Evaluating the Aesthetics of Websites

Using materials analysis and visual design heuristics

by
Matthew Theuma

Project report submitted in part fulfilment of the requirements for the degree of Master of Science (Human-Computer Interaction with Ergonomics) in the Faculty of Life Sciences, University College London, 2007.

MSc Project: Evaluating the Aesthetics of Websites

Using materials analysis and visual design heuristics

by Matthew Theuma

Abstract

Two studies were conducted to explore different but complimentary areas of visual design in relation to HCI, to further the recent expanse of research into non-utilitarian design.

The first study used a materials analysis approach to validate the AttrakDiff Lite questionnaire and explore the nature of immediate affective reactions, building upon prior research by Plumley (2006), who had used subjects analysis. Participants used the questionnaire to rate 30 web pages that were presented for 50 ms each. A factor analysis of questionnaire responses revealed there were two factors that explained 86.7% of the variance. These two factors did not see a clear split between pragmatic and hedonic components however, which may present support for Lavie & Tractinsky's (2004) theory of “classical” and “expressive” aesthetics.

In the second study, 17 visual design heuristics were created based on the principles in “Designing Visual Interfaces” by Mullet and Sano (1995). A pilot study was used to test the inter-rater reliability, after which the heuristics were used to analyse all 30 of the web pages used in the first study. Factor scores were calculated using data from the first study and compared with the heuristic scores. The results showed that there was a correlation between the judgements made by participants in the first study, and those made by the HCI practitioners performing a heuristic analysis in the second study. This not only validates the AttrakDiff Lite questionnaire further, but also suggests that visual appeal can be assessed within 50 ms, which mirrors the findings of Lindgaard et al. (2006).

Acknowledgements

Firstly, I would like to take this opportunity to thank my project supervisor Paul Cairns for all his help and support.

Secondly, I would also like to thank all the participants who took part in the studies. Thirdly, I would like to thank Ryan Barkataki, Angeliki Chrysafidi, Julie Schiller and David Whittle for helping with the development and testing of the visual design heuristics.

Finally, I would also like to thank my dad for his tireless proof-reading efforts.

Contents

Figures

1. Introduction

Since the 1980s, the field of Human-Computer Interaction (HCI) has been providing theories, concepts, methods, and practices for the design of interfaces, focusing on usability of digital products (Udsen & Jørgensen, 2005). Traditionally studies of usability have focused on the pragmatic quality of products, which include: efficiency, effectiveness and learnability.

The proliferation of digital products and the World Wide Web within people's homes and places of work, means that the usability of digital products and services is becoming of paramount importance.

The international standard ISO 9241-11 ("ISO 9241: Part 11," 2007) states that there are three cornerstones of usability: efficiency, effectiveness and satisfaction. It is often the case that efficiency and effectiveness are given more focus, because they can be objectively assessed using measurable attributes such as learnability, speed of user task performance and user error rates (Hartson, 1998; Norman, 2004a). Satisfaction on the other hand is subjective, (Hassenzahl, 2001) and sometimes seen as a by-product of general usability, it is not only dependent on the person, but also on the context in which the product is being used (Lindgaard & Dudek, 2003).

There is an emerging trend in research literature that acknowledges the need to explore more creative and innovative forms of human-computer interaction (Udsen & Jørgensen, 2005). Recently there has been a drive to look beyond function and explore the different elements that influence people's perceptions of usability, one such area is that of aesthetics.

There are two very compelling reasons why aesthetics of interactive systems is drawing increasing attention. Several studies have shown that there is a correlation between aesthetics and perceived usability of a system (Kurosu & Kashimura, 1995; Tractinsky, 1997; Tractinsky, Katz & Ikar, 2000) although some doubt that such a link exists (Hassenzahl, 2004). Secondly, it has been theorised that aesthetic elements of an interface are processed first and influence subsequent perceptions of the system. Studies have shown that these aesthetics impressions are formed in as little as 50 milliseconds (Lindgaard, Fernandes, Dudek, & Brown, 2006; Tractinsky, Cokhavi, Kirschenbaum, & Sharfi, 2006).

While existing literature about aesthetics in relation to consumer products is quite substantial (e.g. Norman, 2004b), empirical research in the field of HCI is just beginning (Hassenzahl, 2004; Tractinsky, 1997). Research into the influence of hedonic qualities, as well as beauty and goodness, has attempted to define what these principles mean for the HCI community and create a groundwork for future studies.

This study has covered two different aspects of the relationship between aesthetics and usability. The first half has looked at immediate affective reactions in users and how these can be measured using a short-form materials analysis questionnaire, which follows on from some previous research (Plumley, 2006). The second half proposes a set of heuristics for analysing visual design based on the principles outlined by Mullet and Sano (1995), in their book "Designing Visual Interfaces".

These two studies explore different but complimentary areas of visual design in relation to HCI and more specifically websites. While the first study looks at how aesthetics can be measured in a way that allows different websites to be compared, the second study provides a set of

heuristics that can be used to analyse and improve visual design. Each study explores a part of HCI that has been often neglected in research. Individually the studies provide insights into how aesthetics can be measured and applied to websites. Combined, the studies provide a comprehensive look at how website aesthetics can be assessed.

2. Background

2.1 A brief history of aesthetics

Aesthetics has a very far reaching history and its development was largely influenced by art in six great ancient civilisations; Egypt, Mesopotamia, Greece, Rome, India, and China. It was Greece however that had the greatest influence on the development of aesthetics in the West, where originally aesthetic values appeared as a reformulation of ideas about beauty (Lavie & Tractinsky, 2004). For example, the Greek philosopher Plato felt that an object's beauty came from its proportions, harmony and unity, while Aristotle believed beauty was derived from order, symmetry and definiteness.

Modern Western aesthetic thinking is thought to have first emerged in 1750, when the German philosopher Baumgarten published the book Aesthetica, giving rise to aesthetics as a philosophical discipline (Udsen & Jørgensen, 2005). Baumgarten introduced the idea of aesthetics as the study of how people experience the world through their senses. He argued that perfection of sensory awareness is to be found in the perception of beauty. By the end of the eighteenth century, "aesthetics" had become an integral part of general language (Lavie & Tractinsky, 2004).

Beauty for the sake of beauty is a relatively new concept in Western society. Greek philosophers judged artefacts primarily on how suitable they were for their intended use and on the quality of their craftsmanship. During the renaissance, beauty was seen as a part of nature and beauty in design was a result of natural beauty (Lavie & Tractinsky, 2004).

Traditionally, approaches to aesthetics have been developed by studying works of art, architecture and natural landscapes. This has lead to much debate between philosophers attempting to define aesthetics and resulted in extensive empirical studies that have tried to take a more scientific approach to aesthetics.

The role of aesthetics in design started out in a simple fashion, because beauty and usefulness were seen as one and the same. However, the industrial revolution placed an emphasis on mass production and aesthetic considerations were neglected. In the early twentieth century industrial design pioneers Loewy and Dreyfuss, introduced an aesthetic approach to their mass produced designs because they recognised its marketing value (Petroski, 1993 as cited in Lavie & Tractinsky, 2004).

Aesthetics can play a key role in satisfying the human requirements of a design. Yamamoto and Lambert (1994) discovered that aesthetically pleasing properties have a positive influence on people's preference of products, leading them to conclude that industrial design can be a "competitive weapon" (p. 317).

In his book "The Design of Everyday Things", Donald Norman (1998, pp. 151-155) ridicules designers who put aesthetics before usability and while this may be the case in some situations, recently it has been argued that too much emphasis is placed on assessing performance of products and not enough on aspects that enhance user satisfaction, like aesthetics.

2.2 The rise of non-utilitarian design

Recently theorists have been discussing how non-utilitarian aspects, such as beauty, enjoyment and fun can be integrated into HCI (Blythe, Overbeeke, Monk, & Wright, Eds., 2004; Hassenzahl, 2004; Shneiderman, 2004). Previously HCI has been suspicious about aesthetics, and highly aesthetic products have sometimes been accused of hiding "harm behind [their] beauty" (Russo & De Moraes, 2003, p. 146, as cited in Hassenzahl, 2004).

Shneiderman (2004) discusses at length how interfaces can be designed to be more fun. He states that designers are now "beginning to develop theories of user engagement through fun-features: alluring metaphors, compelling content, attractive graphics, appealing animations, and satisfying sounds" (p. 49). Shneiderman's view is that these features should be attended to after functionality and usability have been accommodated in a product's design, but others would argue that a more holistic view should be taken of the elements that fulfil users' needs and make a product successful (Hassenzahl, 2004).

Some theorists have suggested that even the broadest definitions of usability, lack consideration of enjoyment and satisfaction (Hassenzahl, 2001). As a result there has been an attempt to introduce new terms within HCI that can better explain usability factors, such as: pleasure (Jordan, 2000), emotion (Norman, 2004b), beauty (Tractinsky, 1997) and goodness (Hassenzahl, 2004).

In his book "Emotional Design", Norman (2004b) states that aesthetic design can be more influential in affecting user preferences than functional usability. This claim reflects well-established knowledge in marketing, product design, and social psychology, which is that beauty influences people's reactions. When applied to a product, it is a major factor that can determine marketplace success or failure (De Angeli, Sutcliffe, & Hartmann, 2006).

This view is supported by some empirical studies, such as the research by Schenkman and Jönsson (2000) who asked participants to rate several web pages based upon: overall impression, beauty, meaningfulness, comprehension, order, legibility, and complexity. The results indicated that beauty was the closest related to overall impression.

Aesthetics has been shown to go beyond usability in affecting user perceptions. A study by Karvonen (2000) documented how "beauty may be the decisive factor when wondering whether or not to trust a service enough to conduct business online" (p. 87). Fogg, et al. (2003) had similar findings when they carried out a large study of over 2,500 participants, which investigated the factors that affected user perceptions of website credibility. They found that it was the "design look" (p. 5) that generated most user comments, and had the greatest influence on user opinions.

Non-utilitarian aspects of design are being taken extremely seriously by theorists and practitioners alike. This is because studies have shown that aesthetics can be a good (if not the best) indicator of a product's overall impression and general user satisfaction.

2.3 The link between aesthetics and usability

In 1995 two researchers Kurosu and Kashimura presented a conference paper that explored a possible link between aesthetics and perceived usability. Their study involved the use of different ATM machines, which were identical in function, had the same number of buttons and all operated in the same way. However, some had the buttons and screen layouts arranged attractively and others unattractively. Kurosu and Kashimura (1995a) found that the attractive ones were perceived to be easier to use and there was a high correlation between judgements of aesthetics and apparent usability ($r = 0.589$).

In the second part of their study, Kurosu and Kashimura (1995b) asked participants to perform ratings of usability and aesthetics before and after interacting with their ATM designs. They found that pre and post interaction ratings of aesthetics were stable ($r = 0.896$), but the correlation between perceptions of usability and actual usability were negligible ($r = 0.286$).

This suggests that while apparent usability, based on visual assessment was influenced by aesthetics the ratings of actual usability were not biased in the same way. Participants changed their ratings based on their interaction experience.

Noam Tractinsky (1997; Tractinsky et al., 2000), conducted two further studies in order to validate and replicate the findings of Kurosu and Kashimura. In the first study Tractinsky (1997) employed the same procedure, but also added an additional experiment that separated aesthetics and usability evaluations into two separate stages to eliminate any response dependency. The results matched those of Kurosu and Kashimura, with participants' ratings of usability showing a strong positive correlation to aesthetics.

In the second study Tractinsky et al. (2000) explored both pre and post interaction ratings of aesthetics and usability. Once again there was a strong relationship between ratings of interface aesthetics and perceptions of usability, however, contrary to Kurosu and Kashimura's findings, the relationship was sustained even after actual system use. Tractinsky et al. (2000, p. 140) goes so far as to say that "post-experimental perceptions of system usability were affected by the interface's aesthetics and not by the actual usability of the system".

De Angeli, Sutcliffe, and Hartmann (2006) performed a similar study to Kurosu, Kashimura and Tractinsky using websites instead of ATM machines. The two websites used had identical content, but radically different user interface design. One was a traditional menu-based style, the other exploited animated metaphors and more aesthetic features (De Angeli et al. 2006, p.272).

De Angeli et al. (2006, p.279) state that they discovered a "framing effect" whereby user judgements of aesthetics and overall preference varied according to the task they were set. Overall users preferred the more aesthetic metaphor-based website, even though it had been rated as having worse usability. However, when given a task that involved more serious use the traditional menu-based site was favoured. De Angeli et al. (2006) argue that judgements of

aesthetics are contextually dependent and that users' perceptions of aesthetics in relation to website quality are more complex then suggested in previous studies.

Not all research has replicated these findings. Hassenzahl (2004) carried out two studies which aimed to explore the interplay between user-perceived usability, hedonic attributes, satisfaction (goodness) and beauty of four different MP3-player skins, both pre and post interaction.

In the first study, participants were asked to rate the qualities of the MP3-player skins based on visual inspection alone, using the AttrakDiff 2 questionnaire (see section 2.5). The results showed that judgements of beauty were unrelated to perceived usability of the software interface. In the second study, participants were asked to rate the qualities of the MP3-player skins before and after interaction, again using the AttrakDiff 2 questionnaire. The results showed that while ratings of beauty were not affected by interaction, the judgements of usability and overall goodness were significantly altered. Hassenzahl's (2004) results indicate that users are capable of changing their opinions of a system quite readily after interaction and their judgements are not biased by appearance.

Hassenzahl (2004) concluded that "no substantial relation between actual or perceived usability and beauty was found" (p. 345). However, he did find that the aesthetically pleasing skins meant that the MP3-players were judged to be better than the less-pleasing ones, even if they were equally easy to use. Hassenzahl (2004) suggests that the difference is not in the beauty, but rather the goodness. He also suggests that goodness is a combination of hedonic identification, perceived usability and mental effort (which is related to actual usability), while beauty is solely related to hedonic identification.

Despite Hassenzahl's findings, a recent study by (Schrepp, Held, & Laugwitz, 2006) (2006) into the effect of hedonic quality on the "attractiveness" of business management software, found a high correlation between hedonic and pragmatic qualities. The study used three variants of business management software, which were all identical in functionality, but had different visual design and navigation elements. Users rated screen-shots of the interfaces showing a typical business scenario using the AttrakDiff 2 questionnaire.

Schrepp et al. (2006) were expecting to find that pragmatic quality would have a stronger influence than hedonic quality on the user preferences of the software, due to the fact that "users perceive business management software mainly as a tool which helps them to manage their daily work" (p. 1056), but this was not the case. The results of the study indicate that both pragmatic quality and hedonic quality seem to have nearly the same impact on user preferences.

These different results may be explained by the findings of Lindgaard and Dudek (2003), which showed that the relation between appeal and perceived usability does not exist for all types of interfaces. Schrepp et al. (2006) argue that the importance of user experience factors in the evaluation process may vary depending on the qualities of a particular user interface.

Although the link between aesthetics and perceived usability is still being debated, researchers such as Tractinsky (2000) and Norman (2004b) argue that an interface's aesthetic qualities may trigger an affective response in users that influences subsequent cognitive processes. As a result of this, a user's first impressions may greatly influence their overall judgment of a

system, because the decision to like or dislike an incoming stimulus is based on the interpretation of a visceral response that is felt rather than thought (Damasio, 2000, as cited in Lindgaard, 2006).

2.4 The importance of first impressions

First impressions can have a very powerful influence, for example social psychology research has shown that initial social perceptions persist even after strong evidence is presented to the contrary (Tractinsky et al., 2000). In marketing research literature, this effect is sometimes referred to as a 'halo effect'. The halo effect occurs when a first impression affects the evaluation of other attributes of a product. Alternatively the phenomenon is sometimes referred to as a cognitive 'confirmation bias' (Mynatt, Doherty & Tweeney, 1977, as cited in Lindgaard & Dudek, 2003) whereby people tend to search exclusively for evidence that supports their view, while ignoring evidence to the contrary, meaning that they will not revise their judgements even if strong opposing evidence is present.

As a result of this influence, a very positive first impression may mean that a person disregards negative aspects encountered later. Alternatively a negative first impression, may lead to a user overlooking positive attributes and therefore fail to revise their initial judgement.

A study carried out by Zajonc (1980, as cited in Lindgaard et al., 2006) demonstrated that preferences for stimuli could develop in as little as 1-5 milliseconds. These preferences can be strengthened by repeat exposure, in what is known as "mere exposure" (p. 116) effect. Therefore, not only can first impressions influence subsequent interactions with a product, but they can also be formed within milliseconds.

Norman (2004b) has written extensively about the way in which emotional responses to a product can affect the way in which it is perceived. He states that there are three levels of human information processing, which give rise to three different levels of beauty. These are the visceral and behaviour levels which are subconscious and the reflective level which is conscious (Norman, 2004a).

The level most closely tied to first impressions is the visceral level. It is perceptually based and gives rise to immediate judgements such as good or bad, safe or dangerous. Norman (2004a, p. 5) states that "if there is any level at which beauty is associated with the object itself, it is at the visceral level". Only simple pattern recognition is possible at this level, but it can bias the other levels of processing.

Every action a person makes had a cognitive component to assign meaning and an affective component to assign value. Isen (1983, as cited in Tractinsky, Katz, & Ikar, 2000) demonstrated that positive affect can improve decision making and creativity. Based on these findings, Tractinsky et al. (2000) states that the positive affect created by aesthetically pleasing interfaces may therefore help to improve users' performance. Norman (2004b) comments on how a state of positive or negative affect begins at the visceral level, therefore it takes only milliseconds to form. He also suggests that someone who is in a happy and relaxed mood is more creative and more able to overlook and cope with usability problems, due to neurotransmitters broadening brain processing. By contrast, when someone is anxious or otherwise unhappy, neurotransmitters bias the brain to become more focused and thus users

will be more aware of any problems encountered. Based on this fact, Norman (2004b, p. 26) states that "designers can get away with more if the product is fun and enjoyable".

A recent paper by Lindgaard et al. (2006) has begun to uncover what rapid information processing and the halo effect can mean for HCI. Three studies were conducted to discover how quickly people can form an opinion about the visual appeal of web pages. In the first study, participants rated the visual appeal of web pages that were presented to them for 500 milliseconds each. The second study replicated the first, but each web page was also rated on seven design dimensions. The third study again replicated the 500 milliseconds test, but also added a 50 millisecond condition.

Throughout each of the studies, ratings of visual appeal were highly correlated, including between the 50 ms and 500 ms conditions ($r = 0.947$). Lindgaard et al. (2006) suggests that this means that visual appeal can be assessed in as little as 50 ms and therefore the first impression of a web page can be formed in the same amount of time.

Tractinsky et al. (2006) replicated and extended the study that Lindgaard et al. (2006) carried out. In their experiment, the average attractiveness ratings given after 500 ms were highly correlated with those given after 10 seconds. In addition, extreme evaluations (both positive and negative) were made faster than moderate evaluations, which Tractinsky et al. (2006) suggests supports the hypothesis of immediate impression. Perhaps more importantly, it is also proposed that the results indirectly reveal that visual aesthetics have an important influence of user's evaluations of "IT artefacts" and in their attitudes towards interactive systems.

A halo effect resulting from good aesthetics cannot cover severe problems however, as Hartmann, Sutcliffe, and De Angeli, (2007) point out, the effect may be lost altogether if users lose trust in a system.

While it seems clear that HCI researchers understand the importance of taking a more holistic view of user experience, there are still ongoing efforts to try and define the elements that contribute to it, such as aesthetics.

2.5 Measuring aesthetics

Some efforts have been made by researchers to quantify aesthetics and apply scientific measurements to the concept. In 1932, American mathematician Birkhoff attempted to produce a mathematical theory for measuring aesthetics. He devised a formula to measure the relative importance of aesthetic factors involved in an aesthetic experience (Solomon, 2004). Birkoff's endeavour to devise this universal formula was incredibly ambitious and did not succeed in becoming widely adopted. Approaches which attempt to enforce an "absolute measure" (Solomon, 2004) for aesthetics are often argued against because of the subjective nature of the topic.

As discussed in Section 2.3, in the field of human-computer interaction, most of the research into aesthetics has concentrated on exploring and defining the link between aesthetics and usability. These types of studies centre on gathering the perceptions of users, not investigating the aesthetics of the artefacts in question. There have been however, a few studies which have looked at uncovering ways of measuring aesthetics in an HCI context.

A study by Lavie and Tractinsky (2004) aimed to develop a way of measuring perceived website aesthetics. The four experiments conducted in the study identified two factors that were affecting participant's ratings of aesthetics. They are described as "classical aesthetics" which emphasises orderly and clear design and "expressive aesthetics" which comes from a designer's creativity and originality. Both of these aesthetic dimensions were measured on a five item scale and proved to be clearly distinguishable from each other. The intention of the study was to reinforce aesthetics as "a distinguishable, measurable construct in the field of human–computer interaction" (p. 292) and explore the relationship between the perception of aesthetics and usability.

In a similar vein, while researching the interplay between beauty, goodness and usability, Hassenzahl (2003, 2004) developed the AttrakDiff 2 questionnaire. The questionnaire makes use of a model also developed by Hassenzahl, which assumes users define product attributes by combining a product's features with personal experiences and standards (Hassenzahl, 2004). The questionnaire was used by Hassenzahl (2004) in his study of MP3-payer skins and it has also been used by Schrepp, et al (2006) in their study of business management software. However, the aim of these studies was to discover the elements that affect user perceptions, not analyse the artefacts used as stimuli.

The AttrakDiff 2 questionnaire consists of twenty-one 7-point items with bipolar verbal anchors, which measure pragmatic qualities, hedonic qualities and overall appeal. Originally the questionnaire was in German and has subsequently been translated into English. The result is that some of the questions can be unclear or seem out of place when analysing websites.

Professor Monk of York University (personal communication, June 13, 2007) developed the questionnaire further and created the AttrakDiff Lite questionnaire. His aim was to create the shortest useful version of AttrakDiff for evaluating websites. Monk's questionnaire also measures pragmatic qualities, hedonic qualities and overall appeal but has been refined so that it only contains ten items (see Section 4).

The type of studies undertaken by Lavie and Tractinsky (2004) and (Schrepp, Held, & Laugwitz, 2006)) rely on a high number of participants, with a relatively low number of artefacts to analyse. Although the studies did explore the measurement of aesthetics, they primarily concentrated on the perceptions of participants. In contrast, the AttrakDiff Lite is a materials analysis questionnaire, so it is intended for use when there are a greater number of artefacts than participants. This is because it is intended to be used for comparing websites not participants.

A study by Plumley (2006) made use of the AttrakDiff Lite questionnaire. The study explored the association between perceptions of aesthetics and usability by performing a subjects analysis, and at the same time began work on validating AttrakDiff Lite. This study continued work on validating the questionnaire using a materials analysis approach. Details of the study are discussed in section 5.

(Lavie & Tractinsky, 2004) emphasise that developing valid aesthetic measures is essential for future study, which will facilitate better understanding of human-computer interaction. By

measuring aesthetics, a greater understanding of how to approach visual design in HCI can be gained.

2.6 Understanding visual design

Aesthetics is an elusive and confusing construct, furthermore the similarity and overlap between aesthetics and beauty is unclear. As Lindgaard et al. (2006) highlights, they can be interpreted as properties of objects, subjective experiences, emotional reactions or creative judgements.

Although the amount of attention paid to aesthetics is growing in the field of HCI, traditional methods of usability assessment, such as heuristic evaluation only contain minimal visual design components.

In their book "Designing Visual Interfaces", Mullet and Sano (1995) describe a set of "fundamental techniques used routinely by practitioners of communication-oriented visual design" (p. xi). These techniques are intended to be used by anyone who wants to enhance the visual quality of a graphical user interface. As Sutcliffe (2001) points out, interaction designers such as Mullet and Sano (1995) have given guidance in the form of examples and scenarios, but it is up to the practitioner to interpret the techniques and apply them in a design context.

More recently Lawrence and Tavakol, (2006) published a book entitled "Balanced Website Design", which promises to help in "optimising aesthetics, usability and purpose". While Lawrence and Tavakol (2006) provide a good procedure for creating a website while keeping aesthetics in mind, they do not provide the same focus on visual design provided by Mullet and Sano (1995).

Other sources of guidance in visual design for HCI practitioners tend to come from style guides, which are aimed at specific development environments, such as Apple's OS X (Apple Human Interface Guidelines, 2007) or Java (Java Look and Feel Design Guidelines, 2001).

While there is a large amount of literature available that can help designers (particularly web designers) to understand principles of usability, the same cannot be said for principles of visual design. Sutcliffe (2001, p. 189) argues the reason for this is because the visual design community follows an "experiential approach" rather than an engineering design philosophy. This means that articulating design principles is contentious and not encouraged. Despite this, he has developed a set of heuristics for attractiveness and aesthetic design based on recommendations by Kristof and Satran (1995, as cited in Sutcliffe, 2001) and Mullet and Sano (1995).

The heuristics created by Sutcliffe (2001, 2002) have been used to assess the visual design of web pages in a few studies (De Angeli et al., 2006; Hartmann et al., 2007; Sutcliffe, 2001, 2002) and are intended to be combined with additional usability heuristics. There are however only 7 heuristics specifically for visual design and Sutcliffe (2001) admits that "further research is required on articulating the design properties of attractiveness and aesthetics across different designs and user groups" (p. 197).

Another problem in trying to understand the nuances of visual design is identified by Pandir and Knight (2006). They carried out a study into web page aesthetics, to try and determine the factors that shape user preferences. They discovered low levels of agreement on pleasure and

interestingness judgements between subjects, leading them to conclude that aesthetic preferences are highly subjective and it is very hard, if not impossible to make generalisations about aesthetic qualities. Furthermore, Korsmeyer (2001, as cited in Pandir and Knight, 2006) suggests that aesthetic judgements are more diverse then non-aesthetic judgements, because there is more variation in taste.

While visual design is a complex and diverse field in itself, researchers such as Hartson (1998), have recognised that in order to advance the field of HCI, practitioners are going to have to learn new skills such as visual design, or at least have an understanding of the principles so that they can work alongside visual designers more effectively. After all, if you cannot attract a user to stay on a website, it does not matter how well designed operational usability may be (Sutcliffe, 2001).

2.7 Summary

Non-utilitarian design is receiving a great deal of attention at present in the field of human-computer interaction. As a result, the uneasy relationship between usability and aesthetics is starting to solidify and become more defined. Despite these recent improvements, a lot of research is still in its infancy and other areas are yet to be explored.

An important part of understanding how aesthetics relates to user experience is being able to measure aesthetics and offer guidance in applying aesthetic design to user interfaces. At present, there are not many methods or tools at the disposal of HCI practitioners that are suited to these tasks.

Researchers such as Sutcliffe (2001), Lavie and Tractinsky (2004) have recognised how essential it is to develop aesthetic measures in order to progress the understanding of user experience. Yet it is not only important to be able to measure aesthetics, the lessons learnt from research must also be able to be applied in design. To facilitate this it is important to consider approaches that are used in industry, such as heuristic evaluations.

3. The Present Study

This study consists of two parts, designed to explore two different, but complementary areas of aesthetics in relation to website design.

The first part of this study aims to explore the immediacy of first impressions of web pages, while also validating the AttrakDiff Lite questionnaire developed by Monk (personal communication, June 13, 2007). As discussed in Section 2.5, very few studies in the field of HCI have explored ways in which aesthetics can be measured and even fewer have taken a materials analysis approach. The study by Plumley (2006) began work on validating the constructs of AttrakDiff Lite using subjects analysis and this study will build upon Plumley's previous work, using a similar paradigm to carry out materials analysis using the questionnaire.

In this first study, participants were shown 30 web pages for 50 milliseconds each and the web pages were then assessed using the AttrakDiff Lite questionnaire. As the questionnaire is still being developed, a factor analysis was carried out using the data gathered during the study to see if responses to the questions loaded onto the expected factors.

The second part of this study involved creating a set of heuristics for analysing visual design based on the principles outlined by Mullet and Sano (1995), in their book "Designing Visual Interfaces". The aim was to create a set of visual design guidelines that HCI practitioners can use to perform a rapid formative or summative evaluation of visual design, as a supplement to a usability evaluation.

In this second study, after the visual design heuristics had been developed, they were piloted by 12 HCI practitioners, who used them to analyse five of the web pages from the first part of the study. After confirming that the judgements for each of the heuristics were stable across participants, all 30 of the web pages from the first study then underwent expert review by 2 HCI practitioners using the heuristics.

The results from both the first and second studies were then compared to reinforce the validation of both the AttrakDiff Lite questionnaire and the visual design heuristics.

4. The AttrakDiff Lite Questionnaire

The AttrakDiff Lite questionnaire is a shortened version of the AttrakDiff 2 questionnaire (Hassenzahl et al. 2003; Hassenzahl, 2004), developed by Professor Monk of York University (personal communication, June 13, 2007). It contains ten items presented on a 7-point bi-polar verbal anchor scale, designed to measure hedonic qualities, pragmatic qualities, as well as overall beauty and goodness (see Figure 1).

Pragmatic Quality Anchors
Confusing – Clearly Structured
Unpredictable – Predictable
Simple – Complicated
Practical – Impractical
Hedonic Quality – Identification Anchors
Stylish – Tacky
Cheap – Premium
Hedonic Quality – Stimulation Anchors
Dull – Captivating
Creative – Unimaginative
Overall Appeal
Bad – Good
Ugly – Beautiful

Figure 1: Items of the AttrakDiff Lite Questionnaire

The questionnaire contains four attribute groups: pragmatic, hedonic identification, hedonic stimulation and overall appeal. Pragmatic attributes are related to the user's need to achieve behavioural goals, which requires utility and usability. Thus a product which has the traditional usability aspects of efficiency, effectiveness and learnability is perceived as pragmatic. Hedonic

attributes on the other hand are related to the user's self. In a product, stimulation comes from a sense of novelty and challenge, while identification comes from an ability to effectively communicate identity and values to others. The questionnaire's fourth attribute group is overall appeal, which aims to measure beauty and goodness. Beauty and goodness are high level evaluative constructs which have varying definitions in HCI. Hassenzahl (2004) states that goodness is a combination of hedonic identification, perceived usability and mental effort (which is related to actual usability), while beauty is solely related to hedonic identification, however there are undoubtedly other factors at work, such as cultural influences.

5. Study 1 – Materials Analysis

5.1 Participants

There were 15 participants, 7 men and 8 women. All participants had previous experience with using websites and had normal or corrected-to-normal vision.

5.2 Apparatus

All participants were tested on a Toshiba Satellite A100-049 laptop with a 1.66 GHz Intel Core 2 Duo processor, 2Gb RAM, NVIDIA GeForce Go 7300 graphics card and a 15.4" WXGA TFT display with a native resolution of 1280 x 800 pixels in 32bit colour.

The experiment was designed using the software package E-Prime, version 2.0, which presented screen-shots of the chosen web pages, controlled display time and collected user responses to the questionnaire.

Screen-shots of web pages were collected using a software package called SnagIt, version 8.2.

5.3 Materials

Screen-shots of 100 web pages, which varied in both aesthetic appeal and usability, were gathered at the beginning of the study. The web pages were selected at random from a number of sources, including the Webby Awards, which are "international award honouring excellence on the Internet", WebPagesThatSuck.com which contains examples of bad aesthetics and usability and also from pre-existing bookmarks/favourites.

From the initial sample of 100 web pages, 30 were chosen as stimuli for the experiment. It was determined that 30 web pages were enough to make the study statistically sound, as 15 participants completing the questionnaire for each web page would generate a total of 450 responses. 20 of the websites were chosen because they had good-to-excellent aesthetics and the remaining 10 had poor aesthetics. The aesthetic ratings were determined based on the awards and reviews from the two source websites, together with a brief expert review. A full list of the web pages used can be seen in Appendix 4.

The web pages covered a variety of topics including: design, e-commerce, news and travel. Screen-shots of the web pages were taken within an Internet Explorer 7 browser at a resolution of 1024 x 768 pixels in 32bit colour. The screen-shots were captured using SnagIt software, which captured only the content of the web pages, discarding the interface of the browser.

5.4 Procedure

Participants were tested individually, with sessions lasting 20 minutes on average. After reading a briefing explaining the purpose of the study, participants then signed a consent form if they were happy to continue.

Participants were then seated in front of the computer and allowed to adjust the seating and monitor until they were comfortable.

So that participants could become fully accustomed to the format of the experiment, including the speed at which the stimuli would be presented and the questionnaire style, they completed 4 practice runs of the trail. The participants were presented with 4 web pages for 50 ms, which were selected at random and not part of the main trial. Each web page was followed by the 10 questions that are part of the AttrakDiff Lite questionnaire. The participants answered each question by clicking on a corresponding 7-point Likert scale.

After completing the practice runs, participants were informed that the actual trial was about to begin and their responses would be recorded. Participants were then shown 30 web pages for 50 ms in a random order, each web page was followed by the AttrakDiff Lite questionnaire.

Users controlled their own progression through the trial, after beginning the trial a blank screen was shown for 1 second, and then a web page was displayed for 50 milliseconds followed instantly by the questionnaire, which acted as a visual mask, preventing any after-image. After completing all the questions, the participant clicked “Next” at which point a blank screen was shown for 1 second, then the next web page was displayed for 50 milliseconds followed instantly by the questionnaire.

After they had completed the session, the participants were debriefed and thanked for their participation.

5.5 Results

A factor analysis was performed in order to try and validate the constructs present in the AttrakDiff Lite questionnaire. Each of the 15 participants had completed the questionnaire for 30 web pages, making a total of 450 responses. The data from these responses was used to perform a principal components analysis, the results of which can be seen in Figure 2.

The initial principle components analysis revealed that there were definitely two clear factors present, the two-factor solution explained 86.7% of the total variance. A good guide for the quality of the factors is to consider the amount of variance the factors account for. A satisfactory result in social sciences is 60% or above (Hair, Black, Babin, Anderson, & Tatham, 2005). The factors found here account for 86.7% of the variance and thus have good explanatory power for the data collected.

However, two of the questionnaire items, Practical – Impractical and Unpredictable – Predictable were loading on both factors (cross-loading). To try and resolve this conflict, a second analysis was performed, this time using direct oblimin rotation. The results can be seen in Figure 3.

Component Matrix

	Component	
	1	**2**
Confusing – Clearly Structured	**.872**	-.379
Dull – Captivating	.338	**.869**
Practical – Impractical	***-.627***	***.679***
Stylish – Tacky	**-.923**	-.290
Cheap – Premium	**.949**	.094
Bad – Good	**.958**	.137
Unpredictable – Predictable	***.759***	***-.577***
Simple – Complicated	**-.593**	.317
Ugly – Beautiful	**.919**	.330
Creative – Unimaginative	-.325	**-.913**

Figure 2: Table of Component Matrix showing the loadings of the items from the AttrakDiff Lite questionnaire

Pattern Matrix

	Component	
	1	**2**
Confusing – Clearly Structured	**.954**	-.021
Dull – Captivating	-.143	**.941**
Practical – Impractical	**-.891**	.395
Stylish – Tacky	***-.661***	***-.624***
Cheap – Premium	**.782**	.451
Bad – Good	**.767**	.494
Unpredictable – Predictable	**.954**	-.249
Simple – Complicated	**-.679**	.070
Ugly – Beautiful	**.636**	**.659**
Creative – Unimaginative	.177	**-.977**

Figure 3: Table of rotated Pattern Matrix showing the loadings of the items from the AttrakDiff Lite questionnaire. Rotation Method: Oblimin with Kaiser Normalization.

The rotation cleared up the results a little but there were still some anomalies.

All of the pragmatic quality items, as well as the overall appeal item Bad – Good, load onto factor 1. Both of the hedonic identification items, Stylish – Tacky and Cheap – Premium also load onto factor 1, although Stylish – Tacky loads onto factor 2 almost equally.

Factor 2 contains all the hedonic stimulation items and the overall appeal item Ugly – Beautiful.

Factor 1 consists of primarily pragmatic components and factor 2 consists of hedonic components, but there was not a clear split between pragmatic and hedonic items as expected. Figure 4 shows the split between the questionnaire items.

Factors

Factor 1		**Factor 2**	
Item	***Type***	***Item***	***Type***
Confusing – Clearly Structured	PQ	Dull – Captivating	HQS
Practical – Impractical	PQ	Creative – Unimaginative	HQS
Stylish – Tacky	*HQI*	*Ugly – Beautiful*	*OA*
Cheap – Premium	HQI		
Bad – Good	OA		
Unpredictable – Predictable	PQ		
Simple – Complicated	PQ		

Figure 4: Table showing loading of items onto the two extracted components, together with their expected type.

Note: Items in *italics* are cross-loaded.

Key: PQ = Pragmatic Quality, HQI = Hedonic Quality Identification, HQS = Hedonic Quality Stimulation, OA = Overall Appeal

5.6 Discussion

The factor analysis has produced some unexpected and interesting results. After performing principle components analysis with direct oblimin rotation, all of the items expected to be pragmatic components were clearly loaded onto factor 1. However, the same cannot be said for the items that were expected to be hedonic components. The hedonic identification item, Stylish – Tacky was quite evenly loaded onto both factors, although slightly more so onto factor 1. While the other hedonic identification item, Cheap – Premium loaded even more highly on factor 1. This is in contrast to both the hedonic stimulation items, Dull – Captivating and Creative – Unimaginative, which both loaded highly on factor 2. The two overall appeal items loaded as expected. Bad – Good was quite evenly loaded onto both factors with a slightly higher loading on factor 1, and Ugly – Beautiful loaded onto factor 2.

Despite the cross-loadings of two items, it should be noted that the two factors discovered are clearly defined in terms of their independence as there is negligible general correlation between factor 1 and 2 ($r = .135$).

A previous study by Plumley (2006) which employed a similar paradigm in order to validate the AttrakDiff Lite questionnaire, had found that “all items on the questionnaire behaved as they were expected to” (p. 22). This meant that all items believed to be pragmatic components loaded onto one factor, while all items believed to be hedonic components loaded onto

another. It is important to note however, that Plumley performed a subjects analysis, where as this study used materials analysis.

These differences raise more questions. For example, Plumley (2006) found that the overall appeal item, Bad – Good loaded more heavily on the hedonic factor and argued that this showed that the judgements of goodness before interaction were based mainly on hedonic qualities. However, this study found that Bad – Good was quite evenly loaded onto both factors with a slightly higher loading on factor 1, which contains mostly pragmatic components. This finding is similar to that of Hassenzahl (2004, p.330) who states that "goodness seems to be a consequence of the presence of pragmatic and hedonic attributes".

Although it seems strange that both hedonic identification items would load onto the same factor as all of the pragmatic items, there is some evidence to suggest that this is more than just an anomalous occurrence.

The items Stylish – Tacky and Cheap – Premium seem to rate very similar attributes based on their wording, they are also highly correlated ($r = -.924$, $p = .000$). The negative correlation comes from the fact that their positive and negative anchors are at opposite ends. See Appendix 1 for a full table of correlations.

Another interesting point to note is that both hedonic identification items, Stylish – Tacky and Cheap – Premium are highly correlated with the overall appeal item Ugly – Beautiful ($r = -.947$ and $r = .898$ respectively). This relates to the findings of Hassenzahl (2004), who states that "beauty solely depends on the product's apparent ability to communicate identity" (p. 330). Furthermore, both Stylish – Tacky and Cheap – Premium are also highly correlated with the overall appeal item Bad – Good ($r = -.917$ and $r = .943$ respectively). Again, Hassenzahl (2004) also found hedonic quality identification was significantly correlated with goodness.

The results also indicate that the items Stylish – Tacky and Ugly – Beautiful have a complex structure because of their cross-loadings.

Although it is not possible to fully explain the reason for the factor analysis result without further testing, there are a number of possible explanations.

The differences between the factor loadings in this study and that of Plumley (2006) are most-likely a result of the fact that Plumley's study used subjects analysis, whereas this study used materials analysis. Another difference in Plumley's study is that the web pages used were all related in terms of category, as they were all taken from travel websites. This study however, used web pages covering a variety of different topics including: design, e-commerce, news and travel. It is possible that certain website categories have a style that is predisposed to communicating visual attributes after a very short exposure time. Further analysis of the results for individual classes of websites would need to be undertaken to determine if this was a factor.

It may be that some of the questions are not working in the way that they should. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA) indicates that this is not the case because the values for the individual variables range from .648 to .858 and the overall MSA for the set of variables is .789 (Hair, Black, Babin, Anderson, & Tatham, 2005).

Perhaps 50 milliseconds was not enough time for participants to make accurate ratings across each of the questionnaire items for the web pages. However, this seems unlikely given previous findings (Plumley, 2006) and the high correlations between similar questionnaire items.

Another possibility is that the exposure time of 50 milliseconds was not enough for participants to make *individual* judgements about all of the items on the AttrakDiff Lite questionnaire. Instead the web page images left each participant with an overall impression, which they then relayed when answering the questionnaire. Evidence for this comes from the fact that questionnaire items which were intended to be measuring the same factors are highly correlated, even though the components did not split across factors in the way that was expected.

It may be that the two hedonic identification items, Stylish – Tacky and Cheap – Premium need further development within the AttrakDiff Lite questionnaire. However, a study by Lavie & Tractinsky (2004) defined aesthetics as having two different dimensions, which they referred to as "classical aesthetics" and "expressive aesthetics". Classical aesthetics emphasises order and clear design, while expressive aesthetics comes from a designer's creativity and originality. The results from their study showed that the "classical aesthetics dimension is more closely related to perceived usability relative to the expressive aesthetics dimension" (p. 292). By comparing the findings of Lavie & Tractinsky (2004) with those of this study, it could be argued that the two hedonic stimulation items are measuring expressive aesthetics, whereas the two hedonic identification items may be measuring classical aesthetics. This could explain why the hedonic identification components in this study are more highly correlated with some of the pragmatic components than they are with the hedonic stimulation components.

Further testing would need to be carried out in order to determine exactly why the factor analysis results are this way. A study using the same paradigm, but with a larger sample size may help. In addition, further studies with varying exposure times to the stimuli could also be beneficial to understanding the results.

The second study attempted to explore the results of the factor analysis further by comparing factor scores with the data from a heuristic analysis of visual design.

6. Study 2 – Developing Heuristics for Visual Design

6.1 Background

Despite the fact that non-utilitarian aspects of design have been gaining more attention from HCI researchers and practitioners, there are few analytical processes or tools for assessing web page aesthetics and none that have been widely accepted.

Design of information systems, particularly the user interface is often compared to architectural design (Tractinsky et al., 2000) which is one of the greatest showcases of aesthetics. Buildings that are purely functional and lack aesthetic considerations are generally viewed as soulless, dull and a detriment to their surroundings. Yet, despite this comparison there is still an uneasy relationship between form and function in HCI. This relationship needs to be addressed, as Tractinsky et al. (2000) suggest, people are introduced to a building by its

facade, and they also experience the “facade” (p. 140) of an interface first. It is what presents the first visual indications of what a user can expect from a system.

Guidelines exist for usability (e.g. “Top Ten Guidelines for Homepage Usability,” 2002) and accessibility (“Web Content Accessibility Guidelines 1.0,” 1999), therefore it would seem the next logical step to provide HCI related guidelines for visual design and other non-utilitarian aspects. Of course preferences for non-utilitarian aspects are arguably much more subjective, but this does not mean it is not worth developing guidelines to aid HCI practitioners in analysing these elements.

It is clear from emerging literature that aesthetics can affect perceptions of usability, although the exact reasons as to how and why are still in dispute. As a result, it would be helpful for HCI practitioners to be able to assess visual design, even if it is at a basic level, in order to combine this assessment with their knowledge of usability. This can ensure that visual design elements are working to enhance, rather than hinder usability. Another compelling factor for taking a greater interest in visual design is discussed by Hassenzahl (2001), who found that incorporating hedonic quality into usability assessments fostered a better relationship with clients’ marketing departments. It may also help to improve the relationship between usability engineers and graphic designers.

As discussed in Section 2.5, there have been some attempts to develop means of measuring website aesthetics, but most of these have been focused on understanding the effect of aesthetics on usability, and have not sought to aid HCI practitioners with visual design. Sutcliffe (2001, 2002) developed two sets of heuristics, which aimed to address the lack of guidance on aesthetic design for HCI practitioners. The first set of heuristics covered general aesthetic design while the second set covered content. While Sutcliffe’s efforts are to be commended in identifying and attempting to address a need in the HCI community, the heuristics themselves lack impact, focus and detail. There are only 7 heuristics related to aesthetic design, and because there are so few, they are disjointed and miss out several key visual design principles. As a result, the heuristics are of limited usefulness and have only been used in studies where Sutcliff himself has been involved (De Angeli et al., 2006; Hartmann et al., 2007; Sutcliffe, 2001, 2002).

Study 2 pursued similar goals to Sutcliffe (2001, 2002), which lead to the development of visual design heuristics that are based on the principles in “Designing Visual Interfaces” by Mullet and Sano (1995). The main priority for the heuristics was for them to be cohesive and understandable by HCI practitioners with little existing visual design knowledge. Given the subjective nature of aesthetics the aim was not to create a rigid formula for visual design of web pages, instead the heuristics are intended to be used as a rapid formative or summative evaluation of visual design and as a supplement to usability assessments.

Due to the nature of visual design, the heuristics that were developed still rely somewhat on the practitioner’s personal judgement, although the same can be said for usability heuristics.

The nature and scope of this study means that the visual design heuristics that were developed are not presented as a definitive answer, but rather as exploratory guidelines that also provide groundwork for future development and refinement.

6.2 Developing the heuristics

The heuristics were developed by first assessing the 22 visual design principles in "Designing Visual Interfaces" by Mullet and Sano (1995). The principles were created to help anyone "to enhance the visual quality of graphical user interfaces" (p. xi), by using techniques that are often used by practitioners of communication-oriented visual design. Mullet and Sano developed their principles by drawing on aesthetics approaches in graphic design, industrial design and architecture.

Although the principles are described as being "truly generic" (Mullet & Sano, 1995, p. xiii), they were written before widespread adoption of the Internet and thus do not take websites into account. Therefore the first step was to assess which principles could be applied to visual design of websites.

18 out of the 22 visual design principles were chosen as a result of this first analysis. These 18 principles were then used to develop the heuristics. This was done by first researching existing usability heuristics, including Nielsen's Heuristics for User Interface Design (Nielsen, 2005) and the heuristics found in ISO 9241-110 ("ISO 9241: Part 110," 2007). In addition to this, Sutcliffe's (2001) Attractiveness Heuristics and the Apple Human Interface Guidelines (2007) were also reviewed. The 18 visual design principles were then reformulated and written out in the style of heuristic statements, similar to those in the usability heuristics.

After the first draft of the heuristics was complete, they underwent review by 3 HCI practitioners who also had experience with visual design. As a result of this review, one of the heuristics was removed because it was deemed too similar to another. Also, some of the wording was revised and examples were added to make the heuristics easier to understand and use. The first draft of the heuristics can be found in Appendix 2.

The expert review resulted in a final set of 17 visual design heuristics, which can be found in Appendix 3.

6.3 Pilot

After the visual design principles had been developed, a pilot study was carried out to confirm their validity and inter-rater reliability.

12 participants took part in the pilot, 5 men and 7 women, all of whom were HCI practitioners.

The participants completed a web-based questionnaire, which consisted of the 17 heuristics, each accompanied with a 7-point Likert scale, with bi-polar verbal anchors ranging from "Extremely Bad" to "Extremely Good". The participants were asked to use the visual design heuristics to rate 5 of the web pages from the first study. The 5 web pages used varied in their usability, aesthetics and style.

The data gathered from the pilot was then used to generate statistics. The range and interquartile range (IQR) of the ratings given by participants was analysed to determine whether the participants had made similar judgements using the heuristics.

Ranges of Ratings: Websites

Web page	Average Range	Average IQR	Heuristic Score
brown.edu	4	2	76.84
direct.gov.uk	4	2	75.24
guardian.co.uk	2	1	96.4
wetestit.com	3	2	36.59
yellowstonepark.com	3	1	90.76

Figure 5: Average ranges of web page ratings and heuristic scores from the pilot heuristic analysis

Figure 5 shows the average range and interquartile range of responses across each of the web pages. It also shows the heuristic score, which is the sum of the average rating for each of the 17 heuristics.

The results show that the web pages, guardian.co.uk and yellowstone.com received the highest aesthetics ratings, with wetestit.com receiving the lowest rating. This result was in line with expectations.

There was a differing amount of agreement between participant's ratings across each of the web pages, although the interquartile range shows that overall they were very similar.

It was the two most aesthetic web pages, guardian.co.uk and yellowstone.com that saw the highest agreement between participants' ratings, with the least aesthetic web page receiving the next highest agreement of ratings. The two web pages with middle ratings had a wider range of responses, but the interquartile range still showed a good amount of agreement.

This demonstrates that web pages which provoked a stronger positive or negative feeling in participants received the most similar ratings. Ritterfeld (2002, as cited in Tractinsky et al., 2006) states that negative influences are stronger during the decision making process, but although the least aesthetic web page was clearly agreed upon by participants, the highly aesthetic web pages received more similar ratings.

Figure 6 (on page 25) shows the average range and interquartile range of the average ratings for each of the 17 heuristics, across all five web pages used in the pilot.

The results show there is a varying amount of agreement between participant's use of the 17 heuristics, however much like with the ratings of web pages, the interquartile range shows that overall they were very similar.

Ranges of Ratings: Heuristics

Heuristic	Average Range	Average IQR
Elegance and Simplicity		
Unity	2	1
Refinement	4	2
Fitness	3	2
Scale, Contrast, Proportion		
Clarity	3	2
Harmony	3	2
Activity	3	2
Restraint	3	2
Organisation and Visual Structure		
Grouping	4	2
Hierarchy	4	2
Relationship	3	2
Balance	4	2
Module and Program		
Focus	4	2
Flexibility	4	2
Consistent Application	2	1
Style		
Distinctiveness	4	2
Integrity	3	1
Appropriateness	3	2

Figure 6: Average ranges of heuristic scores from the pilot heuristic analysis

6.4 Heuristic Analysis

The pilot study had shown that the heuristics were generating valid judgements and the inter-rater reliability was at an acceptable level. As a result, the heuristics were then used to analyse all 30 of the web pages used in the first study.

2 HCI practitioners carried out the heuristic analysis, using the questionnaire designed during the pilot study.

This analysis was carried out in order to compare the heuristic scores, with factor scores based on data from the first study. This was done to further develop the heuristics, but also to try and learn more about the results of the factor analysis from the first study.

6.5 Results

In order to compare the factor analysis results from the first study with the heuristic analysis conducted in the second study, scores for factors and heuristics were calculated for each of the 30 web pages.

The factor scores were generated by adding up the average values for each of the questions that made up a factor. Whenever a question was negatively phrased, the value was reversed by subtracting it from eight. The heuristic score was generated by calculating the sum of the average rating for each of the 17 heuristics. (See Appendix 4 for the table of factor scores alongside the heuristic scores).

After the factor and heuristic scores had been calculated, a correlational analysis was performed on the data. Figures 7, 8 and 9 (on pages 26 and 27) show the results that were obtained.

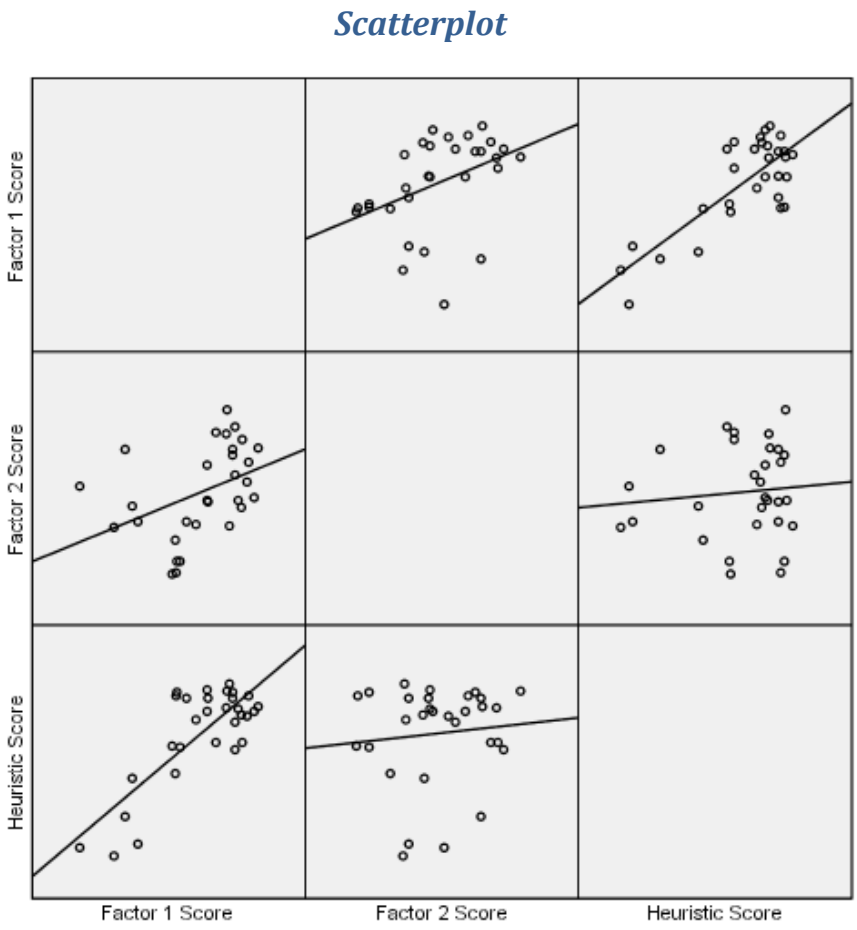

Figure 7: Scatterplot showing the relationship between factor scores and heuristic scores

Correlations

		Factor 1 Score	Factor 2 Score	Heuristic Score
Factor 1 Score	Pearson Correlation	1	.416(*)	.788(**)
	Sig. (2-tailed)		.022	.000
	N	30	30	30
Factor 2 Score	Pearson Correlation	.416(*)	1	.103
	Sig. (2-tailed)	.022		.589
	N	30	30	30
Heuristic Score	Pearson Correlation	.788(**)	.103	1
	Sig. (2-tailed)	.000	.589	
	N	30	30	30

Figure 8: Correlations of factor and heuristic scores

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Figure 8 shows that factor 1 scores are highly correlated with the heuristic scores, and it also shows that factor 1 scores are moderately correlated with factor 2 scores. The correlation between factor 2 scores and the heuristic scores is negligible.

Nonparametric Correlations

			Factor 1 Score	Factor 2 Score	Heuristic Score
Spearman's rho	Factor 1 Score	Correlation Coefficient	1.000	.511(**)	.447(*)
		Sig. (2-tailed)	.	.004	.013
		N	30	30	30
	Factor 2 Score	Correlation Coefficient	.511(**)	1.000	.085
		Sig. (2-tailed)	.004	.	.655
		N	30	30	30
	Heuristic Score	Correlation Coefficient	.447(*)	.085	1.000
		Sig. (2-tailed)	.013	.655	.
		N	30	30	30

Figure 9: Nonparametric correlations of factor and heuristic scores

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

The nonparametric correlations in figure 9 show that factor 1 scores are moderately correlated with factor 2 scores, and it also shows that factor 1 scores are moderately correlated with the heuristic scores. The correlation between factor 2 scores and the heuristic scores is negligible.

6.6 Discussion

The correlational analysis revealed a high correlation between factor 1 scores and the heuristic scores ($r = .788$, $p = .000$). However, the scatterplots in figure 7 display a fair degree of skew, which suggests that perhaps the more conservative nonparametric correlation ($r = .447$) is a more accurate measure.

Both sets of analysis show moderate correlations between the factor 1 and factor 2 scores. As expected, this indicates that they are both measuring aesthetics, albeit different aspects.

The moderate to high correlation between factor 1 scores and heuristic scores is particularly interesting. Participants in the first study were exposed to the stimuli for just 50 milliseconds, whereas the heuristic analysis in the second study had no restrictions on exposure time. Participants could observe the web page for as long as they wanted (although they were still unable to interact with it). Despite these differences, many of the web pages that had high factor 1 scores, also had high heuristic scores and those with low factor 1 scores had low heuristic scores (see Appendix 4). This implies that the participants who answered the questionnaire with only 50 ms exposure to stimuli made similar judgements to the participants who performed the heuristic analysis, at least with the questions that made-up factor 1. Although the results are not directly comparable in terms of their magnitude, this does show that participants in the first study where getting a positive or negative impression within 50ms, which was consistent with the impression of the participants performing the heuristic analysis. This supports the findings of Lindgaard et al. (2006) which state that visual appeal can be assessed in as little as 50 ms. Furthermore, discussion with participants who had completed the AttrakDiff Lite questionnaire in the first study, revealed that they were able to recognise colours and prominent images on the web pages.

While the heuristic scores were moderately correlated with the factor 1 scores, the correlation between factor 2 and the heuristic score is negligible. This suggests that the heuristics are measuring order and clear design, which are associated with factor 1, rather than novelty, challenge and creativity which are measured by factor 2. This also means that the heuristics are measuring what Lavie & Tractinsky (2004) would refer to as “classical aesthetics” which are more closely tied to traditional usability.

Another interesting point that has emerged from the data relates to the comparative ranking of the web pages used in the study. Without a set exposure time, the heuristic analysis generated a steady scale of web pages, ranked from most aesthetic to least aesthetic. As participants were able to progress through the analysis at their own discretion, they were actively comparing the visual qualities of the web pages, which is reflected in the heuristic score. The factor scores generated from the AttrakDiff Lite questionnaire do not provide the same fixed rankings. Having a 50 ms exposure time meant that participants were not comparing web pages, but rather providing individual ratings for each page. This measure was necessary in order to avoid biasing the questionnaire responses and subsequent factor analysis, but it does make the factor scores harder to compare. Furthermore, as factor 1 and 2

are independent, a web page with a high factor 1 score will not necessarily have a high factor 2 score and any comparison of scores must take this into account.

These results show that both the AttrakDiff Lite questionnaire and the visual design heuristics are both performing well in their measurement of aesthetics. The extent of their agreement is a promising indicator of their success.

7. Limitations of the Present Study

This study was investigating a relatively new area in the field of HCI and as such was very exploratory in nature. As a result, there were some limitations.

The first part of this study assumed that 50 milliseconds was enough time to make sound judgements about web pages based on previous research (Lindgaard, et al., 2006) and more specifically make these judgements consistently using the AttrakDiff Lite questionnaire (Plumley, 2006). Future research into the AttrakDiff Lite questionnaire using a materials analysis approach should look into how different exposure times affect responses to stimuli.

To ensure that participants' responses to the AttrakDiff Lite questionnaire were based on aesthetics alone, only screen-shots of single web pages were assessed. In order to compare results from the first and second studies, the same web page stimuli had to be used for the heuristic analysis. This made some of the heuristics difficult to apply, as they are intended to be used across whole websites. Furthermore, the questionnaire had to limit participant's responses to a numerical scale in order to facilitate the analysis of inter-rater reliability. Ideally the heuristic analysis responses should not be limited to numerical values.

These limitations were necessary in order to ensure that there were no confounding factors and keep the study at a manageable scale.

It was also not possible to investigate the affect of interaction on participants' responses using the AttrakDiff Lite questionnaire or the heuristic analysis. To do so would have made it impossible to judge whether the responses made were a result of aesthetics or a combination of aesthetics and usability. Although it was not possible in this study, future research should look into how interaction can affect the measurement of aesthetics.

8. Conclusions

These two studies were conducted to explore different but complimentary areas of visual design in relation to HCI, to further the recent expanse of research into non-utilitarian design. Individually the studies have provided insights into how aesthetics can be measured and applied to websites.

The first study used a materials analysis approach to validate the AttrakDiff Lite questionnaire and explore the nature of immediate affective reactions, building upon prior research by Plumley (2006), who had used subjects analysis. A factor analysis of questionnaire responses revealed there were two factors that explained 86.7% of the variance. However, the questionnaire items did not all load as expected based on Plumley's findings. While Plumley found that that all pragmatic items loaded onto one factor and the hedonic items onto another factor, this study found that there was not a clear split. Factor 1 contained pragmatic

components, hedonic identification components and goodness, while factor 2 contained hedonic stimulation components and beauty.

These differences in findings may be a result of the fact that Plumley's (2006) study used subjects analysis, whereas this study used materials analysis. The high correlations between the two hedonic identification components indicate that this is not an anomalous result and that they are measuring very similar attributes, which accounts for them being loaded on the same factor. It may be that the two hedonic identification items need further development, or it is possible that they could be measuring "classical aesthetics" (Lavie & Tractinsky, 2004), which is closely linked to usability. This would explain the correlation of the hedonic identification items with pragmatic items. In order to try and gain a full understanding of these results, further testing is required. Future research involving the AttrakDiff Lite questionnaire should consider exploring how questionnaire answers are affected by varying exposure times to stimuli.

In the second study, 17 visual design heuristics were created based on the principles in "Designing Visual Interfaces" by Mullet and Sano (1995). These were developed in response to a perceived lack of guidelines for HCI practitioners that could help them to understand visual design in a user experience context. The heuristics underwent expert review and were validated using a pilot study, after which they were used to perform a heuristic analysis on the web pages from the first study.

Heuristic scores were then compared to scores based on the factors from the first study. The heuristic scores were moderately correlated with the factor 1 scores, but the correlation between factor 2 and the heuristic score is negligible. This suggests that the heuristics are measuring order and clear design, which are associated with factor 1, rather than novelty, challenge and creativity which are measured by factor 2. Furthermore, the correlation between the heuristic scores and factor 1 scores indicates that participants who answered the AttrakDiff Lite questionnaire with only 50 ms exposure to stimuli, made similar judgements to the participants who performed the heuristic analysis, which had no fixed stimuli exposure time. This suggests that visual appeal can be assessed within 50 ms, which mirrors the findings of Lindgaard et al. (2006).

Although the 17 visual design heuristics provided promising results, they need further development. Within the scope of this study, they were only used to assess individual web pages and responses were restricted to a 7-point Likert scale. Although it was necessary for this study, the visual design heuristic analysis should not be limited to numerical responses, as they do not provide enough information for improving user experience. Future work on the heuristics should involve testing them on full websites and exploring the affect that interaction may have on analysis.

The studies that were undertaken offer a comprehensive look at how website aesthetics can be assessed by undertaking materials analysis using the AttrakDiff Lite questionnaire, and through the development of visual design heuristics. These are essential steps in helping to increase the understanding of how aesthetics can affect user experience.

9. References

Apple Human Interface Guidelines. (2007). Retrieved August 4, 2007, from http://developer.apple.com/documentation/UserExperience/Conceptual/OSXHIGuidelines/index.html#//apple_ref/doc/uid/20000957

Basso, A., Goldberg, D., Steven Greenspan, & Weimer, D. (2001). First impressions: emotional and cognitive factors underlying judgments of trust e-commerce. 3rd ACM conference on Electronic Commerce, 137 - 143. Retrieved August 1, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/510000/501173/p137-basso.pdf?key1=501173&key2=8683895811&coll=Portal&dl=ACM&CFID=25592133&CFTOKEN=80762907

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, 13(2), 210 - 234. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1170000/1165737/p210-ben-bassat.pdf?key1=1165737&key2=4032561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

Blythe, M. A., Overbeeke, K., Monk, A. F., & Wright, P. C. (Eds.). (2004). Funology: From Usability to Enjoyment. Kluwer Academic Publishers.

De Angeli, A., Sutcliffe, A., & Hartmann, J. (2006). Interaction, usability and aesthetics: what influences users' preferences? 6th ACM conference on Designing Interactive systems, 271 - 280. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1150000/1142446/p271-de-angeli.pdf?key1=1142446&key2=6992561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

First Impressions Count in Website Design - visual appeal, beauty and aesthetics, halo effect, cognitive perception, webpage judgments of credibility. (2007). WebSiteOptimization.com. Retrieved June 12, 2007, from http://www.websiteoptimization.com/speed/tweak/blink/

Fogg, B., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., Tauber, E. R., et al. (2003). How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. Designing For User Experiences, 1 - 15. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1000000/997097/p1-fogg.pdf?key1=997097&key2=1082561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

Hair, J., Black, B., Babin, B., Anderson, R., & Tatham, R. (2005). Multivariate Data Analysis. Prentice Hall.

Hartmann, J. (2006). Assessing the attractiveness of interactive systems. CHI '06 extended abstracts on Human factors in computing systems, 1755 - 1758. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1130000/1125781/p1755-hartmann.pdf?key1=1125781&key2=6012561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

Hartmann, J., Sutcliffe, A., & De Angeli, A. (2007). Investigating attractiveness in web user interfaces. SIGCHI conference on Human factors in computing systems, 387 - 396. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1250000/1240687/p387-hartmann.pdf?key1=1240687&key2=8723561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

Hartson, H. R. (1998). Human–computer interaction: Interdisciplinary roots and trends. The Journal of Systems and Software, (43), 103-118. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6V0N-3V3Y5VX-4-B&_cdi=5651&_user=125795&_orig=search&_coverDate=11%2F30%2F1998&_sk=999569997&view=c&wchp=dGLbVlb-zSkWb&md5=c9d843d535939a470813ad49b1acc30e&ie=/sdarticle.pdf

Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. Retrieved June 12, 2007, from http://www.tu-darmstadt.de/fb/fb3/psy/soz/veroeffentlichungen_mh/vt_hassenzahl_final.pdf

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness and Usability in Interactive Products. Human-Computer Interaction, 19, 319-349.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality] (G. Szwillus & J. Ziegler, Eds.). Mensch & Computer 2003: Interaktion in Bewegung, 187 - 196. Retrieved August 16, 2007, from http://mc.informatik.uni-hamburg.de/konferenzbaende/mc2003/konferenzband/muc2003-18-hassenzahl.pdf

ISO 9241: Part 11. (2007). Retrieved August 3, 2007, from http://www.userfocus.co.uk/resources/iso9241/part11.html

ISO 9241: Part 110. (2007). Retrieved August 20, 2007, from http://www.userfocus.co.uk/resources/iso9241/part110.html

Java Look and Feel Design Guidelines. (2001). Retrieved August 4, 2007, from http://java.sun.com/products/jlf/ed2/book/index.html

Karvonen, K. (2000). The Beauty of Simplicity. 2000 conference on Universal Usability, 85 - 90. Retrieved July 22, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/360000/355478/p85-karvonen.pdf?key1=355478&key2=3977215811&coll=Portal&dl=ACM&CFID=29266786&CFTOKEN=22790792

Kurosu, M., & Kashimura, K. (1995a). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. Human Factors in Computing Systems, 292 - 293. Retrieved July 22, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/230000/223680/p292-

kurosu.pdf?key1=223680&key2=7801115811&coll=Portal&dl=ACM&CFID=29248196&CFTOKEN=93983763

Kurosu, M., & Kashimura, K. (1995b). Determinants of the apparent usability. Systems, Man and Cybernetics, 1995. 'Intelligent Systems for the 21st Century', 1509 - 1514. Retrieved July 22, 2007, from http://ieeexplore.ieee.org.libproxy.ucl.ac.uk/iel3/4021/11540/00537986.pdf?tp=&arnumber=537986&isnumber=11540

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. Human-Computer Studies, (60), 269-298. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6WGR-4B4XRTB-1-7&_cdi=6829&_user=125795&_orig=search&_coverDate=03%2F31%2F2004&_sk=999399996&view=c&wchp=dGLbVtz-zSkWW&md5=c0ca76137d5b88dc387ee56b23f4c658&ie=/sdarticle.pdf

Lawrence, D., & Tavakol, S. (2006). Balanced Website Design: Optimising Aesthetics, Usability and Purpose. Springer-Verlag London Ltd.

Lindgaard, G. (2006). Aesthetics, visual appeal, usability, and user satisfaction: What do the user's eyes tell the user's brain? Retrieved July 22, 2007, from http://hot.carleton.ca/hot-topics/articles/aesthetics-visual-appeal/

Lindgaard, G., & Dudek, C. (2003). What is this evasive beast we call user satisfaction? Interacting with Computers, (15), 429–452. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6V0D-48V6JNS-2-S&_cdi=5644&_user=125795&_orig=search&_coverDate=06%2F30%2F2003&_sk=999849996&view=c&wchp=dGLbVlz-zSkWz&md5=c0b9a1eadcc2244d4a0142d63b29c793&ie=/sdarticle.pdf

Lindgaard, G., Fernandes, G., Dudek, C., & Brown, J. (2006). Attention web designers: You have 50 milliseconds to make a good first impression! Behaviour & Information Technology, 25(2), 115-126.

Mahlke, S. (2002). Factors influencing the experience of website usage. CHI '02 extended abstracts on Human factors in computing systems, 846 - 847. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/510000/506628/p846-mahlke.pdf?key1=506628&key2=1162561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

Mullet, K., & Sano, D. (1995). Designing Visual Interfaces. Prentice Hall.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. SIGCHI conference on Human factors in computing systems: celebrating interdependence, 152-158. Retrieved July 19, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/200000/191729/p152-nielsen.pdf?key1=191729&key2=8756684811&coll=Portal&dl=ACM&CFID=28941838&CFTOKEN=22454386

Nielsen, J. (2005). Heuristics for User Interface Design. Retrieved August 20, 2007, from http://www.useit.com/papers/heuristic/heuristic_list.html

Norman, D. A. (1998). The Design of Everyday Things. Basic Books Inc., New York.

Norman, D. A. (2004a). Beauty, Goodness, and Usability. Human-Computer Interaction, 19(4). Retrieved June 12, 2007, from http://hci-journal.com/editorial/si-beauty-intro.pdf

Norman, D. A. (2004b). Emotional Design: Why we love (or hate) everday things. Basic Books Inc., New York.

Pandir, M., & Knight, J. (2006). Homepage aesthetics: The search for preference factors and the challenges of subjectivity. Interacting with Computers, (18), 1351–1370. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6V0D-4K428JR-2-1&_cdi=5644&_user=125795&_orig=search&_coverDate=12%2F31%2F2006&_sk=999819993&view=c&wchp=dGLbVzb-zSkWA&md5=a5eb41b72b0204f142c57cfc1a5ed122&ie=/sdarticle.pdf

Park, S., Harada, A., & Igarashi, H. (2006). Influences of personal preference on product usability. CHI '06 extended abstracts on Human factors in computing systems, 87 - 92. Retrieved July 22, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1130000/1125475/p87-park.pdf?key1=1125475&key2=7816115811&coll=Portal&dl=ACM&CFID=29253829&CFTOKEN=47469949

Plumley, E. (2006). The Effect of Affect: Exploring the Association of Aesthetics and Usability in Perceptions of Web Pages. University College London.

Schenkman, B., & Joènsson, F. U. (2000). Aesthetics and preferences of web pages. Behaviour & Information Technology, 19(5), 367-377. Retrieved June 12, 2007, from http://www.idemployee.id.tue.nl/g.w.m.rauterberg/BIT-25/schenkman-jonsson-2000.pdf

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), 1055–1069. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6V0D-4JHMHRJ-1-5&_cdi=5644&_user=125795&_orig=search&_coverDate=09%2F30%2F2006&_sk=999819994&view=c&wchp=dGLbVlz-zSkWA&md5=146bb63304d3093efcb1fc792da2d873&ie=/sdarticle.pdf

Shneiderman, B. (2004). Designing for fun: how can we design user interfaces to be more fun? Interactions, 11(5), 48-50. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/1020000/1015552/p48-shneiderman.pdf?key1=1015552&key2=4010661811&coll=Portal&dl=GUIDE&CFID=21154900&CFTOKEN=93928009

Solomon, N. W. (2004). The Hindu : Measuring aesthetics. Retrieved August 16, 2007, from http://www.hindu.com/edu/2004/02/23/stories/2004022300160400.htm

Sutcliffe, A. (2001). Heuristic Evaluation of Website Attractiveness and Usability. , 2220/2001. Retrieved July 17, 2007, from http://springerlink.metapress.com.libproxy.ucl.ac.uk/content/hpbfalklyku4bwnt/fulltext.pdf

Sutcliffe, A. (2002). Assessing the Reliability of Heuristic Evaluation for Website Attractiveness and Usability. 35th Hawaii International Conference on System Sciences, 1838 - 1847. Retrieved June 12, 2007, from http://ieeexplore.ieee.org.libproxy.ucl.ac.uk/iel5/7798/21442/00994098.pdf?tp=&arnumber=994098&isnumber=21442

Top Ten Guidelines for Homepage Usability. (2002). Retrieved September 1, 2007, from http://www.useit.com/alertbox/20020512.html

Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. SIGCHI conference on Human factors in computing systems, 115 - 122. Retrieved June 12, 2007, from http://delivery.acm.org.libproxy.ucl.ac.uk/10.1145/260000/258626/p115-tractinsky.pdf?key1=258626&key2=0291561811&coll=Portal&dl=GUIDE&CFID=21140796&CFTOKEN=49619173

Tractinsky, N. (2004). Toward the Study of Aesthetics in Information Technology. , 771-780. Retrieved June 12, 2007, from http://www.ise.bgu.ac.il/faculty/noam/papers/04_nt_icis.pdf

Tractinsky, N., Cokhavi, A., & Kirschenbaum, M. (2004). Using Ratings and Response Latencies to Evaluate the Consistency of Immediate Aesthetic Perceptions of Web Pages. Third Annual Workshop on HCI Research. Retrieved June 12, 2007, from http://www.ise.bgu.ac.il/faculty/noam/papers/04_ntacmk_final.pdf

Tractinsky, N., Cokhavi, A., Kirschenbaum, M., & Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. Human-Computer Studies, (64), 1071–1083. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6WGR-4KPX929-1-K&_cdi=6829&_user=125795&_orig=search&_coverDate=11%2F30%2F2006&_sk=999359988&view=c&wchp=dGLbVlb-zSkzS&md5=f428446197f7b3bcc755054d7a862a57&ie=/sdarticle.pdf

Tractinsky, N., Katz, A., & Ikar, D. (2000). What is beautiful is usable. Interacting with Computers, (13), 127-145. Retrieved June 12, 2007, from http://www.sciencedirect.com.libproxy.ucl.ac.uk/science?_ob=MImg&_imagekey=B6V0D-41TMVFN-2-1C&_cdi=5644&_user=125795&_orig=search&_coverDate=12%2F31%2F2000&_sk=999869997&view=c&wchp=dGLbVlz-zSkzS&md5=73b05432cbe8bbd2e58b29c4b187f954&ie=/sdarticle.pdf

Udsen, L. E., & Jørgensen, A. H. (2005). The aesthetic turn: unravelling recent aesthetic approaches to human-computer interaction - Digital Creativity. Retrieved June 12, 2007, from http://www.informaworld.com.libproxy.ucl.ac.uk/smpp/section?content=a743863866&fulltext=713240928

User Interface Design GmbH. (2007). AttrakDiff Fragebogen, User Interface Design GmbH. Retrieved August 16, 2007, from http://www.attrakdiff.de/

Web Content Accessibility Guidelines 1.0. (1999). Retrieved September 1, 2007, from http://www.w3.org/TR/WCAG10/

Web Pages That Suck: Learn usability and good Web design by looking at bad Web design. (2007). Retrieved August 1, 2007, from http://www.webpagesthatsuck.com/

Welcome to the Webby Awards. (2007). Retrieved July 17, 2007, from http://www.webbyawards.com/

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. Retrieved July 20, 2007, from http://www.blackwell-synergy.com.libproxy.ucl.ac.uk/doi/pdf/10.1111/1540-5885.1140309

Appendix 1: Correlations of AttrakDiff Lite Questions

	Confusing – Clearly Structured	Dull – Captivating	Practical – Impractical	Stylish – Tacky	Cheap – Premium	Bad – Good	Unpredictable – Predictable	Simple – Complicated	Ugly – Beautiful	Creative – Unimaginative
Confusing – Clearly Structured	1	-0.036	-0.775	-0.691	0.762	0.797	0.877	-0.563	0.685	0.072
Dull – Captivating	-0.036	1	0.323	-0.52	0.336	0.424	-0.224	-0.025	0.567	-0.902
Practical – Impractical	-0.775	0.323	1	0.369	-0.511	-0.52	-0.843	0.545	-0.319	-0.387
Stylish – Tacky	-0.691	-0.52	0.369	1	-0.924	-0.92	-0.511	0.418	-0.947	0.546
Cheap – Premium	0.762	0.336	-0.511	-0.924	1	0.943	0.665	-0.447	0.898	-0.383
Bad – Good	0.797	0.424	-0.522	-0.917	0.943	1	0.632	-0.409	0.924	-0.412
Unpredictable – Predictable	0.877	-0.224	-0.843	-0.511	0.665	0.632	1	-0.587	0.509	0.268
Simple – Complicated	-0.563	-0.025	0.545	0.418	-0.447	-0.41	-0.587	1	-0.419	-0.048
Ugly – Beautiful	0.685	0.567	-0.319	-0.947	0.898	0.924	0.509	-0.419	1	-0.585
Creative – Unimaginative	0.072	-0.902	-0.387	0.546	-0.383	-0.41	0.268	-0.048	-0.585	1

Appendix 2: Draft Visual Design Heuristics

Elegance and Simplicity

- ***Unity***
  The design components should all have a similar look and feel and work together towards a common purpose.

- ***Refinement***
  The design should not contain anything that is not essential to the communication task.

  Note: Removing visual cues can be disorienting, therefore removing or hiding elements should not be done at the expense of visual affordances.

- ***Fitness***
  The design should fit the task and the target audience.

Scale, Contrast, Proportion

- ***Clarity***
  The message or purpose of the design should be easy to perceive.

- ***Harmony***
  The relationship between all elements should be clearly ordered and fit together to form a cohesive design.

- ***Activity***
  Contrast between graphical elements should be used to help users maintain orientation and context.

- ***Restraint***
  The design should not contain unnecessary elements used simply for decoration or because the technology is available.

Organisation and Visual Structure

- ***Grouping***
  Similar elements should be grouped together contextually and effective use should be made of negative space.

  Note: Grouping elements into units helps users deal with large amounts of information more effectively.

- ***Hierarchy***
  Elements should be ordered in a hierarchy of perceptual prominence, corresponding to the intended reading sequence.

- ***Relationship***
  Visual elements should be positioned so that the relationship between each element is discernable with only a minimal use of labelling.

- ***Balance***
  Design elements should be balanced in either a symmetrical or asymmetrical layout, throughout the site.

  Note: Symmetrical layouts achieve balance automatically, whereas asymmetrical layouts require careful manipulation of design elements to compensate for differences in size, position and value.

Module and Program

- ***Focus***
  The design's structure should be predictable and explicit, by providing consistent sizing and positioning of user interface elements.

- ***Flexibility***
  The layout of the design should allow for flexibility and change without disrupting the harmony of the design.

- ***Consistent Application***
  Visual elements should be consistent throughout the design, in appearance, placement and meaning.

  Note: This makes it easier for users to interpret and respond to the interface.

Style

- ***Distinctiveness***
  In order to establish a visual and conceptual identity, a style must be easily recognisable throughout the design.

- ***Integrity***
  The design elements should form a style that is internally consistent and each of the elements should be well integrated to ensure a unified visual experience.

  Note: This means creating a Website that conforms to relevant standards and has a consistent appearance across multiple operating systems and browsers.

- ***Comprehensiveness***
  The style used must be applied throughout the design.

- ***Appropriateness***
  Visual language elements should be matched to the target audience, ensuring that they enhance rather than hinder communication with the user.

Appendix 3: Final Visual Design Heuristics

Visual Design Heuristics

Version 1.2
Based on "Designing Visual Interfaces" by Mullet and Sano (1995)

Elegance and Simplicity

- ***Unity***
  The design components should all have a similar look and feel and work together towards a common purpose.

  Example: The same visual language elements should be used throughout the Website.

- ***Refinement***
  The design should not contain anything that is not essential to the communication task.

  Note: Removing visual cues can be confusing, therefore removing or hiding elements should not be done at the expense of visual affordances.

  Example: Extraneous details which do not form an essential part of the Website, either functionally or aesthetically should be removed.

- ***Fitness***
  The design should fit the task and the target audience.

  Example: Design elements should relate to the main purpose of the Website and the expected visitors.

Scale, Contrast, Proportion

- ***Clarity***
  The message or purpose of the design should be easy to understand.

  Example: The purpose of the Website should be easy for a visitor to determine. In addition, interactive elements should be clearly distinguishable in both form and function.

- ***Harmony***
  The relationship between all elements should be clearly ordered and fit together to form a cohesive design.

  Example: Users should be able to see clearly how individual parts of the Website relate to each other.

- ***Activity***
  Contrast between graphical elements should be used to help users maintain orientation and context.

  Example: Contrasting colours can be used to make it easy for users to distinguish between different areas of focus.

- ***Restraint***
  The design should not contain unnecessary elements used simply for decoration or because the technology is available.

  Example: Using animation just because it can be done, has no real benefit to the design.

Organisation and Visual Structure

- ***Grouping***
  Similar elements should be grouped together contextually and effective use should be made of negative space.

  Note: Grouping elements into units helps users deal with large amounts of information more effectively.

  Example: High level structures help to familiarise the user with the interface and allow them to plan where they want to focus their attention.

- ***Hierarchy***
  Elements should be ordered in a hierarchy of perceived importance, corresponding to the intended reading sequence.

  Example: The most important elements must draw the user in first.

- ***Relationship***
  Visual elements should be positioned so that the relationship between each element is discernable with only a minimal use of labelling.

  Example: Aspects such as position, size and colour can be used to imply relationships between elements and the eye is particularly sensitive to alignment.

- ***Balance***
  Design elements should be balanced in either a symmetrical or asymmetrical layout, throughout the site.

  Note: Symmetrical layouts achieve balance automatically, whereas asymmetrical layouts require careful manipulation of design elements to compensate for differences in size, position and value.

  Example: If the visual weight of elements on each side of the centre axis is equal, then an impression of balance is guaranteed.

Module and Program

- ### Focus
  The design's structure should be predictable and explicit, by providing consistent sizing and positioning of user interface elements, so that points of focus are easy to determine.

  Example: The Website should have a rhythm and regularity, in order to assist the user to focus their visual attention.

- ### Flexibility
  The layout of the design should allow for flexibility and change without disrupting the harmony of the design.

  Example: The design template should allow for future changes and expansion, both in the design and potential content.

- ### Consistent Application
  Visual elements should be consistent throughout the design, in appearance, placement and meaning.

  Note: This makes it easier for users to interpret and respond to the interface.

  Example: Using the same visual elements throughout the Website will mean that they become familiar to the user even after minimal exposure.

Style

- ### Distinctiveness
  In order to establish a visual and conceptual identity, a style must be easily recognisable throughout the design.

  Example: A design which has visual impact can be easily recognised at a glance.

- ### Integrity
  The design elements should form a style that is internally consistent and each of the elements should be well integrated to ensure a unified visual experience.

  Note: This means creating a Website that conforms to relevant standards and has a consistent appearance across multiple operating systems and browsers.

- ### Appropriateness
  Visual language elements should be matched to the target audience, ensuring that they enhance rather than hinder communication with the user.

  Example: Children cannot make subtle perceptual distinctions required by a sophisticated design, while adults can become distracted and annoyed by visual design aimed at children.

Appendix 4: Factor and Heuristic Scores

Website	Factor 1 Score	Factor 2 Score	Heuristic Score
adobe.com	29.27	10.87	96.50
amazon.co.uk	31.53	12.13	86.50
apple.com/macosx/leopard	30.87	15.20	99.50
atomfilms.com	29.93	14.13	78.00
brown.edu	31.53	14.40	75.00
channel4.com	32.13	13.80	78.00
digg.com	26.67	8.07	99.00
direct.gov.uk	28.27	9.80	87.50
dokimos.org/ajff	18.53	11.60	34.00
ebuyer.com	27.47	9.93	96.50
expedia.co.uk	32.53	11.80	89.00
gamasutra.com	26.53	9.07	65.00
globalaigs.org	22.93	10.67	63.00
guardian.co.uk	32.67	12.73	97.50
kayak.co.uk	26.60	7.53	97.50
last.fm	29.20	10.93	100.00
longscycle.com	21.40	9.67	30.50
microsoft.com/windowsvista	31.33	13.33	96.50
mozilla.org	31.07	9.73	102.50
nationalgeographic.com	33.47	13.40	93.00
news.bbc.co.uk	32.07	10.60	89.50
newzealand.com/travel/UK-Ireland	31.33	13.07	99.00
nytimes.com	33.13	11.07	91.00
photography.si.edu	29.20	12.60	91.00
saltinstitute.org	22.33	13.33	47.00
uclic.ucl.ac.uk	31.80	10.93	92.00
useit.com	26.93	8.07	76.00
w3.org	26.27	7.47	76.50
wetestit.com	23.40	9.93	35.50
yellowstonepark.com	30.80	14.07	92.50