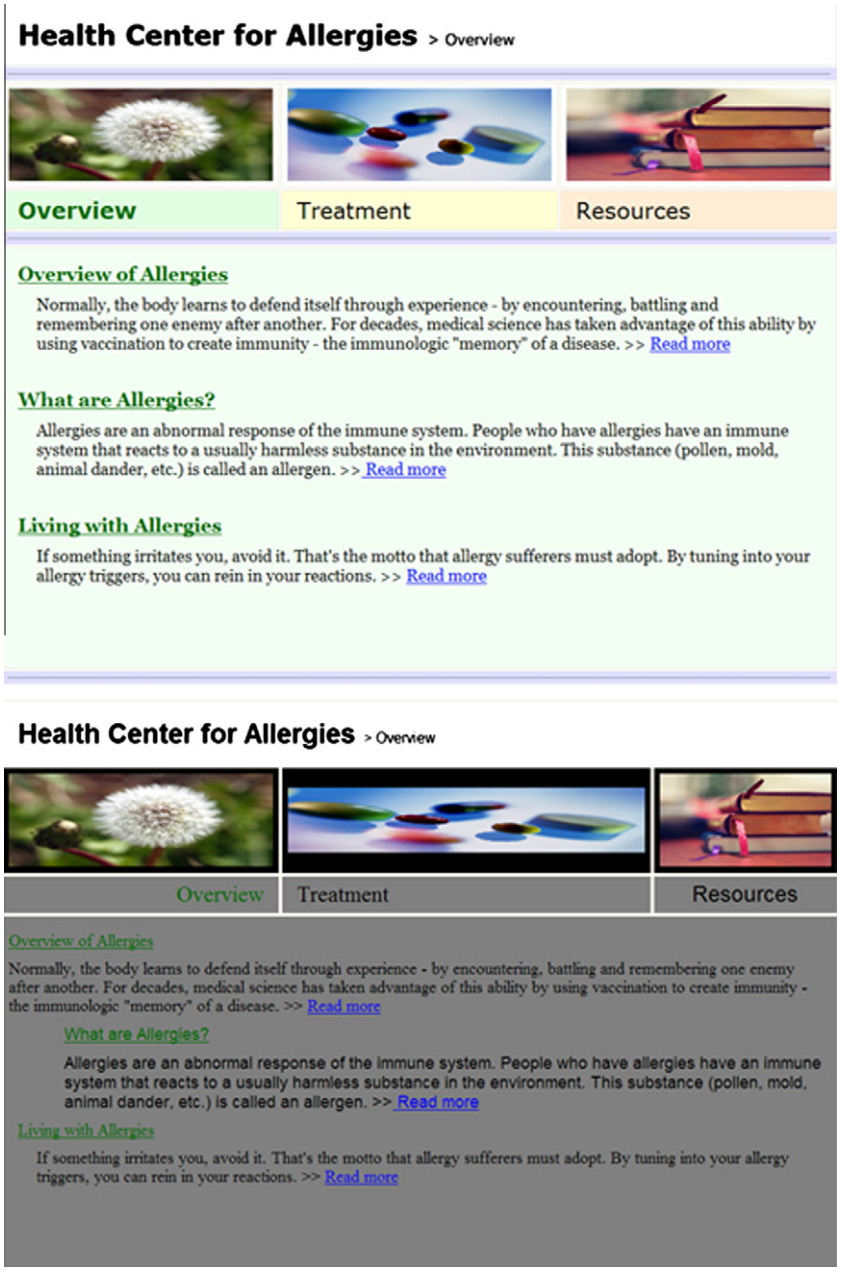

Fig. 2. Sample screenshots for System 4 (high aesthetics; the upper figure) and System 3 (low aesthetics; the lower figure).

tasks on a system; and (3) to assess perceived usability, perceived aesthetics, and user preference after actual use.

The first task was to rate an assigned system with regard to usability, aesthetics, and user preference before actual use. After scanning a system for 1 min without any specific limitations, participants indicated how strongly they agree or disagree with 20 statements in a pre-use evaluation form on a 7-point Likert scale (1 – worst (strongly disagree); 7 – best (strongly agree)). There were 8 statements for perceived usability, 11 statements for perceived aesthetics, and 1 statement for user preference (see Table 2).

The statements for perceived usability were a variation on Items 1 through 8 of the Post-Study System Usability Questionnaire (PSSUQ) (Lewis, 1991, 1992, 1995). Lewis (2002) found that the PSSUQ had a three-factor structure (i.e., system usefulness, information quality, and interface quality), through factor analyses of the PSSUQ data from 5 years of usability studies. Of the 18 items in the PSSUQ, we used only the items for system usability because information quality and interface quality were not related to evaluation of perceived usability. Items 1 through 8 of the PSSUQ were modified so that they could reflect users' expectation for a system in terms of usability.

Moreover, the statements for perceived aesthetics in the pre-use evaluation form were developed based on Lavie and Tractinsky's (2004) findings that users' perceptions of aesthetics consisted of two main dimensions termed "classical aesthetics" and "expressive aesthetics". The classical aesthetics dimension includes the items emphasizing orderly and clean design, and the expressive aesthetic dimension contains the items focusing on creative and original design. Of the statements for perceived aesthetics in the pre-use evaluation form, Items 2 through 6 were related to classical aesthetics and Items 7 though 11 were associated with expressive aesthetics. Similar to the first statement for perceived usability, Item 1 for perceived aesthetics

Table 7
User preference after actual use by conditions.

Condition	N	Mean	Std. Dev.
User preference after actual use			
High aesthetics	36	4.611	1.517
Low aesthetics	37	3.243	1.706
High usability	35	4.714	1.690
Low usability	38	3.184	1.468
Total	73	3.918	1.746

icant interaction effect between the aesthetics and the usability factors on user preference after actual use ($F(1, 69) = 0.570$; p-value = 0.453).

5.2.4. Hypothesis 4: Correlation between perceived aesthetics and perceived usability after actual use

A Pearson correlation was computed to examine the relationship between users' perceptions of aesthetics and usability after actual use. Table 8 presents the descriptive statistics for participants' post-use responses for perceived aesthetics and perceived usability. The scatter plot of perceived aesthetics vs. perceived usability after actual use is displayed in Fig. 4. This figure shows a tendency that the higher a rating of perceived aesthetics was, the higher a rating of perceived usability was. For the 73 participants' ratings, the Pearson correlation coefficient between these two perceptions was 0.730 (p-value < 0.001). After actual use, perceived aesthetics and perceived usability were strongly correlated. However, this strong correlation was not found for all of the four systems. The correlation coefficients between perceived aesthetics and perceived usability after actual use were very high for System 2 (0.727; p-value < 0.001), System 3 (0.665; p-value = 0.003), and System 4 (0.804; p-value < 0.001). However, the correlation coefficient for System 1 was relatively low (0.258; p-value = 0.287). This means that, in the condition of low aesthetics and low usability, perceived aesthetics and perceived usability may not be in a high correlation or in a specific relationship.

5.2.5. Hypotheses 5-1, 5-2, and 5-3: Comparison of interactions before and after actual use

Hypotheses 5-1, 5-2, and 5-3 were proposed to examine the effects of actual use on perceived aesthetics, perceived usability, and user preference. To test these hypotheses, we combined users' responses before and after actual use, and considered the occurrence of actual use as an additional factor. The basic descriptive statistics of perceived aesthetics, perceived usability, and user preference are presented in Table 9.

Table 8
Perceived aesthetics and perceived usability after actual use.

	N	Mean	Std. Dev.
Perceived aesthetics after actual use	73	3.574	1.292
Perceived usability after actual use	73	4.449	1.490

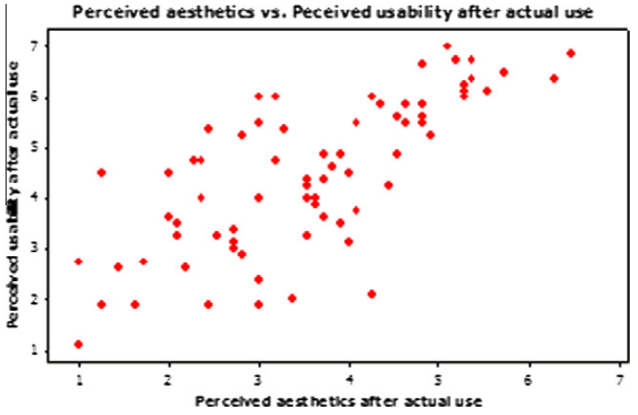

Fig. 4. Scatter plot of perceived aesthetics vs. perceived usability after actual use.

5.2.5.1. Hypothesis 5-1. Effect of actual use on perceived aesthetics. In order to address our expectation that perceived aesthetics would be significantly affected by occurrence of actual use, a three-way ANOVA was performed with the three independent variables (aesthetics level, usability level, and actual use) and the one dependent variable (perceived aesthetics). The normality assumption and the homogeneity of variances assumption for the ANOVA were satisfied. The ANOVA results revealed that the difference between participants' ratings for perceived aesthetics before and after actual use was statistically significant ($F(1, 138) = 4.56$; p-value = 0.035). Also, perceived aesthetics was significantly influenced by the aesthetics factor ($F(1, 138) = 86.67$; p-value < 0.001) and the usability factor ($F(1, 138) = 22.68$; p-value < 0.001). However, there was no interaction effect related with actual use on perceived aesthetics: ($F(1, 138) = 0.38$; p-value = 0.536) for the interaction effect between aesthetics and actual use; ($F(1, 138) = 0.94$; p-value = 0.335) for the interaction effect between usability and actual use; or ($F(1, 138) = 0.00$; p-value = 0.990) for the interaction effect among aesthetics, usability, and actual use.

5.2.5.2. Hypothesis 5-2. Effect of actual use on perceived usability. To examine the effect of actual use on perceived usability, a three-way ANOVA with the three independent variables (aesthetics level, usability level, and actual use) and the one dependent variable (perceived usability) was performed. The experimental data satisfied the assumptions of the ANOVA. According to the ANOVA results, the occurrence of actual use had a significant effect on perceived usability ($F(1, 138) = 13.29$; p-value < 0.001). Also, perceived usability was significantly affected by the aesthetics factor ($F(1, 138) = 27.57$; p-value < 0.001) and the usability factor ($F(1, 138) = 22.81$; p-value < 0.001). No interaction effect between aesthetics and actual use ($F(1, 138) = 0.02$; p-value = 0.880) was found, nor was interaction effect among aesthetics, usability, and actual use ($F(1, 138) = 0.35$; p-value = 0.556), on perceived usability. However, the interaction effect between usability and actual use on perceived usability was significant ($F(1, 138) = 7.64$; p-value = 0.006; significant at the 0.05 level). Fig. 5 shows the interaction plot of usability level and actual use for perceived usability. Tukey's post hoc tests were conducted to compare the means of perceived usability before and after actual use in a given usability level. The results indicated that the difference between the mean ratings before and after actual use in the low usability condition was

Table 9
Perceived aesthetics, perceived usability, and user preference before and after actual use.

Condition	N	Mean	Std. Dev.
Perceived aesthetics			
Before actual use	73	3.927	1.286
After actual use	73	3.574	1.292
Total	146	3.750	1.297
Perceived usability			
Before actual use	73	5.200	1.323
After actual use	73	4.449	1.490
Total	146	4.824	1.454
User preference			
Before actual use	73	4.411	1.614
After actual use	73	3.918	1.746
Total	146	4.164	1.694

Interacting with Computers 22 (2010) 530–543

Contents lists available at ScienceDirect

Interacting with Computers

journal homepage: www.elsevier.com/locate/intcom

Understanding user preferences based on usability and aesthetics before and after actual use

Sangwon Lee [a,*], Richard J. Koubek [b,1]

[a] The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, 343 Leonhard Building, University Park, PA 16802, USA
[b] College of Engineering, The Louisiana State University, 3304 Patrick F. Taylor Hall, Baton Rouge, LA 70803, USA

ARTICLE INFO

Article history:
Received 17 August 2009
Received in revised form 24 May 2010
Accepted 25 May 2010
Available online 1 June 2010

Keywords:
User preference
Usability
Aesthetics
Perceived usability
Perceived aesthetics
Actual use

ABSTRACT

Designing a highly preferred product or system is a crucial issue for better information-services and product sales. We attempted to understand the process of users' preference-making based on usability and aesthetics. In the present study, we examined the relationships among usability/aesthetics features, perceived usability/aesthetics, and user preference through an experiment using four simulated systems with different levels of usability and aesthetics. The results showed that, before actual use, user preference was significantly affected by the differences in aesthetics but marginally affected by the differences in usability. On the other hand, after actual use, user preference was significantly influenced by the differences in both usability and aesthetics. Regardless of the occurrence of actual use, user preference was highly correlated with both perceived usability and perceived aesthetics, which were strongly interrelated. Finally, actual use had a significant effect on perceived usability, perceived aesthetics, and user preference. The findings emphasize the importance of considering both perceived usability and perceived aesthetics. They also demonstrate the need for discriminating users' interactions before and after actual use, in developing a more preferable computer-based application.

1. Introduction

In a competitive market of information-services and product sales, most companies aim to maximize profits or increase usage by promoting their computer-based services and products. Non-profit organizations, which have made efforts to develop an information system that more people can use with pleasure and ease, also operate in this sphere. Developing a highly preferred product or system, thus, has become a crucial issue for companies and organizations. A way of assessing a potential user's intention to purchase or use a product or service is user preference, which directly reflects a user's feeling and attitude for applications (van der Heijden, 2003). Users make their preferences for computerized products or systems by consciously or subconsciously assigning different weights to various factors. Of these factors, two of the most influential are usability and aesthetics (Hartmann et al., 2008; Schrepp et al., 2006; Tractinsky et al., 2000).

What constitutes a good product or system has been traditionally explained by relating it to usability. Several studies (e.g., Head, 1999; Nielsen, 1993; Pearrow, 2000; Shneiderman, 1998; Spool, 1999; Wang, 2001) suggested guidelines and requirements to help design a better product or system. Their focuses mainly lie in usability, although a few guidelines include aesthetic aspects. To evaluate how much applications actually satisfy users, usability measurements have been traditionally emphasized. Over time, methods and techniques to measure usability – usability evaluation methods (UEMs) – have been developed (Hartson et al., 2003; Rosson and Carroll, 2002). The goals of UEMs is to make an application more usable and preferable by finding what needs to be changed or developed in the application as much as is possible in terms of time and cost. However, the design guidelines and the UEMs tended to strongly focus on usability, thereby neglecting other relevant aspects. The UEMs, specifically, placed too much emphasis on heuristic approaches to evaluating design quality and measuring user performance to explain the causes of usability problems in a system (Sutcliffe et al., 2000) though the treatment of aesthetics has been tried by heuristic evaluation methods (e.g., "Aesthetics and minimalist design" in Nielsen's 10 usability heuristics (Nielsen, 1994)).

Although performance issues have been emphasized more than aesthetics issues in the literature (Norman, 2002), greater efforts have been made to improve users' aesthetic experience than user performance on a computerized application have been made (Nielsen, 1996; Norman, 2004). Since the first introduction of aesthetic considerations as an explicit marketing instrument in the design of commodities and mass production artefacts, their importance has quickly increased (Tractinsky et al., 2000). Jordan (1998) argued

* Corresponding author. Tel.: +1 814 865 8011; fax: +1 814 863 4745.
E-mail addresses: sangwon.advance@gmail.com, sangwon@psu.edu (S. Lee), rkoubek@lsu.edu (R.J. Koubek).
[1] Tel.: +1 225 578 5701; fax: +1 225 578 9162.

0953-5438/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2010.05.002

that usable products do not necessarily mean pleasurable ones, and usability and aesthetics should be considered in developing pleasurable products which affect future purchase decisions. Indeed, the usability and aesthetics concepts are two principal dimensions in design (Hartmann et al., 2008; Schrepp et al., 2006; Tractinsky et al., 2000).

Usability can be defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO, 1998). This definition includes both objective measures and subjective feelings of how usable an application is. The other dimension, aesthetics, can be described as "a predominantly affect-driven evaluative response to the visual Gestalt of an object" (Hassenzahl, 2008, p. 291). This description of aesthetics, unlike the ISO definition of usability, only focuses on users' subjective judgments of how aesthetic a product or a system is. To consistently refer only to subjective judgments, perceived usability and perceived aesthetics are defined as a user's subjective perceptions and judgments of usability attributes (e.g., usefulness, ease-of-use, and functionality) and aesthetics features (e.g., colour and layout), respectively. These concepts were well established in previous studies (e.g., Hassenzahl, 2004; Thüring and Mahlke, 2007). Objective measures for usability are separately considered from perceived usability. In that user preferences are made based on their subjective judgments of several criteria, perceived usability and perceived aesthetics need to be considered as important in understanding users' preference-making.

There have been a few studies that have simultaneously considered perceived usability, perceived aesthetics, and user preference (or user satisfaction or overall appraisal) in one experiment. Dillon (2001) introduced perceived usability and perceived aesthetics as two of the measures of user experience. He stated that users' perceptions were informed by aesthetics. The experimental data in his study indirectly showed a potential mismatch among perceived usability, perceived aesthetics, and user preference in users' interactions with computer-based interface designs. In another study, Lindgaard and Dudek (2003) demonstrated that, while a high level of perceived aesthetics positively affected user satisfaction, it did not lead to a high level of perceived usability, in their experiment using different types of web sites. De Angeli et al. (2006) had users evaluate two web sites presenting the same information through different user interaction styles. The experimental results revealed that expressive aesthetics and perceived usability were rated higher in users' preferred interaction style. Moreover, Thüring and Mahlke (2007) proposed a model integrating users' emotions and perceptions of instrumental and non-instrumental qualities. These two qualities can correspond to perceived usability and perceived aesthetics, respectively. These researchers found that the two types of qualities influenced users' emotions in different ways. Consequently emotional reactions and instrumental and non-instrumental qualities affected his or her overall appraisal of the system.

These experiment-based studies have methodological limitations in explaining the process of users' preference-making based on usability and aesthetics, even though they have yielded some revealing insights. Dillon (2001) did not directly deal with user preferences based on perceived usability and perceived aesthetics, but rather published some related experimental data. He did not go further to clearly describe the interrelationship among the factors in users' perceptions and judgments. Furthermore, the experimental conditions in Lindgaard and Dudek's (2003) study were limited in terms of usability and aesthetics. Rather than arranging the variables in all possible combinations, they only considered the usability factor (low/high usability), or specific conditions (high usability and high aesthetics/low usability and high aesthetics) in their experiments. It is difficult to apply the findings of Lindgaard and Dudek's (2003) study directly to other systems with different usability or aesthetics conditions. Also, in de Angeli et al.'s (2006) experiment, the meaning of user preference was limited to a user's favourite interaction style. Finally, the model proposed by Thüring and Mahlke (2007) was based only on specific relationships that could be proven through their experiments using simulated mobile phones and portable audio players and was made general by the use of comprehensive terminology (e.g., instrumental and non-instrumental qualities) instead of more specific terms (e.g., perceived usability and perceived aesthetics).

In answer to the methodological limitations of the previous studies, we introduce a new methodology to examine the relationships among usability/aesthetics features, perceived usability/aesthetics, and user preference in a more concrete experimental setting (see Section 4). To better comprehend user preferences, we consider the full combination of usability and aesthetics levels and examine users' interactions with the consideration of the occurrence of actual use. The discrimination of users' responses before and after actual use can be meaningful because the occurrence of actual use can influence to which factors users are more exposed and by which they are affected.

2. Literature review

We reviewed previous studies on the relationships between any two of the factors – perceived usability and user preference, perceived aesthetics and user preference, and perceived usability and perceived aesthetics. Each relationship is divided into "before actual use" and "after actual use" by when users make their judgments for the factors.

2.1. Perceived usability and user preference

Perceived usability has been considered as one of the major determinants of user preference, but its explanatory power on user preference can be limited. This is because users often consider factors other than usability when making their preferences.

2.1.1. Before actual use

A few studies have investigated the relationship between perceived usability and user preference in users' evaluations before actual use. In one such study, Keinonen (1997) examined users' perceptions of usability-related attributes (perceived affect, perceived usefulness, perceived ease-of-use, perceived functionality, perceived operational logic, and perceived quality of presentation) in relation to their preference-making before actual use, in an experiment using six heart rate monitors. In his experiment, participants completed questionnaire-based subjective ratings for perceive usability dimensions and preference. Keinonen (1997) found that users' evaluations for the dimensions of perceived usability after the search phase but without actual use were highly interrelated, but the users' perceived usability evaluations could not sufficiently reflect user preferences. He stated that the low explanatory power of perceived usability for user preference may be caused by lacking motivation to search for product information, reliance on the interface design quality, and unclear differences among user mental models. Dillon (2001) showed the experimental data from which we can infer the relationship between perceived usability and user preference before actual use, in his spreadsheet task experiment using seven interface designs (drag/drop on approach, drag/drop between approach, icons approach, radio buttons approach, menus approach, one entry area approach, and four entry areas approach) that were originally presented in a study by Tullis and Kodimer (1992). According to Dillon (2001), before actual use, there was no strong correlation between perceived usability (termed likely usability in his study) and user preference.

However, reasons for this finding were not presented because the main focus of the study was not on the relationship between perceived usability and user preference.

2.1.2. After actual use

For the relationship between perceived usability and user preference after actual use, Hassenzahl (2004) presented that, after actual use, goodness (user satisfaction; which can correspond to user preference) was more strongly affected by pragmatic quality (perceived usability) than hedonic attributes (perceived aesthetics). In his second experiment, participants assessed their mental effort with the Subjective Mental Effort Questionnaire (SMEQ) and revised their former ratings before actual use for each MP3-player skin after completion of two scenario tasks. Hassenzahl (2004) emphasized that the not-stable properties of goodness and pragmatic quality were affected by user experience. Later, this was supported by Schrepp et al.'s study using (Schrepp et al., 2006) using a business software and Van Shaik and Ling's study (van Schaik and Ling, 2008) using an intranet site.

2.2. Perceived aesthetics and user preference

As previously stated, we agree with Tractinsky et al. (2000) that aesthetics needs to be considered as one of the two major dimensions (the other being usability) in users' interaction with an application. In the same vein, Bloch (1995) and Lavie and Tractinsky (2004) asserted that both consumers' purchasing decisions and the desirability and commercial value of an interactive engineering product or system were strongly affected by aesthetic quality. In addition, Pandir and Knight (2006) emphasized the effects of individual differences in personal interests and tastes on preferences through an experiment using the screenshots of 12 homepages. According to the researchers, preference judgments were influenced the subjective factors, and thus it was difficult to obtain an agreement on user preference based on aesthetics among individuals.

2.2.1. Before actual use

There have been several studies that have demonstrated the effect of perceived aesthetics on user preference in users' interaction with computer-based applications before actual use. Yamamoto and Lambert (1994), for example, looked at the effect of seven products' aesthetics on user preference without actual use. Participants viewed the product photographs (small gearmotors, small DC motors, precision dispense pumps, solenoid valves, stepper motor controllers, oscilloscopes, and multimeters) with eight product profiles. Data was gathered via personal interviews and from users' subjective rakings for three to five sets of eight profiles in terms of preference. The results showed a weak but non-ignorable effect of perceived aesthetics on user preference, though it varied depending on what the product was. According to these researchers, the effect of perceived aesthetics on user preference did not exceed those of performance or price concerns. Nonetheless, the importance of product aesthetics can be very high in a marketplace facing increasing standardization. In another study, Schenkman and Jönsson (2000) stated that they had seen a significant influence of perceived aesthetics on users' overall impressions and preferences. Their experiment involved 13 existing web pages (eight web pages from telecommunication or electronics companies and five web pages from other fields (National Geographic, Disney, Green Peace, L'Oreal Cosmetics, and Krook Consulting)). After participants did pairwise comparisons of the web pages to indicate similarity and preference, they assigned subjective ratings in seven categories (complexity, legibility, order, beauty, meaningfulness, comprehension, and overall impression) for each web page. A principal component analysis of the experimental results indicated that the variables related to beauty, meaningfulness, and overall impression were grouped together, and the property fitting (PROFIT) analysis showed that the beauty scale was notably related to the preference judgments. The experimental data supported the idea that beauty was the best predictor for overall judgments and user preference.

In a study by Hassenzahl (2004), participants rated four MP3-player skins on four variables (pragmatic quality, hedonic quality, beauty, and goodness) without using the products. He stated that the relationship between perceived aesthetics and user preference was strong and its degree was greater in products for a private use than in those for a public use. At the same time, however, he emphasized the difference between goodness and beauty in nature. In users' interactions after actual use, goodness was significantly associated with pragmatic attributes, while beauty was largely related with hedonic attributes. This can be connected to the relatively more stable properties of beauty than that of goodness over time (van Schaik and Ling, 2008). In another study, Hall and Hanna (2004) examined the effects of the text-background colour combinations in two web pages, one with educational content and one with commercial content. Each web page had four colour conditions. The researchers looked at perceived aesthetics and intention to purchase before actual use, and found a significant relationship between these and users' colour preference. Those two studies indicate that, before actual use, when a user perceives a product to be aesthetically-pleasing and stimulating, he or she can have a positive decision to use a computer-based product or service.

2.2.2. After actual use

The relationship between perceived aesthetics and user preference after actual use of an application has been examined by a few researchers. In one study, Tractinsky et al. (2000) illuminated the relationship between perceived aesthetics and user preference. They designed an experiment using nine automated teller machines (ATMs) with the same components in different arrangements. Each participant was asked to complete experimental tasks on an ATM with specific levels of usability and aesthetics (low and high usability/low, medium, and high aesthetics) and then to rate the ATM on several dimensions. These researchers found a strong relationship between perceived aesthetics and user satisfaction after actual use. Another key finding was that aesthetics design seemed to be associated mainly with systems used voluntarily, but its importance needed to be emphasized for involuntary system use. Van der Heijden (2003) designed an experiment applying the Technology Acceptance Model (TAM) to a Dutch generic portal site. The study showed that perceived visual attractiveness affected users' perceptions of usefulness, ease-of-use, and enjoyment, and finally influenced users' attitudes towards use. Based on the experimental results, he argued that the role of the attractiveness/enjoyment couple in users' behaviours and judgments needed be considered as the intrinsic motivation counterpart of the ease-of-use/usefulness. In addition, De Angeli et al. (2006) found that perceived aesthetics had a significant effect on user preference. Participants compared two web sites with the same information but with different user interface styles (menu-based style and metaphor-based style). After completing the experimental tasks, participants expressed their overall preference for the two web sites. The results showed that users' evaluations of the design's creativity and originality (termed expressive aesthetics in Lavie and Tractinsky's (2004) study) were useful in predicting overall preference. Specifically, participants had a tendency to discount negative attributes of systems with a preferred interaction style.

2.3. Perceived usability and perceived aesthetics

In this section, we look at the literature on the relationship between perceived usability and perceived aesthetics, most of which has not focused on user preference.

2.3.1. Before actual use

Several studies have examined the relationship between perceived usability and perceived aesthetics before actual use. Kurosu and Kashimura (1995) investigated the effect of these factors on each other (in their study, termed apparent usability and apparent beauty), based on participants' subjective ratings on functional and aesthetics aspects of 26 ATM layouts. The results of the experiment, which was conducted in Japan, indicated that apparent usability was positively correlated with and was strongly affected by apparent aesthetics. Tractinsky (1997) got similar results from a similar experiment in Israel. In his experiment, participants evaluated the ATM layouts on how usable each appeared to be and how beautiful it was, after scanning each layout for about 20 s.

In a later study, Tractinsky et al. (2000) examined the relationship among pre-experimental measures (aesthetics, ease-of-use, and amount of information) for nine ATM layouts. They found that perceived aesthetics was strongly correlated with perceived usability before actual use. The findings of these studies suggested that users tended to strongly consider aesthetic attributes even when they evaluated usability and functional aspects of systems, and the usability–aesthetics association in users' perceptions was a genuine phenomena without biases caused by experimental methodologies. In addition, Ben-Bassat et al. (2006) looked at the relationship between perceived usability and perceived aesthetics. They asked participants to use four simulated computer-based phone books for data entry tasks and then to fill out a concluding questionnaire about usability and aesthetics for each experimental system after twice reading instructions describing the use of that system. The experimental results confirmed that perceived usability and perceived aesthetics were interrelated; that is, perceived aesthetics affected perceived usability, or vice versa. In that this finding was consistently found regardless of both the degrees of experience with the system and monetary incentives based on performance, the interdependence between perceived usability and perceived aesthetics seemed to be robust.

2.3.2. After actual use

The relationship between perceived usability and perceived aesthetics after actual use has also been studied by many researchers. In the above-mentioned study, Tractinsky et al. (2000) demonstrated that the strong correlation between perceived usability and perceived aesthetics before actual use remained the same after actual use. This suggests that the occurrence of actual use is not an influential factor in determining the interrelationship between perceived usability and perceived aesthetics. Participants were assigned to one of the experimental conditions (low and high usability/low, medium, and high aesthetics). After performing four types of tasks, users were asked to rate the assigned system on post-experimental measures. The results showed that the degree of an application's aesthetic quality had a more significant effect on perceived usability and perceived aesthetics than the level of its actual usability. This shows the relative importance of users' subjective perception over objective performance measures in design. Likewise, Van der Heijden (2003) found that perceived aesthetics (perceived visual attractiveness) had a significant effect on perceived usability (perceived usefulness and perceived ease-of-use) after actual use, in his Dutch generic portal site experiment. He emphasized perceived aesthetics and perceived usability as two principal criteria of users' judgments for computer-based applications. In another case, when De Angeli et al. (2006) evaluated two web sites with different interface styles, they found that perceived usability after actual use was more correlated with expressive aesthetics in the menu-based style than in the metaphor-based style. This implies that the relationship between perceived usability and perceived aesthetics may vary depending on systems' features, such as interaction style. In addition, Ben-Bassat et al. (2006) found that a strong interrelationship between perceived usability and perceived aesthetics was not affected by occurrence of actual use in their experiment with four computerized systems for data entry tasks. This finding is consistent with that of Tractinsky et al.'s (2000) experiment. Thüring and Mahlke (2007) reiterated the strong effect of perceived aesthetics on perceived usability after actual use in their study, in which users operated four portable digital audio players before rating each system on perceived usability and expressing their overall judgments. According to these researchers, perceived usability and perceived aesthetics need to be considered important in that both of them affect users' emotional experience and overall satisfaction with the system.

Despite all the convincing evidence connecting perceived usability and perceived aesthetics after actual use, a few studies have questioned this strong relationship. For example, Lindgaard and Dudek (2003) claimed that these two factors may not be significantly correlated, arguing that the findings of Tractinsky et al. (2000) may not be true. Lindgaard and Dudek conducted two different experiments. The first experiment used a portal site with high usability and a "travel" site with low usability, and the second used a "florist" site with high usability and high aesthetics (appeal) and a "pen" site with low usability and high aesthetics. After using each site for 10 min, participants were given an unstructured interview and asked to complete the Web site Analysis MeasureMent Inventory (WAMMI). One of the findings from these two experiments indicated that the correlation between perceived usability and perceived aesthetics after actual use was not significant in a web site with a low level of usability and a high level of aesthetics. Similarly, Hassenzahl (2004) stated that to be beautiful did not mean to be usable, supporting his statement by pointing out reasons to believe the experimental design or the analysis method of Tractinsky et al.'s (2000) study had potential to skew the results. According to Hassenzahl's second study, perceived usability (pragmatic quality) and perceived aesthetics (beauty) did not show a significant correlation. He asserted that, after actual use, perceived usability and perceived aesthetics were more associated with the product's pragmatic quality and the product's ability to explain its identification and to make it familiar, respectively.

One possible reason for the conflicting findings on the relationship between perceived usability and perceived aesthetics could be different experimental domains. For example, a "florist" site and a "pen" site in Lindgaard and Dudek's (2003) experiment can be considered as commerce web sites where a user can purchase something for others, while ATMs (Tractinsky et al., 2000), educational web sites (de Angeli et al., 2006), and computer-based phone books (Ben-Bassat et al., 2006) cannot. A user's emotions may be strongly engaged in the purchase of a product for someone. Such emotions may influence his or her perceived aesthetics positively or negatively. Another reason for the conflicting findings may be task characteristics (or interaction type) in the experiments. Hassenzahl et al. (2002) found that the degree of a relationship between ratings of usability and aesthetics after actual use depended on interaction types with a system – goal-directed or leisure-related. Their finding suggests that the relationship between the two was strong in the case of a goal-directed interaction type, whereas it was not strong in the case of a leisure-related interaction type. We can also consider the degree of participants' expertise and familiarity with systems as a possible explanation for the conflicting findings. If users perform tasks well on a system due to their familiarity with it, they may perceive it as usable regardless of its actual usability level. Also, participants with some knowledge of visual interface design

principles may recognize the good or bad qualities of a system, and this can influence how they perceive the aesthetics of the system. However, only Lindgaard and Dudek (2003) stated that participants in their experiments had little or no user interface design and evaluation experience. Finally, the relationship between perceived usability and perceived aesthetics may depend on usability and aesthetics conditions of the systems used in the studies. Excepting Lindgaard and Dudek (2003), the other studies did not substantially deal with usability and aesthetics levels, nor simultaneously consider these two attributes in designing their experiments. As a contribution to resolving this issue, we designed an experiment with a full combination of the usability and the aesthetics factors.

2.4. Summary of relationships among perceived usability, perceived aesthetics, and user preference

The general relationships among perceived usability, perceived aesthetics, and user preference are summarized based on the literature review above, with the consideration of occurrence of actual use (see Fig. 1).

In users' interactions with an application before actual use, the relationship between perceived aesthetics and user preference was relatively stronger than that between perceived usability and user preference. This means that user preference before actual use can be more strongly influenced by how aesthetic users perceive an application to be than by how usable they feel it to be. Also, it appears that perceived usability and perceived aesthetics were strongly correlated. This implies that, before actual use, there is a high possibility that users who perceived an application as aesthetically-pleasing rated it as more usable than did users who perceived it as aesthetically-displeasing, regardless of the level of its actual usability. To date, the literature shows that the relationship between perceived aesthetics and user preference remained strong after actual use and the relationship between perceived usability and user preference after actual use proved to be strong. It is necessary to deeply examine the relative importance of usability and aesthetics in users' preference-making after actual use. Also, there are conflicting findings on the relationship between perceived usability and perceived aesthetics after actual use. Thus, this relationship remains an open question and needs to be clarified. We attempt to examine these relationships – to confirm the previous studies' findings or to clarify the unclear relationships – through an experiment simultaneously considering usability, aesthetics, and occurrence of actual use in users' preference-making.

3. Hypotheses

Nine hypotheses were derived to understand user preferences based on usability and aesthetics. The hypotheses can be divided into three parts: (1) interaction before actual use (Hypotheses 1-1, 1-2 and 2); (2) interaction after actual use (Hypotheses 3-1, 3-2 and 4); and (3) comparison of interactions before and after actual use (Hypotheses 5-1, 5-2, and 5-3).

Hypothesis 1-1. Before actual use, user preference is significantly affected by differences in aesthetics.

Hypothesis 1-2. Before actual use, user preference is not significantly affected by differences in usability.

Before actual use, users mainly interact with the visual attributes of a system. Users cannot properly recognize the level of a system's actual usability because they do not actually use the system. Thus, we expect that user preference before actual use is significantly affected by differences in aesthetics (Hypothesis 1-1) but is not significantly influenced by differences in usability (Hypothesis 1-2).

Hypothesis 2. Before actual use, perceived aesthetics and perceived usability are highly correlated.

Hypothesis 2 attempts to confirm the strong relationship between perceived aesthetics and perceived usability before actual use. Several of the previous studies (Ben-Bassat et al., 2006; Kurosu and Kashimura, 1995; Tractinsky, 1997; Tractinsky et al., 2000) consistently demonstrated that, before actual use, perceived aesthetics was highly correlated with perceived usability.

Hypothesis 3-1. After actual use, user preference is significantly affected by differences in aesthetics.

Hypothesis 3-2. After actual use, user preference is significantly affected by differences in usability.

Hypotheses 3-1 and 3-2 try to illuminate the effects of both aesthetics and usability features on user preference after actual use. Through actual use, users interact with both the visual and the functional attributes of a product or a system. Unlike users' responses before actual use, in this case, users have been sufficiently exposed to the application's actual usability. Thus, it is reasonable to expect that user preference after actual use is significantly affected by both the aesthetics and the usability factors.

Hypothesis 4. After actual use, perceived aesthetics and perceived usability are highly correlated.

This hypothesis focuses on the relationship between perceived aesthetics and perceived usability after actual use. As mentioned in the literature review, the existence of this relationship remains unclear. In this study, we expect to find that, after actual use, perceived usability is highly correlated with perceived aesthetics.

Hypothesis 5-1. Perceived aesthetics is significantly affected by occurrence of actual use.

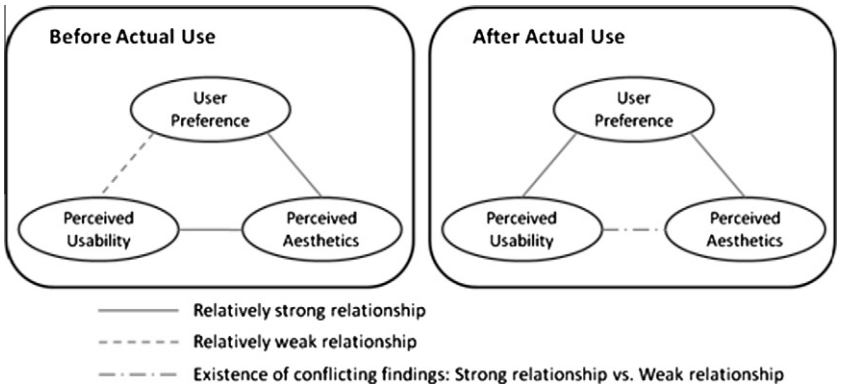

Fig. 1. Relationships among perceived usability, perceived aesthetics, and user preference based on previous studies' findings.

Hypothesis 5-2. Perceived usability is significantly affected by occurrence of actual use.

Hypothesis 5-3. User preference is significantly affected by occurrence of actual use.

Users are continuously exposed to aesthetics features, irrespective of occurrence of actual use, but not exposed to usability attributes. We predict that perceived usability is different by occurrence of actual use (Hypothesis 5-2). Also, considering the strong correlation between perceived aesthetics and perceived usability, it is expected that perceived aesthetics is significantly influenced by occurrence of actual use (Hypothesis 5-1). Consequently, user preferences before and after actual use will be significantly different (Hypothesis 5-3).

4. Method

To test the nine hypotheses, an experiment using four simulated systems with different usability and aesthetics levels was conducted. This section describes participant characteristics, usability and aesthetics manipulations, tasks and variables, and experimental design and procedure.

4.1. Participants

From the user background survey which was mainly based on Part 1 (system experience) and Part 2 (past experience) of the Questionnaire for User Interaction Satisfaction (QUIS) (Chin et al., 1988), we obtained participants' information. Seventy-three undergraduate or graduate students majoring in engineering participated in the experiment: 59 males and 14 females; aged 23.68 years on average; 37 North Americans, 31 Asians, and 5 Europeans. For screening questions for the experiment, participants answered that they had no difficulty in perceiving and distinguishing colours. Regarding of user background, they were familiar with using computers: they spent on average 5.31 h per day at a computer. Also, they had little experience with allergy-related systems (e.g., web sites providing some information about allergies). They were compensated $15 for participation. Of the participants, 19, 19, 18, and 17 participants were assigned to System 1, System 2, System 3, and System 4, respectively, which were different in usability and aesthetics levels.

4.2. Stimuli and manipulations

To facilitate manipulation of usability and aesthetics factors, four simulated systems with different levels of usability (low vs. high) and aesthetics (low vs. high) were created as stimuli in the experiment by two designers (1 Asian and 1 North American) (see Table 1). The system domain was "Health Center for Allergies", providing various information and resources about allergies. The systems were designed with Microsoft Office FrontPage 2003 and JavaScript, and were validated through expert reviews and a pilot test with 10 users before the actual study.

The usability factor was manipulated by providing well-organized content in a high usability condition and not-well-organized content in a low usability condition. Content organization means organizing and placing individual content items in groups properly and understandably (McCracken and Wolfe, 2004). Bernard (2000) stated the need to structure and organize content items according to users' mental model. Also, he argued that content organization was essentially related to navigation issues, which could affect a system's usability. In our experiment, high usability (Systems 3 and 4) was achieved by following users' mental model based on card sorting for allergy-related content items. We selected 24 content items from three existing allergy web sites: (1) WebMD – Allergies Health Center (www.webmd.com/allergies); (2) Health Scout: Health Encyclopedia – Diseases and Conditions (www.healthscout.com/ency/68/98/main.html); and (3) My Allergy Network (www.healthcentral.com/allergy). Ten potential users for our systems did card sorting for the content items. The data were analyzed by the complete linkage algorithm in IBM EZ-Sort. At the other end of the spectrum, low usability (Systems 1 and 2) was manipulated by intentionally disorganizing the content and assigning awkward names to groups in order to conflict users' mental model based on the distance matrix between the content items.

Given the same level of usability, two systems with different levels of aesthetics had the same content items but differed in terms of colour combination, visual layout, and text font. In fact, aesthetics is based on a user's subjective appreciation for visual features. It can be affected by factors such as participants' or designers' cultural background and systems' inherent characteristics. Thus, aesthetics quality cannot be defined as a universally binary status. To minimize such effects, the systems in two extreme points were designed based on findings from previous studies and their interfaces were reviewed by experts with mixed background. In Hall and Hanna's (2004) study, users perceived white–black or black–white colour combinations for web page text-background as less aesthetic than non-greyscale colour combinations. Also, Cheng et al. (2009) showed that warm colours could create the desired atmosphere and enhance the positive emotion, in an experiment using online retail stores. A few studies (Kurosu and Kashimura, 1995; Tractinsky, 1997; Tractinsky et al., 2000) demonstrated the effect of visual layout on perceived aesthetics through experiments using ATMs with different arrangements of content items. Conklin et al. (2006) achieved differences in aesthetics of home robotic control systems by using different colours, layouts, and fonts. The experimental results showed that the manipulation of interface aesthetics with these factors was successful. Moreover, McCracken and Wolfe (2004) recommended using Georgia or Verdana without blending them in the body text over Times New Roman or Arial. The findings of these studies were applied to creating systems with different levels of aesthetics for our experiment. To create an aesthetic system (Systems 2 and 4), we mainly used an analogous colour harmony scheme (orange, yellow, and green with different brightness and saturation), attractive layouts based on the four principles of visual organization (proximity, alignment, consistency, and contrast), and appealing font styles (Georgia and Verdana without intermixing them in the body text). On the contrary, a non-aesthetic system (Systems 1 and 3) was created through use of greyscale colours (white, black, and greys with different brightness), awkward layouts that did not follow the principles of visual organization, and unappealing font styles (Times New Roman and Arial intermixed in the body text). Fig. 2 shows sample screenshots for System 4 (high aesthetics) and System 3 (low aesthetics).

Table 1
Four systems – usability and aesthetics levels.

	Usability/aesthetics levels
System 1	Low usability/low aesthetics
System 2	Low usability/high aesthetics
System 3	High usability/low aesthetics
System 4	High usability/high aesthetics

4.3. Tasks

Participants were required to perform three major experimental tasks: (1) to evaluate perceived usability, perceived aesthetics, and user preference before actual use; (2) to complete four scenario

Table 2
Statements of pre-use evaluation form.

Assessment of "Perceived Usability before Actual Use"
1. Overall, I am satisfied with how easy it will be to use this system
2. It will be simple to use this system
3. I will be able to effectively complete the tasks and scenarios using this system
4. I will be able to complete the tasks and scenarios quickly using this system
5. I will be able to efficiently complete the tasks and scenarios using this system
6. I feel comfortable using this system
7. It will be easy to learn to use this system
8. I believe I can become productive quickly using this system

Assessment of "Perceived Aesthetics before Actual Use"
1. Overall, I am satisfied with the appearance of this system
2. I feel the design of this system is aesthetic
3. I feel the design of this system is pleasant
4. I feel the design of this system is clear
5. I feel the design of this system is clean
6. I feel the design of this system is systematic
7. I feel the design of this system is creative
8. I feel the design of this system is fascinating
9. I feel the design of this system uses special effects
10. I feel the design of this system is original
11. I feel the design of this system is sophisticated

Expression of "User Preference before Actual Use"
1. I prefer to use this system

queried participants' overall satisfaction with the appearance of a system.

In the second task, participants were required to complete four scenario tasks on an assigned system. This was to allow participants to actually use a system. It was necessary to develop task scenarios involving the system's interface and functional features as much as possible. The scenarios are presented in Table 3.

The third task required participants to rate an assigned system in terms of perceived usability, perceived aesthetics, and user preference after actual use. Based on the experience participants had with a system through completing the scenario tasks on it, they responded to 20 statements in a post-use evaluation questionnaire on a 7-point Likert scale. The statements of the post-use evaluation form were the same as those of the pre-use evaluation form, except the statements' tense changed.

4.4. Variables

The experiment involved three independent variables and six dependent variables. The experimental variables are described in Tables 4-1 and 4-2.

4.5. Experimental design and procedure

The experiment followed a $2 \times 2 \times 2$ mixed factorial design with two between-subjects variables and one within-subjects variable. The between-subjects variables were the usability and the aesthetics levels, and the within-subjects variable was the occurrence of actual use. As presented in Table 4-1, each of the variables had two levels.

Table 3
Task scenarios.

Scenario 1	Suppose that you want to learn the types of allergies. Find the information about the types of allergies and Double-click "Eczema" (one of the allergy types)
Scenario 2	Suppose that you want to understand what causes allergies. Find the information about the causes of allergies and Double-click "cytokines" in a paragraph
Scenario 3	Suppose that you want to find an allergy doctor using a tool of the system. Use the given information to find his/her name and contact information
Scenario 4	Suppose that you want to ask a question about "sun allergy" using a tool of the system. Use the given information and question content to send your question about sun allergy to allergy experts

Table 4-1
Variable description – independent variables.

Independent variables	Description
Usability level	Low usability/high usability
Aesthetics level	Low aesthetics/high aesthetics
Occurrence of actual use	Before actual use/after actual use

Table 4-2
Variable description – dependent variables.

Dependent variables	Description
Perceived usability before actual use	Average score of responses for perceived usability in the pre-use evaluation questionnaire (Items 1–8)
Perceived aesthetics before actual use	Average score of responses for perceived aesthetics in the pre-use evaluation questionnaire (Items 9–19)
User preference before actual use	Score of user preference in the pre-use evaluation questionnaire (Item 20)
Perceived usability after actual use	Average score of responses for perceived usability in the post-use evaluation questionnaire (Items 1–8)
Perceived aesthetics after actual use	Average score of responses for perceived aesthetics in the post-use evaluation questionnaire (Items 9–19)
User preference after actual use	Score of user preference in the post-use evaluation questionnaire (Item 20)

At the preliminary session, the experiment was introduced to participants through an informed consent form presented in writing and then orally. Participants then answered a user background survey. From this survey, we could obtain basic information about participants – age, gender, major, English fluency, familiarity with using computers, and prior experience using allergy-related systems. Participants' answers to these items helped us to check their eligibility and provided bases for analyzing the experimental data from various viewpoints.

The experimental session took place in three stages. In the first stage, participants were assigned to one of the four systems in a counter-balanced order. Participants simply navigated the system for 1 min. Based on their short experience through scanning the system, they gave responses to the pre-use evaluation questionnaire. In the second stage, participants completed four scenario tasks on an assigned system. The tasks were given to participants in a counter-balanced order. There was no time limit in carrying out the scenario tasks. Whether or not participants achieved each of the scenarios was checked during this session. In case of an incorrect task result, the experimenter notified participants of the fact and asked them to perform that task again without any time gap. The task completion time as a user performance measure was electronically recorded. As the final stage, after the completion of the four scenario tasks on a given system, participants assessed the system by answering the items on the post-use evaluation questionnaire.

5. Results

This section describes the manipulation check for usability and aesthetics factors and the statistical results for the experimental data with respect to our hypotheses.

5.1. Manipulation check

Before analyzing the experimental data, it was necessary to check the manipulations for aesthetics and usability. The success of the aesthetics manipulation was evaluated from the viewpoint of perceived aesthetics before actual use. This was because perceived aesthetics before actual use was relatively less affected by the actual usability of a system than perceived aesthetics after actual use was. The normality assumption and the homogeneity of variance assumption of a one-way analysis of variance (ANOVA) for the ratings of perceived aesthetics before actual use were satisfied. The ANOVA result revealed a significant effect of the aesthetics factor on perceived aesthetics before actual use ($F(1, 71) = 46.76$; p-value < 0.001). The mean ratings of perceived aesthetics before actual use were 4.742 (standard deviation = 0.975) and 3.133 (standard deviation = 1.034) for the high and the low aesthetics conditions, respectively. This result indicated that the manipulation of aesthetics was successful.

The usability manipulation was checked by comparing the task completion times (the total time required to complete the four scenario tasks) in the high and low usability conditions. A two-way ANOVA for the effects of usability and aesthetics levels on time performance was performed. To satisfy the assumptions of the ANOVA, task completion time was transformed by using log (base 10) transformation. The average task completion times in the high and the low usability conditions were 153.0 s (standard deviation = 63.1) and 299.9 s (standard deviation = 132.7), respectively. The ANOVA result showed that the usability factor had a significant effect on the task completion time ($F(1, 69) = 40.83$; p-value < 0.001). This implies that participants more quickly completed the scenario tasks in the high usability condition than in the low usability condition. At the same time, there was no significant effect of the aesthetics factor on task completion time ($F(1, 69) = 1.10$; p-value = 0.299), nor was there the interaction effect of the usability and the aesthetics factors on task completion time ($F(1, 69) = 0.44$; p-value = 0.507). Thus, it can be concluded that the manipulation of the usability factor was successful and was free from any aesthetics side effects.

5.2. Test of Hypotheses

5.2.1. Hypotheses 1-1 and 1-2: Effects of differences in aesthetics/usability on user preference before actual use

In order to address the first two hypotheses, that differences in aesthetics would significantly influence user preference before actual use but differences in usability would not, a two-way ANOVA was performed with the two independent variables (aesthetics level and usability level) and the one dependent variable (user preference before actual use). The descriptive statistics for participants' ratings of user preference before actual use in each aesthetics/usability condition are presented in Table 5. The data satisfied the assumptions of the ANOVA. The ANOVA results revealed that the aesthetics factor had a significant effect on user preference before actual use ($F(1, 69) = 16.69$; p-value < 0.001), but the usability factor did not ($F(1, 69) = 3.74$; p-value = 0.057). This means that, before actual use, user preference was significantly affected by differences in aesthetics, but only marginally influenced by differences in usability. There was no interaction effect between the aesthetics and the usability factors on user preference before actual use ($F(1, 69) = 0.06$; p-value = 0.800).

Table 5
User preference before actual use by conditions.

Condition	*N*	Mean	Std. Dev.
User preference before actual use			
High aesthetics	36	5.111	1.260
Low aesthetics	37	3.730	1.644
High usability	35	4.743	1.633
Low usability	38	4.105	1.566
Total	73	4.411	1.614

5.2.2. Hypothesis 2: Correlation between perceived aesthetics and perceived usability before actual use

The relationship between perceived aesthetics and perceived usability before actual use was illuminated by conducting a Pearson correlation analysis. The descriptive statistics for participants' ratings are presented in Table 6. Fig. 3 shows that there was a positive correlation between users' perceptions of aesthetics and usability before actual use. For the 73 participants' ratings, the Pearson correlation coefficient between these two perceptions was 0.660 (p-value < 0.001). This implies that, before actual use, perceived aesthetics was strongly correlated with perceived usability. Such a relationship was consistent regardless of the aesthetics and the usability levels. The correlation coefficients between perceived aesthetics and perceived usability before actual use in each system were generally high: 0.768 (p-value < 0.001) for System 1; 0.544 (p-value = 0.016) for System 2; 0.405 (p-value = 0.096) for System 3; and 0.459 (p-value = 0.064) for System 4.

5.2.3. Hypotheses 3-1 and 3-2: Effects of differences in aesthetics/usability on user preference after actual use

We expected that differences in aesthetics would significantly affect user preference after actual use, and so would differences in usability. For these hypotheses, a two-way ANOVA with the two independent variables (aesthetics level and usability level) and the one dependent variable (user preference after actual use) was performed. Table 7 displays the descriptive statistics for participants' ratings of user preference after actual use in each aesthetics/usability condition. The assumptions for the ANOVA were satisfied. According to the ANOVA results, user preference after actual use was significantly affected by both differences in aesthetics ($F(1, 69) = 17.03$; p-value < 0.001) and differences in usability ($F(1, 69) = 21.39$; p-value < 0.001). This means that participants' ratings in a high aesthetics/usability level were significantly higher than those in a low aesthetics/usability level. There was no signif-

Table 6
Perceived aesthetics and perceived usability before actual use.

	N	Mean	Std. Dev.
Perceived aesthetics before actual use	73	3.927	1.286
Perceived usability before actual use	73	5.200	1.323

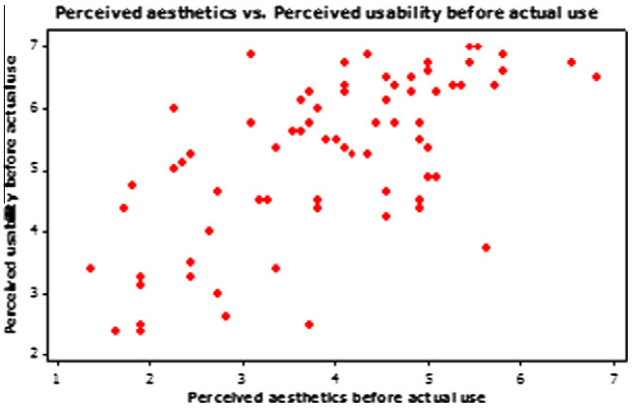

Fig. 3. Scatter plot of perceived aesthetics vs. perceived usability before actual use.

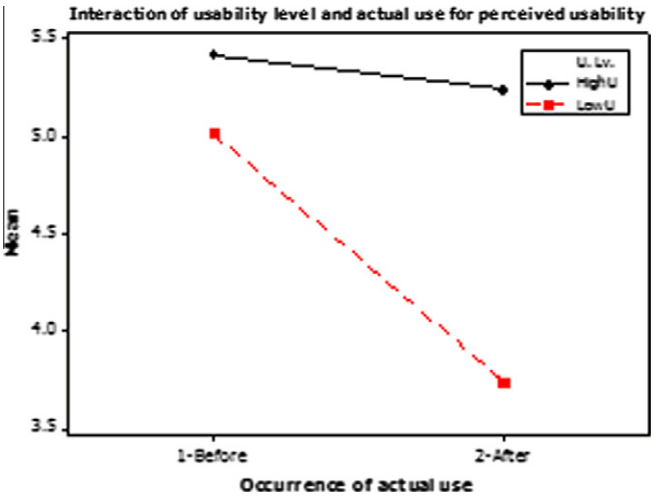

Fig. 5. Interaction plot of usability level and actual use for perceived usability.

significant (t-value = −4.630; adjusted p-value < 0.001), but the difference in the high usability condition was not (t-value = −0.611; adjusted p-value = 0.928).

5.2.5.3. Hypothesis 5-3. Effect of actual use on user preference. A three-way (aesthetics level, usability level, and actual use) ANOVA with user preference as the dependent variable was performed to address the hypothesis that user preference would be significantly affected by occurrence of actual use. The assumptions for the ANOVA were satisfied. The ANOVA results revealed that there was a significant effect of the actual use on user preference ($F(1, 138) = 3.99$; p-value = 0.048). Also, the main effects of the aesthetics factor ($F(1, 138) = 33.71$; p-value < 0.001) and the usability factor ($F(1, 138) = 21.36$; p-value < 0.001) on user preference were significant. There was not (aesthetics × actual use) interaction effect ($F(1, 138) = 0.00$; p-value = 0.987), or (aesthetics × usability × actual use) interaction effect ($F(1, 138) = 0.12$; p-value = 0.729), with regard to ratings for user preference. However, the interaction effect between the usability factor and the occurrence of actual use on user preference was marginally significant ($F(1, 138) = 3.49$; p-value = 0.064; not significant at the 0.05 level). Fig. 6 displays the interaction plot of usability level and actual use for user preference. Tukey's post hoc tests were conducted to compare the mean ratings for user preference before and after actual use in a given usability

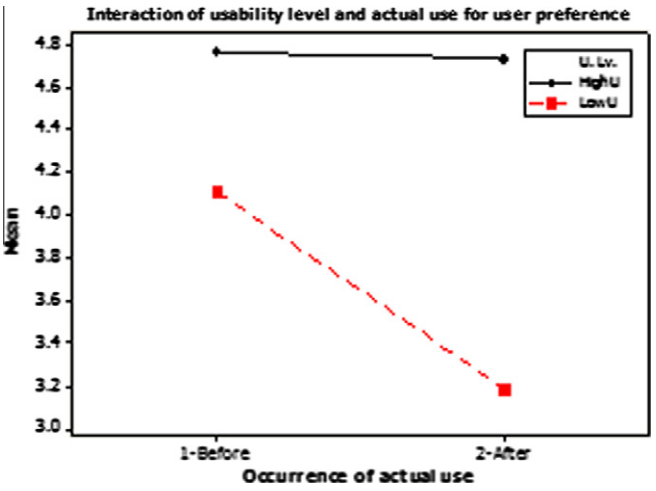

Fig. 6. Interaction plot of usability level and actual use for user preference.

level. According to the test results, the difference between the user preference ratings before and after actual use in the low usability condition was significant (t-value = −2.792; adjusted p-value = 0.030), while the difference in the high usability condition was not (t-value = −0.090; adjusted p-value = 0.999).

6. Discussion

The analysis results above indicated that our hypotheses were, at least partially, supported. Table 10 summarizes the testing outcomes for the hypotheses. In users' preference-making based on usability and aesthetics, usability/aesthetics features affect users' perceptions of usability/aesthetics, and, again, these perceptions influence users' judgments of preference. We discuss the relationships among usability/aesthetics factors, perceived usability/aesthetics, and user preference by separately considering users' responses before and after actual use.

6.1. Interaction before actual use

The result of Hypothesis 1-1 implies that, before actual use, users properly recognized the aesthetics features of a system, and this recognition led to different ratings of user preference. Also, the analysis result for Hypothesis 1-2 showed that user preference before actual use was marginally influenced by the usability conditions. We posit possible reasons for this result. First, users may recognize the level of a system's actual usability during the 1 min scan of the system, at which time they were exposed to the usability features to some extent (but not sufficiently). Second, according to the user background survey, participants had familiarity with using computers, and had been exposed to the visual and the functional attributes of computer-based systems many times. Such user experiences may make it possible for users to identify the usability problems of a system in a relatively short time.

Our expectation was that, before actual use, user preference can be more significantly affected by differences in aesthetics than by differences in usability (Hypotheses 1-1 and 1-2). This implies that user preference can be more strongly influenced by perceived aesthetics than by perceived usability. Indeed, the previous studies' findings partially support this connection. Several researchers (Hall

Table 10
Summary of testing outcomes for hypotheses.

Classification	Hypothesis	Result	Meaning
Before actual use	Hypothesis 1-1	Supported	Before actual use, user preference was significantly affected by differences in aesthetics
	Hypothesis 1-2	Partially supported	Before actual use, user preference was marginally affected by differences in usability
	Hypothesis 2	Supported	Before actual use, perceived aesthetics and perceived usability were strongly interrelated
After actual use	Hypothesis 3-1	Supported	After actual use, user preference was significantly affected by differences in aesthetics
	Hypothesis 3-2	Supported	After actual use, user preference was significantly affected by differences in usability
	Hypothesis 4	Supported	After actual use, perceived aesthetics and perceived usability were strongly interrelated
Before and after actual use	Hypothesis 5-1	Supported	The effect of actual use on perceived aesthetics was significant
	Hypothesis 5-2	Supported	The effect of actual use on perceived usability was significant
	Hypothesis 5-3	Supported	The effect of actual use on user preference was significant

and Hanna, 2004; Hassenzahl, 2004; Schenkman and Jönsson, 2000; Yamamoto and Lambert, 1994) emphasized a strong effect of perceived aesthetics on user preference, while a few other studies (Dillon, 2001; Keinonen, 1997) showed a relatively weak relationship between perceived usability and user preference before actual use. However, before actual use, user preference was highly correlated with both perceived aesthetics and perceived usability. In users' responses before actual use, the Pearson correlation coefficient of perceived aesthetics and user preference was 0.814 (p-value < 0.001), and that of perceived usability and user preference was 0.802 (p-value < 0.001). This can be explained by the strong interrelationship between perceived aesthetics and perceived usability (Hypothesis 2).

Similar findings to that of Hypothesis 2 can be found in the previous studies (Ben-Bassat et al., 2006; Kurosu and Kashimura, 1995; Tractinsky, 1997; Tractinsky et al., 2000). To support the high correlation between perceived aesthetics and perceived usability, we additionally performed two 2-sample t tests for the difference between the mean ratings of perceived usability before actual use by the aesthetics conditions and for the difference between the mean ratings of perceived aesthetics before actual use by the usability conditions. The descriptive statistics and the test results are presented in Table 11. According to the results, before actual use, users who perceived a system as aesthetically-pleasing rated it as more usable than did users who perceived it as aesthetically-displeasing. Interestingly, a similar tendency was found for perceived aesthetics before actual use. In users' responses before actual use, the rating of perceived aesthetics was higher in the high usability condition than in the low usability condition. This result supports the notion that, before actual use, users who perceived a system as usable rated it as more aesthetically-pleasing than did users who perceived it as non-usable. Thus, it can be concluded that perceived aesthetics and perceived usability were interrelated and affected by each other.

6.2. Interaction after actual use

The results of Hypotheses 3-1 and 3-2 mean that, in users' interaction after actual use, users properly identified both the aesthetics and the usability features of a system, and reflected this recognition in making their preference for that system. Especially, the significant effect of the usability factor on user preference after actual use means that users had sufficient opportunity to recognize the usability deficiencies of a system by actually using it.

The experimental results showed that, after actual use, user preference was highly correlated with both perceived aesthetics and perceived usability. For users' responses after actual use, the Pearson correlation coefficients of perceived aesthetics and user preference after actual use and of perceived usability and user preference after actual use were 0.826 (p-value < 0.001) and 0.853 (p-value < 0.001), respectively. Similar results can be found in previous studies ((de Angeli et al., 2006; Tractinsky et al., 2000; van der Heijden, 2003) for the relationship between perceived aesthetics and user preference after actual use; (Hassenzahl, 2004; Schrepp et al., 2006; van Schaik and Ling, 2008) for the relationship between perceived usability and user preference after actual use). This finding may result from the strong interrelationship between perceived aesthetics and perceived usability after actual use (*Hypothesis* 4).

Regarding *Hypothesis* 4, previous studies' findings for the relationship between perceived aesthetics and perceived usability after actual use were conflicting, as discussed in Section 2.3. The analysis result for our experimental data showed that there was indeed a high correlation between perceived aesthetics and perceived usability after actual use. This finding can be supported by two 2-sample t tests for the difference between the mean ratings of perceived usability after actual use in the two aesthetics conditions and for the difference between the mean ratings of perceived aesthetics after actual use in the two usability conditions. Table 12 presents the descriptive statistics and the test results. The results revealed that perceived usability and perceived aesthetics after actual use were significantly affected by the aesthetics factor and the usability factor, respectively. These results supports that, in users' responses after actual use, perceived aesthetics and perceived usability were highly interrelated, just as they were before actual use.

Furthermore, the effect of perceived aesthetics on perceived usability can be supported with the consideration of task completion time. For systems with a high level of usability, the ratings of perceived usability were significantly different by the aesthetics conditions (the mean ratings of perceived usability = 5.721 for the high aesthetics condition and 4.757 for the low aesthetics condition; t-value = 2.34; p-value = 0.026), although the task completion times were not (the task completion times = 149.2 s for the high aesthetics condition and 156.2 s for the low aesthetics condition; t-value = −0.34; p-value = 0.739). A similar result was found for systems with a low level of usability. The difference between the two aesthetics conditions was significant with respect to the perceived usability ratings (the mean ratings of perceived usability = 4.270 for the high aesthetics condition and 3.197 for the low aesthetics condition; t-value = 2.80; p-value = 0.008), but was not significant with regard to the task completion times (the task completion times = 275.6 s for the high aesthetics condition and 324.2 s for the low aesthetics condition; t-value = -1.13; p-value = 0.264). These results show that the effect of perceived aesthetics was stronger than that of objective performance, on perceived usability.

However, in analyzing the correlation between perceived aesthetics and perceived usability for each system, the correlation coefficient for System 1 was relatively low, unlike those for System 2, System 3, and System 4. Especially in the high aesthetics and the low usability conditions, the high correlation between perceived aesthetics and perceived usability after actual use did not support the argument of Lindgaard and Dudek (2003) that the correlation between these two factors was not strong. Instead, the results suggested a weak relationship or non-existence of a specific relationship between perceived aesthetics and perceived usability after actual use in the low aesthetics and the low usability conditions. These conflicting findings may result from the systems' different use purposes. The use purpose of the system in our experiment

Table 11
Two 2-sample t tests for data before actual use.

Condition	N	Mean	Std. Dev.	2-Sample t test
Perceived usability before actual use				
High aesthetics	36	5.570	1.153	t-value = 3.82; p-value < 0.001
Low aesthetics	37	4.666	1.270	
Perceived aesthetics before actual use				
High usability	35	4.236	1.257	t-value = 2.02; p-value = 0.047
Low usability	38	3.641	1.261	

Table 12
Two 2-sample t tests for data after actual use.

Condition	N	Mean	Std. Dev.	2-Sample t test
Perceived usability after actual use				
High aesthetics	36	4.955	1.301	t-value = 3.03; p-value = 0.003
Low aesthetics	37	3.956	1.512	
Perceived aesthetics after actual use				
High usability	35	4.049	1.267	t-value = 3.19; p-value = 0.002
Low usability	38	3.136	1.167	

was to obtain useful information and resources about allergies while that of the "pen" site used in Lindgaard and Dudek's (2003) experiment was to provide a commercial venue via which a product could be purchased and used.

6.3. Comparison of interactions before and after actual use

In Hypothesis 5-1, we see the importance of considering the occurrence of actual use as an influential factor of perceived aesthetics. To wit, the results showed that the difference between the high and the low aesthetics conditions and the difference between the high and the low usability conditions were significant with respect to users' ratings of perceived aesthetics. These significant differences were consistent with the findings described in Sections 6.1 and 6.2. The effect degrees of actual use on perceived aesthetics were not affected by the aesthetics and the usability conditions. Regarding Hypothesis 5-2, interestingly, users' ratings of perceived usability before actual use were significantly higher than those after actual use in systems with a low level of usability. As stated previously, users can better identify the usability problems of a system after actual use than before actual use. More exposure to a system's usability can result in a low rating of perceived usability. In that users can be more exposed to usability problems in the low usability condition than in the high usability condition, the effect degrees of the actual use on perceived usability may differ by the usability levels. For Hypothesis 5-3, only the interaction effect between usability and actual use on user preference was found. This implies that the effect degrees of the actual use on user preference can depend on the usability conditions. According to the Tukey's post hoc tests for the difference between the user preference ratings before and after actual use in the two usability levels, users' ratings of user preference after actual use were significantly lower than those before actual use in systems with low usability. This tendency is similar to that of perceived usability. This indicates the possibility that, for a non-usable system, user preference can be strongly affected by perceived usability.

7. Conclusion

To understand the process of users' preference-making based on usability and aesthetics, we tried to illuminate the relationships among usability/aesthetics features, perceived usability/aesthetics, and user preferences before and after actual use by conducting a systematic and empirical investigation. This study introduced a new methodology where usability, aesthetics, and occurrence of actual use were simultaneously considered in a more complete setting. Also, the previous studies' findings were confirmed and the unclear or under-examined relationships among the factors were clarified. The results were analyzed from different viewpoints based on the experimental data.

In users' interactions before actual use, user preference was significantly affected by the aesthetics factor but marginally affected by the usability factor. On the other hand, in users' responses after actual use, user preference was significantly influenced by both the usability and the aesthetics factors. Overall, a strong interrelationship between perceived usability and perceived aesthetics, as well as high correlations between perceived usability/aesthetics and user preference were found, irrespective of the occurrence of actual use. However, in the low usability and the low aesthetics conditions, perceived usability and perceived aesthetics after actual use were in a relatively low correlation. In addition, the difference between before and after actual use was significant with respect to perceived usability, perceived aesthetics, and user preference.

These findings have several implications in developing a more preferable computer-based product or system:

- Users' subjective perceptions of usability and aesthetics need to be sufficiently considered with usability and aesthetics features based on purely on existing design guidelines or requirements, in that designs following the guidelines or requirements may not be directly linked to high levels of perceived usability and perceived aesthetics.
- Understanding the effects of perceived usability/aesthetics on user preference and the correlations between perceived usability/aesthetics and user preference can be helpful in predicting a level of user preference.
- Designers need to utilize the strong interrelationship between perceived usability and perceived aesthetics in developing a preferable application. For example, users may perceive a product or system as usable more than its actual usability, if it is designed with a high aesthetic quality.
- The occurrence of actual use can be considered as an important factor in understanding user preferences. Although the processes of users' preference-making before and after actual use seemed analogous, the effects of actual use on users' perceptions and judgments (i.e., perceived usability, perceived aesthetics, and user preference) were quite clear in the experiment.

Further studies are needed to resolve certain limitations of this study. First, the experimental domain was "Health Center for Allergies", a system providing some information and resources about allergies. According to the categories by use purposes (Korgaonkar and Wolin, 1999; Lindgaard and Dudek, 2003), the system used in the experiment can be regarded as an information system. In future work, it is necessary to determine what results will be found for other systems in different domains (e.g., systems with a less instrumental context – entertainment, communication, and commerce). This need is supported by the finding of Yamamoto and Lambert (1994), namely, that the impact of perceived aesthetics on user preference varied depending on what the product is. Second, the use situation was contrived, and the content used in the systems was not an influential factor in our experiment. Although participants might not really need "Health Center for Allergies", we needed to assume that they were potential users of our systems and to recruit participants who had no or little prior experience on systems like ours in order to control biases caused by their experience. In a similar vein, the same content was provided regardless of usability conditions; that is, the information quality was not considered in our experiment. To broaden and confirm our findings, it is necessary to consider these issues in future work. Third, it should also be noted that all of the participants in our experiment were engineering students and they were heavily skewed to males. The generalization of the experimental findings can be restricted by this homogeneous nature of the participants. In that we could control the factors related to using computers to some extent by recruiting such participants, however, the results can be sufficiently meaningful. Fourth, cultural effects were not adequately managed though we tried to minimize such effects on users' perceptions and judgments. Actually, there was a trade-off between our criteria and the control of cultural factors in recruiting participants in the experiment. Further studies need to more deeply consider the cultural background of participants and to compare it with that of designers. Finally, further studies need to consider more deeply how to manipulate usability and aesthetics. For more exact analysis, the difference between the perceived usability ratings in the high and the low usability conditions and the difference between the perceived aesthetics ratings in the high and the low aesthetics conditions should be controlled at a more similar level. The gap between these differences may affect per-

ceived usability and perceived aesthetics, and finally user preference.

References

Ben-Bassat, T., Meyer, J., Tractinsky, N., 2006. Economic and subjective measures of the perceived value of aesthetics and usability. ACM Transactions on Computer–Human Interaction (TOCHI) 13 (2), 210–234.

Bernard, M., 2000. Constructing user-centered websites: design implications for content organization. Usability News 2 (2).

Bloch, P.H., 1995. Seeking the ideal form: product design and consumer response. Journal of Marketing 59 (3), 16–29.

Cheng, F.-F., Wu, C.-S., Yen, D.C., 2009. The effect of online store atmosphere on consumer's emotional responses – an experimental study of music and colour. Behaviour & Information Technology 28 (4), 323–334.

Chin, J.P., Diehl, V.A., Norman, L.K., 1988. Development of an Instrument Measuring User Satisfaction of the Human–Computer Interface. ACM Press, New York, NY.

Conklin, S.M., Koubek, R.J., Thurman, J.A., Newman, L.C., 2006. The effects of aesthetics and cognitive style on perceived usability. In: Paper Presented at the Human Factors and Ergonomics Society Annual Meeting, Springer, CA.

de Angeli, A., Sutcliffe, A., Hartmann, J., 2006. Interaction, usability and aesthetics: what influences users' preferences? In: Paper Presented at the 6th ACM Conference on Designing Interactive Systems, University Park, PA.

Dillon, A., 2001. Beyond usability: process, outcome and affect in human computer interactions. Canadian Journal of Information Science 26 (4), 57–69.

Hall, R.H., Hanna, P., 2004. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. Behaviour & Information Technology 23 (3), 183–195.

Hartmann, J., Sutcliffe, A., de Angeli, A., 2008. Towards a theory of user judgment of aesthetics and user interface quality. ACM Transactions on Computer–Human Interaction (TOCHI) 15 (4) (Article No. 15).

Hartson, H.R., Andre, T.S., Williges, R.C., 2003. Criteria for evaluating usability evaluation methods. International Journal of Human–Computer Interaction 15 (1), 145–181.

Hassenzahl, M., 2004. The interplay of beauty, goodness, and usability in interactive products. Human–Computer Interaction 19 (4), 319–349.

Hassenzahl, M., 2008. Aesthetics in interactive products: correlates and consequences of beauty. In: Schifferstein, H., Hekkert, P. (Eds.), Product Experience. Elsevier, San Diego, CA, pp. 287–302.

Hassenzahl, M., Kekez, R., Burmester, M., 2002. The importance of a software's pragmatic quality depends on usage modes. In: Paper Presented at the 6th International Conference on Work With Display Unit (WWDU), Berlin, Germany.

Head, A.J., 1999. Design Wise: A Guide for Evaluating the Interface Design of Information Resources. Information Today, Inc., Medford, NJ.

ISO, 1998. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) – Part 11: Guidance on Usability.

Jordan, P.W., 1998. Human factors for pleasure in product use. Applied Ergonomics 29 (1), 25–33.

Keinonen, T., 1997. Expected usability and product preference. In: Paper Presented at the Conference on Designing Interactive Systems. Amsterdam, The Netherlands.

Korgaonkar, P.K., Wolin, L.D., 1999. A multivariate analysis of web usage. Journal of Advertising Research 39 (2), 53–68.

Kurosu, M., Kashimura, K., 1995. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In: Paper Presented at the SIGCHI Conference on Human Factors in Computing Systems, Denver, CO.

Lavie, T., Tractinsky, N., 2004. Assessing dimensions of perceived visual aesthetics of web sites. International Journal of Human–Computer Studies 60 (3), 269–298.

Lewis, J.R., 1991. User Satisfaction Questionnaires for Usability Studies: 1991 Manual of Directions for the ASQ and PSSUQ (No. 54.609). Boca Raton, FL: Tech. Rep.

Lewis, J.R., 1992. Psychometric evaluation of the post-study system usability questionnaire: the PSSUQ. In: Paper Presented at the Human Factors Society 36th Annual Meeting, Santa Monica, CA.

Lewis, J.R., 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human–Computer Interaction 7 (1), 57–78.

Lewis, J.R., 2002. Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human–Computer Interaction 14 (3&4), 463–488.

Lindgaard, G., Dudek, C., 2003. What is this evasive beast we call user satisfaction? Interacting with Computers 15 (3), 429–452.

McCracken, D.D., Wolfe, R.J., 2004. User-Centered Website Development: A Human–Computer Interaction Approach. Pearson Prentice Hall Inc., Upper Saddle River, NJ.

Nielsen, J., 1993. Usability Engineering. Academic Press, San Diego, CA.

Nielsen, J., 1994. Heuristic evaluation. In: Nielsen, J., Mack, R.L. (Eds.), Usability Inspection Methods. Wiley, New York, NY, pp. 25–62.

Nielsen, J., 1996. Seductive User Interface. <http://www.useit.com/papers/seductiveui.html>.

Norman, D.A., 2002. The Design of Everyday Things. Basic Books, New York, NY.

Norman, D.A., 2004. Introduction to this special section on beauty, goodness, and usability. Human–Computer Interaction 19 (4), 311–318.

Pandir, M., Knight, J., 2006. Homepage aesthetics: the search for preference factors and the challenges of subjectivity. Interacting with Computers 18 (6), 1351–1370.

Pearrow, M., 2000. Web Site Usability Handbook. Charles River Media, Inc., Rockland, MA.

Rosson, M.B., Carroll, J.M., 2002. Usability Engineering: Scenario-based Development of Human–Computer Interaction. Morgan Kaufmann, San Francisco, CA.

Schenkman, B.N., Jönsson, F.U., 2000. Aesthetics and preferences of web pages. Behaviour & Information Technology 19 (5), 367–377.

Schrepp, M., Held, T., Laugwitz, B., 2006. The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers 18 (5), 1055–1069.

Shneiderman, B., 1998. Designing the User Interface: Strategies for Effective Human–Computer Interaction. Addison-Wesley Longman Publishing Co., Inc., Boston, MA.

Spool, J.M., 1999. Web site Usability: A Designer's Guide. Morgan Kaufmann, San Francisco, CA.

Sutcliffe, A., Ryan, M., Doubleday, A., Springett, M.V., 2000. Model mismatch analysis: towards a deeper explanation of users' usability problems. Behaviour & Information Technology 19 (1), 43–55.

Thüring, M., Mahlke, S., 2007. Usability, aesthetics and emotions in human–technology interaction. International Journal of Psychology 42 (4), 253–264.

Tractinsky, N., 1997. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: Paper Presented at the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA.

Tractinsky, N., Katz, A.S., Ikar, D., 2000. What is beautiful is usable. Interacting with Computers 13 (2), 127–145.

Tullis, T.S., Kodimer, M.L., 1992. A comparison of direct-manipulation, selection, and data-entry techniques for re-ordering fields in a table. In: Paper Presented at the SIGCHI Conference on Human Factors in Computing Systems, Monterey, CA.

van der Heijden, H., 2003. Factors influencing the usage of websites: the case of a generic portal in the Netherlands. Information & Management 40 (6), 541–549.

van Schaik, P., Ling, J., 2008. Modelling user experience with web sites: usability, hedonic value, beauty and goodness. Interacting with Computers 20 (3), 419–432.

Wang, C.-H., 2001. A survey of design guidelines for usability of web sites. In: Smith, M.J., Salvendy, G., Harris, D., Koubek, R.J. (Eds.), Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, vol. 1. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 183–187.

Yamamoto, M., Lambert, D.R., 1994. The impact of product aesthetics on the evaluation of industrial products. Journal of Product Innovation Management 11 (4), 309–324.

Sangwon Lee has received his Bachelor degree from Korea University in 2004 and his Master degree from the Pennsylvania State University in 2006. He is currently a Ph.D. candidate of the Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at Penn State. His academic interests lie in Human Factors, especially, usability engineering and Human–Computer Interaction: constructing a conceptual framework for user preferences, developing a heuristic usability evaluation method, and analyzing human cognitive processes based on users' judgments for interactive applications.

Richard J. Koubek is Dean for College of Engineering at the Louisiana State University. He was Professor and Head for the Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at the Pennsylvania State University. He also held the Peter and Angel Dal Pazzo Department Head Chair at Penn State. Dr. Koubek's research focuses on usability, human aspects of manufacturing, and Human–Computer Interaction. Dr. Koubek received his Ph.D. in Industrial Engineering from Purdue University.