

In the past, research on human–technology interaction has almost exclusively concentrated on aspects of usefulness and usability. Despite the success of this line of research, its narrow perspective has recently become a target for criticism. To explain why people prefer some systems over others, factors such as aesthetic qualities and emotional experiences play an important role in addition to instrumental aspects. In the following, we report three experiments that illustrate the importance of such factors. In the first experiment, we study the role of emotions in human–technology interaction by using Scherer's (1984) component theory of emotions as a theoretical foundation. A combination of methods is derived from that theory and employed to measure subjective feelings, motor expressions, physiological reactions, cognitive appraisals, and behaviour. The results demonstrate that the manipulation of selected system properties may lead to differences in usability that affect emotional user reactions. The second experiment investigates the interplay of instrumental and non-instrumental system qualities. The results show that users' overall appraisal of a technical device is influenced by both groups of qualities. In the third experiment, we join the approaches of the first two studies to analyse the influence of usability and aesthetics within a common design. The results indicate that systems differing in these aspects affect the perception of instrumental and non-instrumental qualities as well as the users' emotional experience and their overall appraisal of the system. Summarizing our results, we present a model specifying three central *c*omponents of *u*ser *e*xperience and their interrelations (CUE-Model). The model integrates the most important aspects of human–technology interaction and hints at a number of interesting issues for future research.

Dans le passé, la recherche sur l'interaction entre l'humain et la technologie s'est presque exclusivement concentrée sur les aspects d'emploi et d'utilité. Malgré son succès, ce champ de recherche présente une perspective étroite qui est devenue récemment l'objet de critique. En plus des aspects instrumentaux, des facteurs comme les qualités esthétiques et les expériences émotives jouent un rôle important pour expliquer pourquoi les gens préfèrent certains systèmes mieux que d'autres. Dans ce qui suit, nous rapportons trois expérimentations qui illustrent l'importance de ces facteurs. Dans la première expérimentation, nous étudions le rôle des émotions dans l'interaction entre l'humain et la technologie en utilisant le modèle composant des émotions de Scherer (1984) comme cadre théorique. Une combinaison de méthodes est dérivée de ce modèle théorique et utilisée pour mesurer les sentiments subjectifs, les expressions motrices, les réactions physiologiques, les évaluations cognitives, et le comportement. Les résultats démontrent que la manipulation des propriétés du système qui ont été choisies peut causer des différences dans l'emploi qui affecte les réactions émotives de l'usager. La deuxième expérimentation examine l'interaction entre les qualités instrumentales et non instrumentales du système. Les résultats indiquent que l'évaluation globale des usagers d'un appareil technique est influencée à la fois par les deux groupes de qualités. Dans la troisième expérimentation, nous combinons les approches des deux premières études afin d'analyser l'influence de l'utilité et de l'esthétique au sein d'un design commun. Les résultats indiquent que les systèmes qui diffèrent par ces aspects affectent la perception des qualités instrumentales et non instrumentales ainsi que l'expérience émotive des usagers et leur évaluation globale du système. Pour résumer nos résultats, nous présentons un modèle qui spécifie trois composantes centrales de l'*e*xpérience de l'*u*sager et leurs interrelations (Le modèle CEU ou CUE-Model en anglais). Le modèle intègre les aspects les plus importants de l'interaction entre l'humain et la technologie et suggère quelques sujets intéressants pour la recherche future.

Correspondence should be addressed to Manfred Thüring, Berlin University of Technology, Institute of Psychology and Ergonomics, Franklinstr. 28-29 (FR 2-6), 10587 Berlin, Germany (thuering@gp.tu-berlin.de).

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 Michael Minge and Andreas Wand for their work on the studies and two anonymous reviewers for many clarifying and helpful comments.

http://www.psypress.com/ijp DOI: 10.1080/00207590701396674

En el pasado, los estudios de la interacción humano–tecnología se centraban casi exclusivamente en el aspecto de utilidad y usabilidad. A pesar de los logros de esta línea de investigación últimamente se ha criticado su estrecha perspectiva. Para explicar el por qué las personas prefieren unos sistemas sobre otros hay que tener en cuenta factores como cualidades estéticas y la experiencia emocional, la cual juega un rol importante a parte de los aspectos instrumentales. En este estudio se presenta tres experimentos, los cuales reflejan la importancia de estos factores. En el primero de ellos se estudia el rol de las emociones en la interacción humano–tecnología utilizando como base la teoría de componentes de las emociones de Scherer (1984). De esta teoría se deriva una combinación de métodos empleados para medir los sentimientos subjetivos, expresiones motoras, reacciones fisiológicas, valoraciones cognitivas y comportamiento. Los resultados demuestran que la manipulación de unas propiedades selectivas del sistema puede conducir a diferencias en usabilidad, lo cual afecta las reacciones emocionales del usuario. El segundo experimento estudia la interacción entre las cualidades instrumentales y no instrumentales del sistema. Los resultados demuestran que las valoraciones totales de los dispositivos técnicos son influidas por ambos grupos de cualidades. En el tercer experimento hemos juntado las aproximaciones de los dos primeros estudios para analizar la influencia de la usabilidad y estética dentro de un diseño común. Los resultados indican que los sistemas que difieren en estos aspectos afectan la percepción de las cualidades instrumentales y no instrumentales junto con la experiencia emocional del usuario y la valoración total del sistema. Resumiendo nuestros resultados, presentamos un modelo especificando tres componentes centrales de la experiencia del usuario y sus interrelaciones (el modelo CUE). El modelo integra los aspectos más importantes de la interacción humano–tecnología y señala numerosas e interesantes investigaciones para el futuro.

"Computing is not about computers anymore. It is about living." This claim by Nicholas Negroponte[1] (1995, p. 6) was made more than 10 years ago, and illustrates how vigorously technology impacts on all aspects of modern life. The usage of computers and other technical devices has become both a necessity and a matter of course for almost everyone in today's industrial societies. Hence, it comes as no surprise that a growing amount of research in engineering, computer science, and the humanities has engaged in investigating human–technology interaction.

From the beginning, pragmatic issues such as the utility and usability of technical systems have dominated this research (Nielsen, 1993). Usability, in particular, has served as a key concept for capturing the "quality of use" of interactive systems (Bevan, 1995), inspiring many investigations on the effectiveness and efficiency of system employment. Mostly, performance-based methods were chosen to assess these two usability components. Effectiveness was studied in experiments in which the accuracy and completeness of reaching the goals of a predefined task were measured. Efficiency, on the other hand, was captured by relating the effectiveness of system usage to its costs in terms of effort or time. The third component of the usability concept is user satisfaction (International Organization for Standardization, 1998), and most approaches to capture it used subjective judgments, which again were mostly based on the efficiency and effectiveness of system usage (Lindgaard & Dudek, 2003).

But is usability all that matters in human–technology interaction? For instance, Dillon (2001) proposed that user satisfaction is likely to be influenced by factors such as personal experience with technology, preferred working style, and the aesthetics of system design. Such quality aspects seem to be important for users but are not connected to their performance with the system. Besides, it is important to know how people feel during system usage. Are they empowered, annoyed, frustrated, confident, unsure, or wary? This issue refers to the emotional side of user experience, a side that has been neglected by research on human–technology interaction.

To summarize, while effectiveness, efficiency, and satisfaction are definitely important determinants of human–technology interaction, other aspects, such as the aesthetics of system design and emotional experiences during system usage, certainly impact on the perceived quality of use as well (Hassenzahl, 2006). Like others (e.g., Hassenzahl & Tractinsky, 2006), we therefore argue for a broader perspective that regards user experiences as a compound of three basic elements:

1. the perception of instrumental qualities, such as the controllability or the effectiveness of a system,
2. the perception of non-instrumental qualities, such as visual aesthetics or haptic quality, and
3. the user's emotional responses to system behaviour.

[1]Nicholas Negroponte is founder and former chairman of the Media Lab at the Massachusetts Institute of Technology.

In the following, we report three experiments investigating these components, their interrelations, and their dependency on system features and interaction characteristics. The first experiment is concerned with the relation between usability and emotions during system usage. The second experiment addresses the influence of both usability features and aesthetics on the perception of instrumental and non-instrumental qualities. The third experiment integrates the approaches of the first two studies and aims at providing a comprehensive view on usability, aesthetics, and emotional user reactions.

STUDY 1: USABILITY AND EMOTIONS

Do systems of different usability influence emotional experiences during human–technology interaction? A variety of methods can be employed to answer this question (Picard, 1997). They range from physiological measures, such as heart rate and electrodermal activity (EDA), electromyography (EMG), or pupil responses, to various kinds of survey methods, like questionnaires and interview techniques. In order to choose adequate methods for capturing emotional responses during human–technology interaction, a sound theoretical foundation is required.

A number of psychological theories emphasize the multifaceted character of emotions. Scherer (1984) defines emotions as consisting of five aspects or components (see Figure 1). The ‘‘emotion triad’’ proposed by Izard (1977) is central to his model. It comprises subjective feelings, physiological activation, and motor expressions, and is connected to two other components, i.e., cognitive appraisals and behavioural tendencies.

Since all components of Scherer's model might play a role in human–technology interaction, Study 1 investigates to which extent each of them is influenced when systems of different usability are employed.

Method

Participants. Thirty individuals (half of them women) participated in the study. They were between 20 and 41 years old ($M = 25.9$, $SD = 3.9$).

Stimuli. Two versions of a computer-based simulation of a mobile phone were designed to induce two different degrees of perceived usability. While the well-designed version was highly usable, the ill-designed one had several usability flaws hampering the solution of particular tasks. The menu options were structured in a confusing way and the buttons were arranged in an unusual order. The difference in usability was secured by comparing usability ratings and performance indicators of both versions in a pre-test with eight participants. In all other respects, both versions were identical.

Design. The independent variable of the experiment was the factor usability, consisting of two treatments (well-designed vs. ill-designed). It was assumed that the system with usability flaws would lead to negative emotional experiences while the other one would be experienced as positive or neutral. Hence, the two versions should lead to differences with respect to the five emotional components under investigation.

Dependent variables. Several methods were used to gain information on the components. To measure subjective feelings, participants filled in the Self-Assessment Manikin (SAM) by Lang

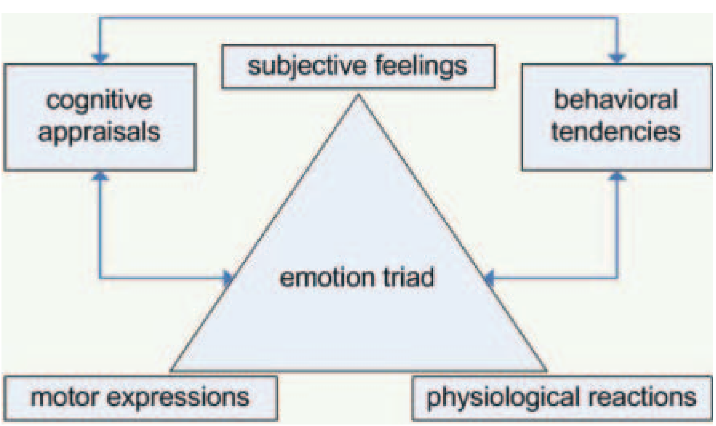

Figure 1. The component model of emotions according to Scherer (1984).

(1980), which is based on the dimensions of valence and arousal (Russell, 1980). SAM consists of a 9-point graphical item ranging from 1 to 9 for each dimension valence and arousal. Heart rate and EDA served as indicators for physiological reactions (Ward & Marsden, 2003). To measure facial expressions, EMG responses were recorded from the *zygomaticus major* and *corrugator supercilii* muscle sites, which control smiling and frowning, respectively (Partala & Surakka, 2004). With respect to behavioural tendencies, the time required for input operations was recorded as an indicator for the effectiveness of system usage. To collect data on cognitive appraisals, participants filled in a short form based on the Geneva appraisal questionnaire by Scherer (2001), which addresses five dimensions: intrinsic pleasantness, novelty, goal/need conduciveness, coping potential, and norm/self compatibility. One item that was applicable in the domain of interactive systems was taken from the appraisal questionnaire for each dimension. The 5-point items ranged from 1 to 5 (*not at all* to *extremely*).

Procedure. The experiment took 75 minutes on average. At the beginning, electrodes for measuring physiological reactions and facial expressions were attached, and baseline values were recorded for 2 minutes. The participants started with one system and completed a first group of tasks. Then they switched to the other version to solve five remaining tasks. The presentation order of versions and tasks was balanced. Heart rate, EDA, and EMG were measured during task completion. The interaction with the system was recorded to analyse users' behaviour. After each task, participants assessed their affective state with the SAM scales. After finishing all the tasks of a set, everyone answered the Geneva appraisal questionnaire. To ensure a realistic emotional involvement, participants were paid depending on their performance, and immediate feedback on their achievements was provided after each task.

Results

The results of the experiment are summarized in Table 1. For analysing our data, we always applied *t*-tests with usability as the independent variable. Significant differences were found between the two versions on the valence and arousal dimensions of the SAM questionnaire, $t(29) = -4.51$, $p < .001$ and $t(29) = 7.90$, $p < .001$. The well-designed system was rated with positive valence and as less arousing, while the ill-designed system got higher scores on the arousal dimension and lower valence scores.

TABLE 1
Mean scores and standard deviations on all dependent variables for levels of usability

	Well-designed version		*Ill-designed version*	
Component & Dependent variable	*M*	*SD*	*M*	*SD*
Subjective feelings				
SAM—valence (1–9)	6.6	1.2	3.8	1.7
SAM—arousal (1–9)	4.1	1.5	5.4	1.5
Physiological reactions				
EDA [µS]	2.1	1.9	2.5	2.0
Heart rate [bpm]	1.9	5.2	2.2	5.9
Motor expressions				
EMG—corrugator supercilii	49.0	4.4	52.0	3.5
EMG—zygomaticus major	47.6	3.6	51.6	4.1
Cognitive appraisals				
Pleasantness (1–5)	3.8	0.8	2.1	0.7
Novelty (1–5)	1.6	0.7	2.7	1.1
Goal relevance (1–5)	3.6	1.1	1.8	0.7
Coping potential (1–5)	3.9	0.8	2.2	0.7
Norm/self compatibility (1–5)	4.0	0.8	2.5	0.9
Behaviour intention				
Time per input [s]	1.7	0.4	3.1	1.2

EDA measures were higher for the ill-designed system, $t(29) = 2.64$, $p<.05$, while the heart rate values did not differ, $t(29) = -0.41$, $p = .68$. The EMG showed heterogeneous results. As expected, the activity of the *corrugator supercilii* was significantly higher for the ill-designed version, $t(29) = -2.19$, $p < .05$ but, contrary to our assumptions, the activity of the *zygomaticus major* was significantly higher as well, $t(29) = -2.96$, $p < .01$.

Regarding cognitive appraisals, significant differences between the versions were found on all five dimensions. The well-designed system was experienced as more pleasant, goal conducive, capable, and norm/self compatible as well as less novel than the ill-designed version. The average time required per input was significantly higher for the ill-designed system, $t(26) = -6.16$, $p < .001$, indicating that single operations performed with this version took longer than a single operation conducted with the well-designed one.

Discussion

Different emotional patterns were found for the well-designed and the ill-designed system. The well-designed version led to more positive and less arousing subjective feelings. Physiological measures showed lower EDA values and less activity of the *corrugator supercilii* when usability was high. A complementary constellation was found for the ill-designed system.

Our results regarding the activity of the *zygomaticus major* differ from other studies, which found higher activity in relation to positive emotions (Partala & Surakka, 2004). Instead, our data point in the same direction as experiments that detected high activity of the *zygomaticus major* for negative emotions (Lang, Greenwald, Bradley, & Hamm, 1993). Hence, it seems that the activity of the *zygomaticus major* is not well suited to discriminate between positive and negative feelings, although it might be a strong indicator for emotional responses in general. Similarly, the difference between the two versions regarding heart rate was not significant. For durations of system usage as long as in our experiment, EDA seems to be more sensitive to arousal than heart rate.

Appraisal processes of emotions in interactive contexts have scarcely been investigated in other experiments. In our study, the well-designed system was experienced as more positive on all appraisal dimensions than the ill-designed version. Goal conduciveness and coping potential in particular can be attributed directly to high usability. Interestingly, a low degree of novelty was associated with more positive experiences. This relationship may contribute to users' reluctance when new systems are introduced. With respect to behavioural tendencies, our study indicates a noteworthy aspect concerning the efficiency of system usage. Since the average time required per input was significantly higher for the system with usability flaws, negative emotions may contribute to slowing down the user. Of course, further research is required to strengthen this assumption.

STUDY 2: USABILITY AND AESTHETICS

In Study 1, we found an influence of usability on users' emotional reactions. But which other types of system qualities may determine the way users experience the interaction with a system and which interaction characteristics may shape their overall judgment? Rafaeli and Vilnai-Yavetz (2004) presented a model suggesting that the appreciation of any artifact—and hence of any technical system as well—is related to three conceptually distinct aspects: instrumentality, symbolism, and aesthetics. While instrumentality corresponds to pragmatic features, such as usefulness and usability, symbolism and aesthetics represent two categories that are independent from instrumental values. Symbolism refers to the meanings and associations a product elicits in the minds of its users. Aesthetics, on the other hand, refer to the sensual experience a product entails, and to the extent to which this experience fits individual goals and preferences.

Tractinsky, Katz, and Ikar (2000) address another important issue with respect to aesthetics. In their study, they investigated the relationship between perceived usability and visual attractiveness and found that the two were related. Their results suggest that the perception of usability is influenced by the aesthetics of an interactive product. Lindgaard and Dudek (2003) also investigated this relationship, but failed to find a clear connection between these two system features.

Taking theses approaches as a starting point, Study 2 investigates if variations in usability and visual aesthetics influence the perception of instrumental and non-instrumental qualities independently and analyses their contribution to the overall appraisal of interactive products.

Method

Participants. Fifty-six individuals (half of them women) participated in the study. They were between 20 and 38 years old (M = 27.0, SD = 3.7).

Stimuli. Portable digital audio players were chosen as the domain of study and simulated on a computer. To produce two versions with different impact on the perceived instrumental qualities, three system features were varied: the number of simultaneously discernible menu lines (five or two), a cue hinting at available but hidden menu items (present or not), and an indicator of the actual position in the menu hierarchy (given or not). The resulting displays are shown in Figure 2. In a pre-test with 10 participants, it was assured that interaction characteristics were generated which affected the usability of the systems differently, i.e., the first version was of higher usability in terms of performance and subjective ratings than the second one.

In another pre-test, participants ranked seven different-looking players with respect to their attractiveness. The body designs were randomly taken from existing products. The two versions that received extreme rankings (*looking best* vs *looking worst*) were chosen for the experiment (see Figure 3).

Figure 2. Variation of usability used in Study 2 (high usability on the left, low usability on the right).

The combination of usability and aesthetics features yielded four distinct versions, which were presented on a hand-held 7" TFT-display equipped with touch screen functionality for receiving input.

Design. A 2 × 2 factorial design was applied in the experiment with the independent variables usability and visual aesthetics. Since each had two treatments (high and low), four combinations could be tested:

1. high usability and high aesthetics,
2. high usability and low aesthetics,
3. low usability and high aesthetics,
4. low usability and low aesthetics.

All participants used and rated two versions of the system, either (1) and (4) or (2) and (3), according to a Latin Square plan for repeated measures (Winer, 1971). We assumed that the factor usability would effect the perception of instrumental qualities, while the factor visual aesthetics would influence the perception of non-instrumental qualities. Both factors should impact on overall judgments.

Dependent variables. Task completion rates and time on task were recorded to ensure that versions of assumed high or low usability differed as planned. Questionnaires were employed to assess the users' perception of instrumental and non-instrumental qualities. Selected subdimensions of the Subjective Usability Measurement Inventory (SUMI; Kirakowski & Corbett, 1993) served to rate instrumental qualities (controllability, effectiveness, helpfulness, learnability). Each dimension consisted of four items. The answering format had three options: disagree, undecided, agree (0 to 2). Scale values were computed for each dimension (Cronbach's alpha .79 for controllability, .75 for effectiveness, .67 for helpfulness, .70 for learnability). Scale values of the four dimensions were added to receive an overall usability rating (range from 0 to 8). One dimension of a questionnaire developed by Lavie and Tractinsky (2004) was used to measure visual aesthetics. The dimension consisted of five 7-point items ranging form 0 to 6 (*disagree* to *agree*). A scale value was obtained by averaging over the five items (Cronbach's alpha .86). The global dimension of the SUMI served for the overall ratings. The four items had the same answering format as the other SUMI items. Scale values were obtained by averaging over the four items (Cronbach's alpha .82).

Figure 3. Variation of visual aesthetics used in Study 2 (high aesthetics on the left, low aesthetics on the right).

Procedure. The experiment took about 45 minutes. Presentation order of the players was counterbalanced. For each system, participants rated the visual aesthetics of the player after it was introduced. Then they had 10 minutes to complete the tasks of a first set before the systems were changed, and the tasks of a second set had to be accomplished in another 10 minutes. The interaction with the system was recorded to analyse user performance. Afterwards, participants filled in the questionnaire on perceived usability and gave an overall rating.

Results

To analyze the data of our Latin Square design for repeated measures, we used mixed linear models according to Winer (1971). Analyses with the independent variables usability and visual aesthetics were conducted for testing the effects on all dependent variables (see Table 2). The behavioural data confirmed the result of the pre-test by showing that the performance with the well-designed systems was better than the performance with the ill-designed systems. We found a significant effect for usability in the predicted direction for the number of accomplished tasks, $F(1, 106) = 9.6$, $p < .001$, as well as for the average time on task, $F(1, 105) = 10.5$, $p < .01$.

An analysis of the ratings of perceived usability showed a highly significant main effect for the factor usability, $F(1, 104) = 23.6$, $p < .001$. For the effect of the factor visual aesthetics on perceived usability, we found a trend, $F(1, 104) = 3.7$, $p < .10$. With respect to perceived visual aesthetics, the ratings for the aesthetically well-designed version were higher than those for the ill-designed version, $F(1, 106) = 14.2$, $p < .001$. The factor usability had no effect on perceived visual aesthetics.

An analysis of the overall judgments showed a highly significant main effect for both factors, usability, $F(1, 105) = 9.6$, $p < .01$, and visual aesthetics, $F(1, 105) = 15.8$, $p < .001$, but no interaction. A regression analysis revealed that the variables perceived usability and perceived visual aesthetics predicted 60% of the variance of the overall judgments, with a greater impact of perceived usability (beta weight .68 for perceived usability versus .24 for perceived visual aesthetics).

Discussion

The behavioural data indicate that the variation of the three selected system features produced different interaction characteristics and substantially influenced the actual usability of the audio players. Although the usability flaws were rather minor, users perceived the difference and rated the versions accordingly. Similarly, the variation of the body design of the players apparently influenced the appeal of the players, leading to less favourable ratings for the less attractive version.

The data revealed a trend for an influence of the factor aesthetics on the perceived usability rating, but showed no influence of the factor usability on attractiveness ratings. This result points in the same direction as the study by Tractinsky et al. (2000), but more data are required to clarify the connection between perceived usability and aesthetics definitively.

Together, the ratings of usability and aesthetics show that the perception of instrumental and non-instrumental qualities are affected by corresponding system features. Moreover, we found an influence of both instrumental and non-instrumental quality perceptions on the overall appraisal of the system. The result of our regression analysis is compatible with the results of other studies (Hassenzahl, 2003; Lindgaard & Dudek, 2003), and demonstrates that a user's judgment of a system relies on both pragmatic and aesthetic features. The influence of perceived usability on the overall appraisal was found to be higher than that of aesthetics. Hassenzahl (2003) argued that the weight of instrumental and non-instrumental quality perceptions depend on situational factors. The fact that our participants were paid according to their performance might be the reason for the higher influence of perceived usability on their overall appraisal.

The results of Study 1 and Study 2 indicate that emotions as well as the perception of two types of qualities are influenced by particular system features and contribute to the overall user experience. If all three components are indeed

TABLE 2
Mean scores and standard deviations on all dependent variables for levels of usability and aesthetics

	Low usability				*High usability*			
	Low aesthetics		*High aesthetics*		*Low aesthetics*		*High aesthetics*	
Component & Dependent variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Performance measures								
No. of accomplished tasks	4.1	1.1	3.9	1.2	4.4	1.2	4.8	1.0
Average time on task (s)	47.2	19.2	49.4	22.1	38.3	22.1	31.6	19.1
Quality perceptions								
Perceived usability (0–8)	2.5	1.8	2.8	1.9	4.0	2.2	5.2	2.7
Perceived visual aesthetics (0–6)	2.9	1.1	4.6	0.7	3.5	1.1	4.8	0.8
Overall judgments								
Global rating (0–2)	0.7	0.7	1.1	0.8	1.0	0.7	1.7	0.7

central for this experience, it should be possible to influence them together. Our third study was designed to test this assumption.

STUDY 3: USABILITY, AESTHETICS, AND EMOTIONS

Do differences in usability and aesthetics influence the perception of instrumental and non-instrumental qualities in a way that is consistent with users' emotional experience and their overall appraisal of a technical device? This issue is investigated in Study 3 by varying usability and design features of an interactive system and by analysing their impact on quality perceptions, emotions, and overall appraisals.

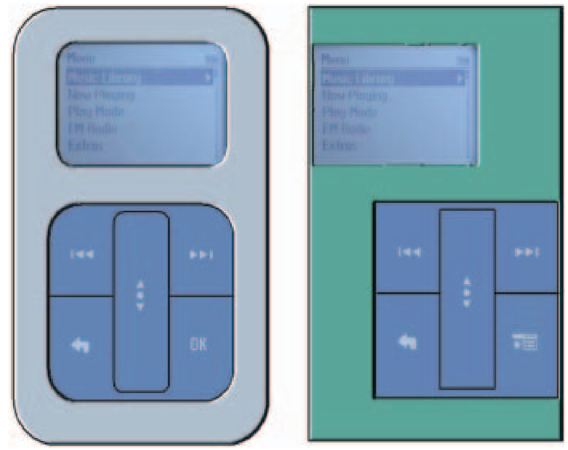

Figure 4. Variations of visual aesthetics used in Study 3 (high aesthetics on the left, low aesthetics on the right).

Method

Participants. Forty-eight individuals (half of them women) participated in the study. They were between 20 and 34 years old ($M = 25.5$, $SD = 3.6$).

Stimuli. Again, portable digital audio players were chosen as the application domain, and the same tasks and variations of usability were used as in Study 2. With respect to system features capable of influencing the perception of non-instrumental qualities, we manipulated the visual aesthetics by focusing on design dimensions derived from Leder and Carbon (2005) as well as Han, Kim, Yun, Hong, and Kim (2004): symmetry (high or low), colour combination (high or low colour differences), and shape (rounded or angular). A pre-test secured the distinctiveness of the two versions resulting from the combination of these features (see Figure 4).

Design. The experimental design was the same Latin Square plan for repeated measures as in Study 2. We assumed that the versions of higher usability and more appealing design would influence the perceptions of attractiveness and usability and result in more positive emotional reactions and more favourable overall judgments.

Dependent variables. We used the same behavioural measures and questionnaires to assess the user's perception of usability (Cronbach's alpha .83 for controllability, .82 for effectiveness, .70 for helpfulness, .70 for learnability) and visual aesthetics (Cronbach's alpha .76) as in Study 2. As in Study 1, subjective data regarding emotional user reactions was measured with the Self-Assessment Manikin (SAM). EDA and heart rate were selected as physiological measures. EMG data of the *zygomaticus major* and the *corrugator supercilii* were recorded. Two techniques served to measure overall judgments: the global dimension of the SUMI (Cronbach's alpha .84) and a ranking of the player versions.

Procedure. The experiment lasted 60 minutes on average. At the beginning, baseline values for heart rate, EDA and EMG were recorded for 2 minutes. The participants started with one player version and completed a first set of tasks. Then they switched to the other version to accomplish the tasks of a second set. Before working on the tasks, subjects rated the visual aesthetics of the version. Behavioural data, heart rate, EDA, and EMG were measured during task completion. After finishing a task, participants filled in the SAM scales. When all tasks of a set were solved, the usability of the system was rated, and at the end of the experiment participants ranked the two versions they had used.

Results

As in Study 2, mixed linear models analyses (Winer, 1971) with the independent variables usability and visual aesthetics served to test our hypotheses for all dependent variables. All results are summarized in Table 3. With respect to the behavioural data, the findings of Study 2 were replicated. The two versions of different usability yielded differences in central interaction characteristics. We found a significant main effect of usability for both the number of accomplished

TABLE 3
Mean scores and standard deviations on all dependent variables for levels of usability and aesthetics

	Low usability				*High usability*			
	Low aesthetics		*High aesthetics*		*Low aesthetics*		*High aesthetics*	
Component & Dependent variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Performance measures								
No. of accomplished tasks	3.8	1.3	3.8	1.2	4.9	0.5	4.9	0.3
Average time on task (s)	47.0	24.3	46.6	20.1	25.0	13.2	22.7	11.4
Quality perceptions								
Perceived usability (0–8)	3.1	2.0	4.0	2.3	6.6	1.0	6.6	1.5
Perceived visual aesthetics (0–6)	2.2	1.2	4.1	1.2	2.7	1.5	3.9	1.0)
Subjective feelings								
SAM—valence (1–9)	4.3	1.8	4.7	1.9	6.2	1.2	7.1	1.5
SAM—arousal (1–9)	5.8	1.4	5.2	1.7	4.4	1.4	3.8	1.5
Physiological reactions								
EDA [µS]	14.9	15.4	9.1	16.2	0.6	8.1	0.9	8.2
Heart rate [bpm]	0.7	7.4	− 6.8	13.8	− 1.4	11.3	2.9	16.1
Motor expressions								
EMG—corrugator supercilii	7.4	15.1	7.4	15.9	2.4	15.8	2.2	15.2
EMG—zygomaticus major	1.0	11.8	0.7	11.1	4.5	15.1	2.9	13.9
Overall judgments								
Global rating (0–2)	0.5	0.5	0.8	0.7	1.4	0.4	1.6	0.5

tasks, $F(1, 92) = 92.2$, $p < .001$, as well as for the average time on task, $F(1, 83) = 44.5$, $p < .001$. Compared to the system of lower usability, the highly usable system led to a greater percentage of correct solutions and to a faster performance.

The analyses of usability and aesthetics ratings showed that our variations of system properties entailed the predicted differences in users' quality perceptions. With respect to perceived instrumental qualities, a significant difference was found for the factor usability in the mean ratings based on the SUMI questionnaire, $F(1, 92) = 62.7$, $p < .001$. With respect to the perception of non-instrumental qualities, there was a significant effect of the factor visual aesthetics, $F(1, 85) = 55.2$, $p < .001$. No other significant differences were found.

The analyses of subjective emotional data revealed significant main effects for the factors usability and visual aesthetics on the dimensions valence and arousal: usability: valence, $F(1, 90) = 38.7$, $p < .001$; arousal, $F(1, 78) = 19.2$, $p < .001$; aesthetics: valence, $F(1, 90) = 4.7$, $p < .05$; arousal, $F(1, 78) = 5.5$, $p < .05$. Lower usability and visual aesthetics led to less positive valence and to higher arousal. No interaction effects were found.

The physiological data partially underlined these results. We found a significant effect of the factor usability on EDA, $F(1, 89) = 17.8$, $p < .001$, but no effect on the heart rate. EDA was higher in the case of low usability. With respect to users' facial expressions, we found a statistical trend for the factor usability concerning the activity of the *corrugator supercilii*, $F(1, 89) = 2.8$, $p < .10$. Activity tended to be higher in the low usability conditions. No significant differences were found for the activity of the *zygomaticus major*.

The ratings on the global dimension of the SUMI showed a significant main effect of the factor usability, $F(1, 89) = 69.5$, $p < .001$, and a trend for the factor visual aesthetics, $F(1, 89) = 3.2$, $p < .10$. All participants preferred the highly usable and attractive version to the version of lowest usability and attractiveness. For mixed combinations, 71% of the participants preferred the system of high usability and low aesthetics to the system of low usability and high aesthetics, while 29% favoured the reverse combination.

Discussion

The rationale underlying the experiment was to vary selected system features in order to produce distinctive interaction characteristics leading to different perceptions of instrumental and non-instrumental qualities, which in turn should cause different emotional reactions and corresponding overall judgments. If the perception of instrumental and non-instrumental together with emotional reactions constitute the major components of user experience, an experiment following this rationale should produce patterns of behavioural data, ratings, and physiological measures that are internally consistent but also distinct for systems of different usability and aesthetics.

Did we find such patterns? First of all, variations of system properties with respect to usability as well as to aesthetics had the predicted impact on the perception of both types of qualities. Systems with features leading to a high degree of usability and attractiveness received better ratings than their impaired counterparts. The results of the SAM questionnaire were consistent with these findings and showed corresponding differences for the subjective feelings of our participants. They also revealed that the effect of usability was greater than the one of visual aesthetics for both valence and arousal. Consequently, the system of high usability and appealing design was experienced as most satisfying, while the system of low usability and least attractiveness was regarded as most annoying. The EMG data and other physiological measures support this interpretation. It must be noted, though, that we did not find the expected differences in all measured variables. The overall judgments pointed in the same direction as the ratings of perceived qualities and emotions, and again revealed a greater impact of usability on the overall appraisal of the systems. In summary, it can be concluded that our experimental variations produced the desired effects in terms of distinct patterns of behavioural data, physiological responses, and ratings.

GENERAL DISCUSSION AND IMPLICATIONS FOR FUTURE RESEARCH

Our studies support the notion of user experience as a compound of emotions and perceptions of instrumental as well as non-instrumental qualities. To summarize our results, we propose a model that integrates these components into a common framework. We call this framework CUE-Model, where CUE stand for *c*omponents of *u*ser *e*xperience (see Figure 5).

User experience is gained in the course of interacting with a technical device. Usually, this interaction aims at solving a particular task, takes place in a certain context, and extends over a limited period time. Attributes of the user—such as knowledge or skills—as well as features of the system—such as functionality and interface design—affect the interaction and determine its major characteristics. Since these characteristics

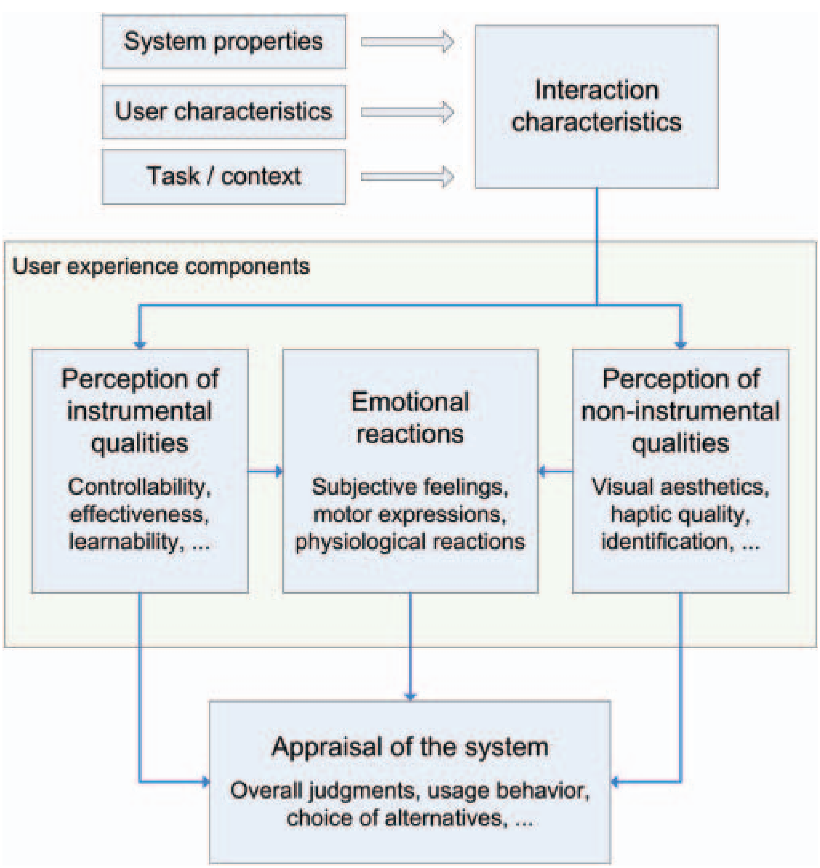

Figure 5. The CUE-Model (*c*omponents of *u*ser *e*xperience).

are frequently encountered throughout system usage, we assume that they are perceived as inherent qualities of the system.

The CUE-Model distinguishes two types of such qualities. Instrumental qualities concern the experienced support the system provides and the ease of its use. Features such as the controllability of the system and the effectiveness of its functionality fall into this category. Non-instrumental qualities, on the other hand, concern the look and feel of the system. Features such as visual aesthetics or haptic quality belong to this class. Hence, while instrumental qualities are closely related to the usability and usefulness of a system, non-instrumental qualities result from its appeal and attractiveness.

The perception of both types of qualities is likely to influence the third component of user experience, i.e., the emotions that accompany the interaction process. For example, sluggish system responses may affect perceived effectiveness and lead to impatience or even to frustration and anger. In contrast, a futuristic and innovative design may impact on perceived visual aesthetics and cause surprise, curiosity, or pleasure. In accordance with Scherer (1984), the CUE-Model characterizes emotions as episodes of subjective feelings accompanied by specific physiological reactions and expressive behaviour. Such episodes may occur repeatedly and shape the user's emotional experience with a technical device. Finally, all three components of user experience should impact on the overall appraisal of the system and thus influence the user's future decisions and behaviour.

For reasons of simplicity, the CUE-Model only specifies relations that can be justified on the basis of our experiments. Therefore, it assumes that the relationships between the components are one-directional (e.g., there is no mutual influence of the two types of quality perceptions). Of course, these assumptions are preliminary since mutual influences as well as feedback loops may exist. We will address these possibilities in our future research to increase the specificity of the CUE-Model.

Further research topics concern the temporal characteristics of human–technology interaction as proposed by Hassenzahl and Sandweg (2004), as well as user characteristics, parameters of the situation, and further dimensions of non-instrumental qualities as discussed by Mahlke (2006). Together, all these topics offer demanding opportunities to reach beyond 'classical' usability approaches in order to gain a more comprehensive view on human–technology interaction than we have today.

REFERENCES

Bevan, N. (1995). Measuring usability as quality of use. *Software Quality Journal*, *4*, 115–130.

Dillon, A. (2001). Beyond usability: Process, outcome and affect in human–computer interactions. *Canadian Journal of Library and Information Science*, *26*, 57–69.

Han, S. H., Kim, K. J., Yun, M. H., Hong, S., & Kim, J. (2004). Identifying mobile phone design features critical to user satisfaction. *Human Factors and Ergonomics in Manufacturing*, *14*, 15–29.

Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology: From usability to enjoyment*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hassenzahl, M. (2006). Hedonic, emotional, and experiential perspectives on product quality. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 266–272). London: Idea Group.

Hassenzahl, M., & Sandweg, N. (2004). From mental effort to perceived usability: Transforming experiences into summary assessments. In *CHI '04 proceedings on human factors in computing systems* (pp. 1283–1286). New York: ACM Press.

Hassenzahl, M., & Tractinsky, N. (2006). User experience—a research agenda. *Behaviour & Information Technology*, *25*, 91–97.

International Organization for Standardization. (1998). *ISO 9241—Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability.* (available from www.iso.org/iso/en/CatalogueListPage.CatalogueList). Geneva, Switzerland; ISO.

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Kirakowski, J., & Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, *24*, 210–212.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*, 261–273.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, *60*, 269–298.

Leder, H., & Carbon, C. C. (2005). Dimensions in appreciation of car interior design. *Applied Cognitive Psychology*, *19*, 603–618.

Lindgaard, G., & Dudek, C. (2003). What is the evasive beast we call user satisfaction? *Interacting with Computers*, *15*, 429–452.

Mahlke, S. (2006). Aesthetic and symbolic qualities as antecedents of overall judgments of interactive products. In N. Bryan-Kinns, A. Blanford, P. Cruzon, & L. Nigay (Eds.), *People and computers XX—Engage* (pp. 57–64). London: Springer.

Negroponte, N. (1995). *Being digital*. New York: Alfred A. Knopf.

Nielsen, J. (1993). *Usability engineering*. San Diego, CA: Academic Press.

Partala, T., & Surakka, V. (2004). The effects of affective interventions in human–computer interaction. *Interacting with Computers*, *16*, 295–309.

Picard, R. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Rafaeli, A., & Vilnai-Yavetz, I. (2004). Instrumentality, aesthetics and symbolism of physical artifacts as triggers of emotion. *Theoretical Issues in Ergonomics Science*, *5*, 91–112.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1281–1288.

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* (pp. 293–317). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

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

Ward, R. D., & Marsden, P. H. (2003). Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies*, *59*, 199–212.

Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.