



VISUAL APPEAL and AFFECT in WEBSITES



A multi-method investigation into the relation
between visual appeal judgements of websites
and affect

Master Thesis by Gijs Huisman

Visual Appeal and Affect in Websites

A multi-method investigation into the relation between visual appeal judgements of websites
and affect

Master Thesis by

Gijs Huisman

s0064254

07-04-2011

University of Twente

Faculty of Behavioural Sciences

Communication studies - New Media, Research and Design

First supervisor: Dr. T.M. van der Geest

Second supervisor: Dr. D.K.J. Heylen

External supervisor: M. van Hout MSc

UNIVERSITY OF TWENTE.



Abstract

The present study investigated the relation between visual appeal judgements of websites and affect. Furthermore, it was investigated which method might prove most reliable in measuring the affective component of visual appeal judgements.

To investigate the relation between judgements of visual appeal and visceral affective responses, which are rapid subconscious good/bad judgements of a stimulus, six high and six low visual appeal website screenshots with a stimulus exposure time of 50ms were presented to participants. During stimulus exposure, electromyography (EMG) measurements of the corrugator supercilii muscle region were taken, and participants were subsequently asked to rate the level of visual appeal on a ten-point scale, and to indicate a LEMtool (Layered Emotion Measurement tool) emotion image. In the second phase of the experiment, six high and low visual appeal websites were displayed for 1s, during which eye-fixations were recorded using eye-tracking. In the final experimental phase the same twelve websites as in the second phase were presented without a time-limit. Participants were asked to give a visual appeal rating, and, for each screenshot, to use LEMtool to select visual elements that they had a particular feeling towards.

Facial EMG measurements showed participants experienced more negative affect, as indicated by heightened corrugator muscle activity, when giving low visual appeal ratings, compared to high ratings. Also, a significant negative correlation was found between visual appeal ratings and facial muscle activity. Furthermore, high visual appeal websites received significantly higher visual appeal ratings and received significantly more positive LEMtool indications, than low visual appeal websites and vice versa. These self-report findings were consistent between a stimulus exposure time of 50ms and free-viewing. Finally, eye-tracking revealed that there was no significant difference in the number of fixations and the relation between fixation and non-fixation duration, between high and low visual appeal websites.

Based on these results it was concluded that there is indeed a relation between visual appeal judgements of websites and affect. Because 50ms was enough time for participants to judge the level of visual appeal of the website screenshots, and because a significant correlation between visual appeal ratings and corrugator muscle activity was found, it is plausible that people make visual appeal judgements based on visceral affect. This notion was supported by LEMtool measurements, and eye-tracking data.

Finally, the triangulated measurements revealed that both facial EMG measurements and visual appeal rating scales proved reliable methods of measuring affective responses to the visual appeal of websites screenshots.

Acknowledgements

I would like to express my deepest gratitude to a number of people, without whom this project would not have reached its full potential.

First of all my supervisors, Thea van der Geest and Dirk Heylen, who overall gave me free reign, but provided the necessary structure, and nudged me in the right direction at points where I veered off course. Having them as my supervisors has been a major component in the successful completion of this thesis.

Next, Marco van Hout, who acted as external supervisor in the application of the LEMtool. Marco, together with Lars Rengersen at SusaGroup, have once again provided me with the best work environment possible. They have been a constant source of inspiration and motivation, and I owe a major part of my personal and professional development to them.

Also, everybody at Monito, Bas Jansen, Erik Orbons, and Martin Schoenmaker, again for providing an incredibly pleasant work environment, but also in setting up the online part of the pre-test, as well as the online environment used in the main experiment. Both these parts of the study were made possible by them.

Then, Matthijs Noordzij, for providing me with the opportunity to work with the EMG equipment and helping me make sense of it. Matthijs has been an incredible help in structuring the EMG data, and making the statistical analysis possible.

In the same light I have to thank Niels Baas, for his explanation of the eye-tracking hardware and software.

And finally, I would like to express my gratitude towards all my friends, and family, who have stood by me. I especially have to thank Merijn Bruijnes for doing a final check of this document, and Christina, who has provided tremendous support throughout, and who I have been able to rely on during the entire process of writing this thesis.

Contents

1. Introduction	11
1.1 Research context and scope	11
1.2 Research questions and basic research model	14
1.3 Thesis structure	16
2. Theoretical framework	18
2.1 Beauty, aesthetics, and visual appeal	18
2.2 The study of visual appeal in HCI	19
2.3 Basic components of emotion	22
2.4 Visceral affective responses	25
2.5 From visual appeal to affect	26
2.6 Measuring emotion	29
2.6.1 <i>The challenges of physiological measurements</i>	29
2.6.2 <i>Motor expression component: Facial EMG</i>	31
2.6.3 <i>Behavioral component: Eye-tracking</i>	33
2.6.4 <i>Subjective feeling component: Self-report</i>	34
2.6.5 <i>Subjective feeling component: The LEMtool</i>	35
2.7 Summary	37
3. Hypotheses	39
3.1 Motor expression component	39
3.2 Behavioral component	40
3.3 Subjective feeling component	40
3.3.1 <i>Visual appeal rating scales</i>	40
3.3.2 <i>The LEMtool</i>	41
3.5 Research model	42
4. Manipulation check	45
4.1 Stimuli pre-selection	45
4.2 Method	47
4.2.1 <i>Participants</i>	47
4.2.2 <i>Apparatus</i>	48
4.2.3 <i>Materials</i>	48

4.2.4 Procedures	48
4.3 Results	49
4.3.1 Visual appeal ratings	49
4.3.2 Perceived usability, content familiarity and structure familiarity	52
4.3.3 Response latency	54
4.3.4 Conclusions	56
5. Main experiment	59
5.1 Method	59
5.1.1 Participants	59
5.1.2 Apparatus	60
5.1.3 Materials	61
5.1.4 Procedures	61
5.1.5 Data reduction	64
5.2 Results	65
5.2.1 Motor expression component: Facial EMG	65
5.2.2 Motor expression component: Influence of mental effort	67
5.2.3 Behavioral component: Eye-tracking	69
5.2.4 Subjective feeling component: Visual appeal rating scales	70
5.2.5 Subjective feeling component: The LEMtool	73
6. Conclusions and discussion	77
6.1 Conclusions	77
6.1.1 General conclusions	77
6.1.2 Visceral affect and emotional categorization	80
6.2 Discussion	81
6.2.1 Limitations	81
6.2.2 Recommendations	83
References	85
Appendices	95
Appendix A - Stimulus set	95
Appendix B - Rating scales used in the pretest	100
Appendix C - Detailed analysis of cognitive constructs	101

1

Introduction

In this chapter, the research context and research scope will be outlined. Human-Computer Interaction and User Experience will be introduced as the central research fields. The concepts of affect and emotion will be related to User Experience, and the presumed relation between visual appeal and affect will be introduced. Subsequently, the central research questions will be presented.

1.1 Research context and scope

Until recently, Human-Computer Interaction (HCI) research has mainly focused on the investigation of the users behavioral goals, and the completion of instrumental tasks in judging the quality of the interaction between a user and a computer system. However, this nearly exclusive focus on effectiveness and efficiency has been called into question, and other, non-instrumental elements have been introduced in the past two decades (see for instance Carroll & Thomas, 1988). An increasing number of HCI researchers has recognized the importance of non-instrumental goals and needs. This field of investigation, that goes beyond the study of instrumental usability metrics to take into account a more holistic experience of technology, has been referred to as User Experience (UX) (Hassenzahl & Tractinsky, 2006).

ISO standard 9241-210 defines UX as “A person's perceptions and responses that result from the use or anticipated use of a product, system or service”. The subjective, experiential nature of UX is stressed in this definition. Furthermore, UX is a direct result of use or anticipated use (see also Law, Roto, Hassenzahl, Vermeren & Kort, 2009). The appreciation of art, for example, is not subsumed under this definition of UX since it does not relate to use or anticipated use. This also indicates a link between UX and usability in the sense that both concepts incorporate a component of actual system usage.

A clear definition is essential in a field where numerous disciplines, such as computer science, psychology, and design, converge. The fact that UX has only recently been defined in more detail, may have added to the fact that the empirical investigation of UX is still in its infancy. Yet, strong efforts have been made to combat the elusiveness and vagueness

sometimes associated with UX. Hassenzahl and Tractinsky (2006) submit several concepts that help to explain UX more fully:

UX is about technology that fulfills more than just instrumental needs in a way that acknowledges its use as a subjective, situated, complex and dynamic encounter. UX is a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organizational/social setting, meaningfulness of the activity, voluntariness of use, etc.). (p. 95)

Current UX research puts a strong emphasis on the user's internal state by looking at the concepts of affect and emotion as the central principles of UX (Hassenzahl & Tractinsky, 2006). Since the late 90's, affect has been studied in relation to computer use in a research field called Affective Computing (Picard, 1997). Affective Computing has greatly aided the understanding of affect in HCI, and has known a strong tradition in dealing with the question of how a computer system can detect (Picard, Vyzas & Healy, 2001) and respond to (Picard & Klein, 2002) negative emotions of the user. Studies in this direction deal with, for example, the role of user frustration (Scheirer, Fernandez, Klein & Picard, 2002) and how user frustration can be managed by a computer system (Klein, Moon & Picard, 2002).

While sharing the same interest in emotions in HCI with Affective Computing, UX research is less interested in the computer's side, but instead focuses on affective consequences on the human's side (Hassenzahl & Tractinsky, 2006). Here, the question is not how a computer could respond to a user's affective state, but how the user affectively experiences the interaction with the computer. UX focuses on positive affect, going beyond the prevention of negative emotions, and trying to elicit positive emotions, such as joy, delight and pride. The experience of such positive emotions may result in enjoyment of a system, and can be considered as a goal in itself in computer interaction (Norman, 2004b), for instance in the case of gaming (Mandryk, Inkpen & Calvert, 2006).

The elicitation of positive affect however, is no simple task. It is comparatively more straightforward to elicit negative affect during computer interaction. Several studies have demonstrated for instance, that users can be frustrated when button delays (i.e. unresponsive buttons) are introduced during a time-constrained task (Partala & Surakka, 2004; Reuderink, Nijholt & Poel, 2009; Scheirer et al., 2002). Absence of such causes of frustration however, is no guarantee for the experience of positive affect by the user.

One approach that has been suggested to elicit positive affect is presenting users with an attractive visual interface (Norman, 2004b; Tractinsky, 2004; Zhang, 2009). Norman

(2004a) proposes three levels of emotional processing. When considering visual aesthetics, the first and third levels are most important (Norman, 2004b). The first level is the visceral level of emotional processing. Norman (2004a) states that stimuli are first processed subconsciously, resulting in an initial good/bad-judgement, or indeed a visually appealing/unappealing-judgement. This visceral judgement happens very rapidly and is biologically determined. The visceral level of emotional processing provides subconscious preference information, early on in the perception process.

The second level is the behavioral level and deals with function, comprehensiveness, and usability (Norman, 2004a). The behavioral level, being related to usability, is less relevant in the current study and will therefore not be discussed further.

The third level of emotional processing is called the reflective level. This level is conscious and intellectually driven. Individuals prior experiences, self-image, and personal meanings shape considered judgements of an object (Norman, 2004a). Thus, at this level, individual preference differences for visual appeal emerge. For example, at the visceral level, everybody likes bright colors. However, due to, for instance, personal preference, not everybody likes brightly colored shoes at the reflective level.

While the visceral and reflective level of emotional processing (Norman, 2004a) may give a simple theoretical explanation of how affect may modulate visual appeal judgements, empirical evidence of the role affect plays in visual appeal judgements in UX, is scarce.

Zhou and Fu (2007) noted this hiatus and conducted a study using visually appealing and unappealing websites as affective primes. Using positive and negative words as targets, the authors looked at response times for the identification of the valence of the targets. Strongly valenced, validated affective images were compared to visually appealing and unappealing websites used as primes, and found that the websites had a priming effect similar to the strongly valenced images. For instance, when a visually appealing website preceded a positive target word, participants identified the word as positive more quickly than when a visually unappealing website preceded a positive target word. This effect is in line with affective priming research that uses validated affective stimuli.

Zhou and Fu's (2007) findings indicate that websites may indeed be processed as affective stimuli. This has also been suggested by Lindgaard, Fernandes, Dudek, and Brown (2006) in a study aimed at discovering how quickly individuals are able to judge a website for its visual appeal. In a series of experiments, participants were shown website screenshots, and were asked to make a judgement on how appealing or unappealing they found the visual design of the websites to be. High correlations were found between the 500ms and 50ms stimulus exposure conditions, suggesting that participants were able to assess the visual appeal of the website screenshots within 50ms.

Lindgaard et al. (2006) propose that this response could be classified as a pre-cognitive, mere exposure effect (Zajonc, 1980), whereby participants are able to indicate preference for a previously seen stimulus without recognizing the stimulus. Lindgaard et al. (2006) went on to suggest their findings relate to Norman's (2004a) visceral level of emotional processing. However, according to Norman, visceral emotional processes, because of their rapid and subconscious nature, can only be measured by using psychophysiological measurement techniques. Indeed, Lindgaard et al. (2006) suggested follow-up studies to incorporate, among other techniques, psychophysiological measurements to determine to what extent their findings can be seen as an affective response.

This suggestion is taken up in the current investigation in an effort to further investigate whether or not judgement of visual appeal of websites are affect related. This main research goal will be addressed in the form of specific research questions described in the next section.

1.2 Research questions and basic research model

It has been recognized that affect plays an important role in UX (Hassenzahl & Tractinsky, 2006). Presenting users with a visually attractive interface may be a way to elicit positive affect, yet the relation between affect and visual appeal of websites is not evident. Lindgaard et al. (2006 and Zhou and Fu (2007) provide some early indications that visual appeal in websites may be judged through visceral affective processing (Norman, 2004a). Still, more empirical evidence is necessary to clarify the relation between visual appeal of websites and affect. Therefore the first research question can be formulated as follows:

RQ1: How are judgements of visual appeal of websites related to affect?

In order to find an answer to this question, the methodology of measuring affect should be considered. As Lindgaard et al. (2006) (see also Norman, 2004a) have noted, psychophysiological measurements may be used to measure visceral affective responses. Moreover, Lindgaard's et al. (2006) findings indicate that consistent judgements of the level of visual appeal of websites can be obtained with self-report methods. Currently, most studies that investigate judgements of visual appeal based on brief stimulus exposures, use preference scales (Lindgaard et al., 2006; Tractinsky, Cokhavi, Kirschenbaum & Sharfi, 2006). However, as Norman (2004a) points out, this may not be a viable approach to measuring visceral emotional responses to visual appeal. The second research question can thus be formulated as follows:

RQ2: Which measurement techniques can reliably measure affective responses to visual appeal of websites?

To provide an answer to both research questions, an experiment will be conducted. Both brief (i.e. 50ms) and extended (i.e. 1s and free-viewing) stimulus exposure times will be used to gain a more complete understanding of the role affect might play in the judgement of high visual appeal and low visual appeal website screenshots.

Because affect is a complex and multifaceted phenomenon, the most complete understanding of affective responses is obtained by looking at multiple components of affect (Scherer, 2005). In the present study the motor expression component, related to facial expressions, will be studied by measuring facial muscle activity. Furthermore, eye-tracking will be used to provide insight into the behavioral component of affect. Finally, self-report will be used to assess the subjective feeling component. Both visual appeal rating scales, and a non-verbal emotion measurement instrument, the LEMtool (Layered Emotion Measurement tool) (Huisman, 2009; Huisman & Van Hout, 2010) will be employed.

Figure 1 visualizes the relations between the main concepts of the current study. The basic research model shows that the perception of visual appeal may lead to affect. Each of the three components of affect, is in fact measurable output from the affective experience that is a result of the perceived level of visual appeal of a stimulus.

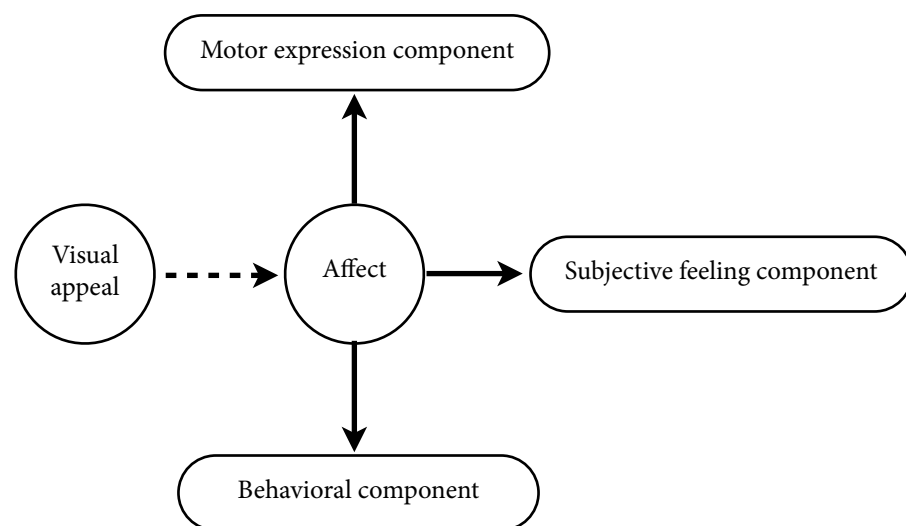


Figure 1 Basic research model

Note that this model is a basic research model and does not explicate the theorized relation between visual appeal and affect. Furthermore, it does not depict all facets of affect, nor does it explain the nature of their relation to each other. The following chapter will present a theoretical framework that will explicate the concepts and relations depicted in the model.

1.3 Thesis structure

This thesis consists of six chapters of which the Introduction is the first. Each chapter was written to take into account readers who want to restrict their reading to a certain section or chapter. Each chapter will therefore start with a short summary of the main topics of the previous chapter, as well as introduce the topics of the chapter itself.

In the second chapter the theoretical foundation of the current study will be outlined. First, a working definition of visual appeal will be given. Then, previous research into visual appeal in HCI will be discussed. This will be followed by an outline of relevant issues in studying emotion and affect. Next, the relation between visual appeal and affect will be explicated. The chapter will conclude with a discussion of relevant emotion measurement techniques.

Chapter 3 will present seven hypotheses that serve to structure the main experiment and provide an answer to the research questions posed in Chapter 1.

The fourth chapter will present the design and results of the manipulation check that was performed. The goal of this manipulation check was to determine if a preselected set of 24 website screenshots differed on visual appeal.

In Chapter 5, the websites that were subjected to the manipulation check were used as the stimuli. Facial electromyography, eye-tracking, rating scales, and LEMtool were used to assess reactions to high visual appeal and low visual appeal websites with differing exposure durations. Findings from the main experiment will be reported in this chapter.

The sixth and final chapter will present the conclusions based on the findings of the main experiment. In addition, limitations of the current study will be discussed and recommendations for future research will be given.

2

Theoretical Framework

In the previous chapter, the field of User Experience (UX) was introduced. It was explained how visual appeal plays a role in UX, and it was proposed that visual appeal is related to affect.

In this chapter, concepts introduced in the first chapter will be further explicated. First, a working definition of visual appeal will be given. Then, previous studies into visual appeal in human-computer interaction will be discussed. Both usability and UX related studies will be discussed, and a hiatus in the current body of research will be identified. Next, the basic components of emotion will be discussed. The innate nature of emotion will be explained, and a basic affective mechanism will be presented. Then, an explanation of the relation between visual appeal and affect will be provided. The chapter will conclude with a discussion of relevant emotion measurement techniques.

2.1 Beauty, aesthetics, and visual appeal

The study of beauty dates back to ancient Greek philosophers and scholars, and has remained a topic of great interest throughout history. During some time periods, beauty has been attributed with divine qualities, while during others, it has been marginalized in arts and science (Frohlich, 2004; see also Lavie & Tractinsky, 2004). The term beauty itself has been used interchangeably with terms such as aesthetics, and visual appeal and has been defined in terms of object properties or as existing purely ‘in the eye of the beholder’ (Frohlich, 2004; Norman, 2004b).

In research into beauty, three general views can be defined (Reber, Schwartz & Winkielman, 2004). The first, called the *objectivist* view, assumes that properties of the object, such as color saturation and contrast, symmetry and simplicity, is what constitutes beauty. The objectivist view proposes that beauty can be ‘constructed’ by combining these object properties in a certain manner. This view was dominant in the 16th century.

According to the second view, called the *subjectivist* view, any object can be beautiful, as long as it pleases the senses of an individual. Beauty is a function of the specific preferences of an individual. Defining these preferences is impossible because they are formed through

personal experiences. The expression ‘beauty is in the eye of the beholder’ fits the subjectivist view of beauty (Reber et al., 2004).

The third view states that beauty emerges in the interaction between people and objects, and is referred to as the *interactionist* view. The patterns that exist between the object and the perceiver is what accounts for beauty. In essence this is a combination of the first two views: elements of objects can be pleasing to everyone (i.e. objective), but these preferences can be modified when individual subjective experiences comes into play (Reber et al., 2004).

The interactionist view of beauty is currently the most prominent (Reber et al., 2004), and matches with theories of perception of aesthetics and affect (Hassenzahl, 2007; Lindgaard & Whitfield, 2004; Norman, 2004a). Still, there remains considerable debate about the exact composition of beauty, especially in HCI (Frohlich, 2004; Hassenzahl, 2004a; Norman, 2004b; Tractinsky, 2004). Because of the ongoing nature of this discussion, the current proposal will adopt the term ‘visual appeal’ in favor of beauty or aesthetics. This has several advantages.

First, the term ‘visual appeal’ denotes a visual experience, whereas aesthetics and beauty can relate to other sensory modalities. Stressing the visual nature of the construct under study, helps to limit the scope of the current investigation to the purely visual.

Second, the term visual appeal is used in studies with a similar approach as the current investigation (Lindgaard et al., 2006; Tractinsky et al., 2006), making comparisons more straightforward.

Third, visual appeal can be described as an attractive-unattractive judgement of an object. It is in essence a statistical construct (Frohlich, 2004) that becomes apparent as a difference in the judgement of attractiveness of two objects. On this basis, no claims can be made about the ‘beauty’ of the objects as such. Thus, using the term visual appeal, distances the current study from the discussion of the essence of beauty or aesthetics.

The interactionist view (Reber et al., 2004) would hold that judgements of visual appeal may be partially stable (i.e. defined by objective properties of the stimulus) but may also be subject to personal preferences (i.e. subjective experience). Both affective and cognitive processes can modulate visual appeal judgements in this view. However, in general as well as regarding visual appeal specifically, much HCI research to date focusses on cognitive processes in relation to usability. The next section will discuss the way visual appeal has thus far been studied in HCI.

2.2 The study of visual appeal in HCI

Visual appeal has been studied in a variety of ways in HCI. Considering that HCI research in general has a strong tradition of investigating usability, it is not surprising that HCI research

focussing on the role of visual appeal in computer interaction is grounded in usability research. This has yielded ample evidence that the visual design of an interface has a profound effect on the use of the system

For example Tractinsky, Katz, and Ikar (2000) found that an attractive visual design impacts the perception of the usability of a computer-simulated automated teller machine (ATM), independent of the actual usability of the ATM. Participants indicated, both before and after using the system, that it was more usable when it had a pleasing visual design. Actual usability, manipulated by introducing system delays and unresponsive buttons, did not have an impact on the perception of either usability or visual appeal.

A similar result was found by Sonderegger and Sauer (2010) who used visually appealing and unappealing simulations of mobile phone interfaces. Participants who used the appealing mobile phone interface rated their device as being more usable, and showed a lower task completion time than did participants who used the unappealing mobile phone interface. Comparable results were found for website stimuli. Moshagen, Musch and Göritz (2009) presented participants with visually appealing and unappealing health-related websites, and found visual appeal to reduce task completion time and error rates for poor usability websites. Apart from corroborating the results of Tractinsky et al. (2000), these results suggest that actual usability metrics, such as task completion time, are influenced by visual appeal.

Other studies, using specific manipulations of visual elements, also found that perceived usability was positively influenced by a visually appealing interface. These manipulations include manipulations of color (Nakarada-Kordich & Lobb, 2005), shape (Ben-Bassat, Meyer & Tractinsky, 2006), and visual complexity (Tuch, Bargas-Avila, Opwis & Wilhelm, 2007) of a visual interface.

These studies indicate a strong relation between perceived usability and visual appeal, as well as provide indications that actual usability metrics are influenced by visual appeal. This seems to suggest that the main purpose of visual design in HCI, is a functional one, aimed at improving usability. Indeed, efficient and effective communication with the user through the appropriate visual design of an interface, is a goal most designers seek to accomplish (Lavie & Tractinsky, 2004). Nonetheless, visual appeal is not purely perceived as functional. Research has indicated that creative (Lavie & Tractinsky, 2004) and self-oriented (Hassenzahl, 2004b) attributes are important in the perception of visual appeal in interfaces as well. These attributes do not necessarily relate to usability, but to the pleasure a high visual appeal interface might evoke in its users.

Lavie and Tractinsky (2004) have made strong efforts to explain the role of visual appeal in HCI more fully. They propose to view visual appeal as consisting of a usability related dimension and a dimension related to the inherent pleasure an interface might evoke.

They coined these dimensions respectively 'pragmatic' and 'expressive'. The pragmatic dimension emphasizes orderly and clean design, as is often advocated in usability literature. The expressive dimension on the other hand, stresses creativity and originality, which relates to the expressiveness of the design.

In a similar fashion Hassenzahl (2003, 2004b) has identified pragmatic and hedonic attributes of general product appeal in HCI. Pragmatic attributes, similar to Lavie and Tractinsky's (2004) pragmatic dimension, are strongly related to usability, highlighting the importance of controllability and functionality. Hedonic attributes on the other hand, emphasize an individuals' psychological well-being (Hassenzahl, 2003, p.35). Hassenzahl (2007) notes that Lavie and Tractinsky's (2004) expressive dimension of visual appeal closely matches his hedonic dimension of the multi-component model of product appeal. Hassenzahl (2007) states that this is problematic, as it indicates that Lavie and Tractinsky's (2004) definition of visual appeal may be too broad, overlapping with definitions of overall appeal (see also Lindgaard et al., 2006). Instead, Hassenzahl (2007) proposes to treat visual appeal as a sensory input for the judgement of an object. The physical, visual nature of the object is stressed here.

The ongoing discussion points to the difficulty of studying visual appeal in HCI, where both usability and UX are concerned. While Hassenzahl's (2003, 2007) and Lavie and Tractinsky's (2004) approaches seem to concur on the pragmatic, usability related dimension of visual appeal, there is far less consensus on the UX related dimension. Moreover, where Lindgaard et al. (2006) and Zhou and Fu (2007) propose visual appeal to be affect-driven, previously mentioned studies only hint at the role of affect. The current investigation therefore aims to provide further evidence for the relation between visual appeal and affect in HCI. This relation has only received marginal attention in the study of visual appeal in HCI. Both theoretical explanations and empirical evidence are lacking.

Looking at emotion research, useful concepts for investigating the relation between affect and visual appeal in HCI can be extracted. Norman's (2004a) distinction between visceral, behavioral, and reflective levels of emotional processing is a tantalizing starting point, and it touches upon some of the main concepts that are most often represented in emotion theory. However, for the purpose of gaining a more detailed view of what an emotion is exactly, how it differs from affect, and which elements relate to visual appeal, central concepts in emotion theory will be discussed more thoroughly in the following sections.

2.3 Basic components of emotion

Currently there is a number of basic aspects of the concept of emotion that receive strong support from the research field in general. First, emotions are considered reactions to events that are relevant to concerns of an individual (Frijda, 1986). Examples of concerns would be a concern for safety, or a concern for social status. Every stimulus (internal or external) is subjected to a relevancy check according to the concerns of an individual. Emotions only occur when the stimulus is seen as relevant; someone does not get emotional about something that has little or no relevance to them personally (Desmet & Hekkert, 2007).

Second, emotions are considered a multifaceted phenomenon consisting of several components (Brave & Nass, 2008; Scherer, 2005). Scherer's (2005, 2009) Component Process Model provides a detailed insight into these central components (Table 1).

Table 1
Relationships between organismic subsystems and the functions and components of emotion (Scherer, 2005, p. 698)

Emotion function	Organismic subsystem and major substrata	Emotion component
Evaluation of objects and events	Information processing (CNS)	Cognitive component (appraisal processes)
System regulation	Support (CNS, NES, ANS)	Neurophysiological component (bodily symptoms)
Preparation and direction of action	Execute (CNS)	Motivational component (action tendencies)
Communication of reaction and behavioral intention	Action (SNS)	Motor expression component (facial and vocal expression)
Monitoring of internal state and organism-environment interaction	Monitor (CNS)	Subjective feeling component (emotional experience)

Note: CNS = central nervous system; NES = neuro-endocrine system; ANS = autonomic nervous system; SNS = somatic nervous system.

To give an example of how an emotion might occur as a change in the five components of Table 1, fear for instance, is an emotional response to a threat to an individual's well-being (i.e. evaluation of the stimulus as a threat to the individuals concern for safety). Once the stimulus is considered relevant, (neuro-)physiological responses such as sweating and the release of hormones, in this case most likely adrenaline, occur. In addition cognitive and physical preparation for action (e.g. fleeing), a distinctive facial expression (e.g wide open eyes), and a strong negative subjective feeling, are hallmarks of the emotion of fear (see also Brave & Nass,

2008). Following Scherer's (2005) approach, one can only speak of an emotion when changes in all of the five organismic subsystems take place. Hence, the term affect is used here to denote elements that relate to emotion as defined by the component process model, but that in themselves cannot be described as a full emotion.

Scherer (2009) explains that there are four different types of appraisal objectives that follow a fixed order, namely: relevance; implications; coping potential; and normative significance. An organism evaluates an event based on a number of stimulus evaluation checks (SEC) in order to attain these objectives. Emotions are a result of a coordination of the five organismic subsystems, driven by the appraisal results. Note, that appraisal can occur with a lot, or a little, cognitive processing (Scherer, 2005, 2009). For the current investigation the SEC of intrinsic pleasantness, subsumed under the relevance appraisal objective, and therefore one of the first SEC's an organism employs in order to appraise an event, is most important. The intrinsic pleasantness SEC asks whether or not an event is intrinsically pleasant, independent of the organisms current motivational state. This pleasantness/unpleasantness check matches closely to what Norman (2004a) describes as visceral affect.

A way to further distinguish between emotions with a stronger or weaker cognitive component, is by looking at the three key areas of the brain that are most relevant for emotions: the thalamus; the limbic system, consisting of the hypothalamus, hippocampus and the amygdala; and the cortex. Environmental information is first processed by the thalamus, which sends information simultaneously to the limbic system and the cortex. The latter is involved in higher level processing, while the former constantly evaluates the relevancy of the received input. If the input is determined as relevant, the limbic system sends signals to the body, coordinating physiological responses, and to the cortex, biasing attention and other cognitive processes (Brave & Nass, 2008). Figure 2 depicts the relations between these different brain structures.

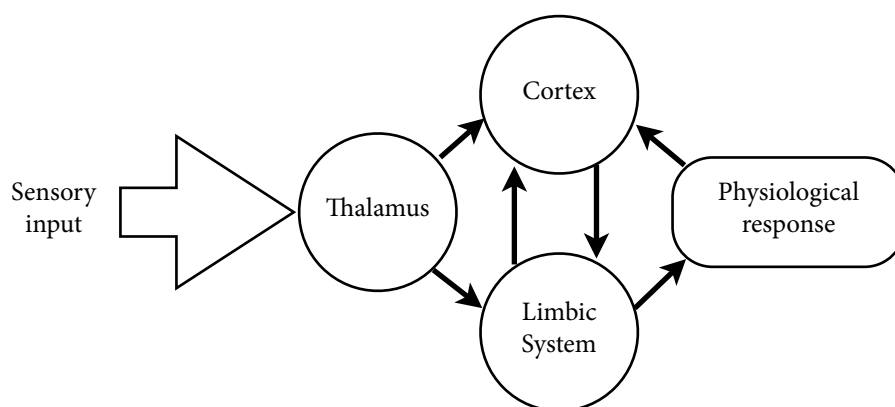


Figure 2 Neurophysiological structure of emotion (adapted from Brave & Nass, 2008, p.78)

The direct link between the thalamus and the limbic system is what accounts for more primitive emotions such as startle-based fear. Indications of these types of emotions in HCI have been found by Reeves and Nass (2002), who showed that users have emotional responses to moving objects on a screen, large images, and images in peripheral vision. The thalamic-limbic path also accounts for what Norman (2004a) refers as subconscious visceral responses. These responses are automatic, and mainly consist of a good or bad judgement that occurs before the stimulus has been cognitively processed by the cortex. Evolutionary theorists emphasize differentiation of emotions in the limbic system (Darwin, 1872/1998; see also Ekman, 1994; Izard, 1994). In this view, each emotion is likely to have a very specific pattern of physiological and cognition-biasing responses (Brave & Nass, 2008).

Different from emotions where the thalamic-limbic pathway plays a central role, are emotions that are a result of extensive cognitive processing. These emotions stem from the link between the limbic system and the cortex. Emotions resulting from this link are sometimes referred to as 'secondary emotions' (Brave & Nass, 2008), being more heavily influenced by memory and personal preferences. Hence, they are more culturally determined and can vary significantly between individuals. This is what Norman's (2004a) reflective level of processing refers to. Some theorists even argue that emotions are entirely learned social constructs. Here, emotions are purely differentiated within the cortex, while the limbic system is merely a sign of emotional valence (i.e. a positive/negative judgement) or arousal (Ortony, Clore & Collins, 1988; Ortony & Turner, 1990; Russell, 1980, 1994). Any consistencies between cultures are explained by common social structures (Brave & Nass, 2008).

In essence, Lindgaard et al. (2006) propose that their findings on visual appeal judgements after a stimulus exposure time of 50ms, are related to affective responses stemming from the thalamic-limbic pathway. However, because of the self-report format used to assess the subjective feeling component of emotion in their study, some cognitive processing (i.e. cortico-limbic pathway) is required to make this conscious judgement. If these judgements are related to affect, they seem to result from an interaction between the limbic system and the cortex (see Figure 2). Indeed there are other theories that do acknowledge the existence of some basic emotions (e.g. happiness, sadness, fear, anger, and disgust), but also recognize the existence of socially learned emotions, such as pride (Ekman, 1992; Elfenbein, Beaupré, Lévesque & Hess, 2007). From this perspective the limbic system is capable of differentiating between a limited number of basic emotions, while the cortex processes socially learned emotions. This view best matches the Component Process Model (Scherer, 2005), in that the model states that emotions can occur with varying levels of cognitive processing.

The interplay between the limbic system (i.e. visceral level of emotional processing) and the cortex (i.e. reflective level of emotional processing) is eloquently illustrated by

Norman (2004a) when he talks about the joy of owning an original piece of artwork. The artwork may be beautiful to look at (i.e. be viscerally pleasing), but if this was all there is to it, a high quality reproduction would elicit the same emotions. The pleasure of owning an original comes from the reflective value of owning it, and is thus a strongly social and cultural affair. What is of interest to the current study, is the initial perception of the artwork. This initial perception, through the tahlc-lymbic pathway, would lead to a rapidly occurring affective response. Numerous studies into affect and emotion have found such responses in differing contexts. The next section will present some of these studies.

2.4 Visceral affective responses

Lindgaard et al. (2006) point out that, when first encountered, young children show fear of large, dark, and noisy objects that move rapidly towards them. Detection of these objects by the sensory system, including the amygdala, is sufficient to initiate an immediate response. Exact recognition of the objects is unnecessary to elicit this response (Lindgaard et al., 2006, p. 176; see also Lindgaard & Whitfield, 2004). A similar response seems to occur when individuals judge the visual appeal of websites with brief stimulus exposure times and subsequently judge the same websites with an extended exposure time (Lindgaard et al., 2006; Tractinsky et al., 2006). Lindgaard et al. (2006) propose a mechanism similar to the mere exposure effect (Zajonc, 1980) to accounts for these findings.

In a series of experiments it was demonstrated that participants were consistently able to indicate preference for a stimulus they were previously exposed to, without actually recognizing the stimulus as familiar (Zajonc, 1980). In some experiments visual stimuli (e.g. random polygons) were presented extremely briefly, in some cases for just 1ms. This led to the conclusion that there must exist some sort of pre-cognitive affective mechanism that allowed participants to judge previously seen stimuli as more preferable than novel stimuli. The mere exposure effect has been replicated in numerous studies and proves to be very robust (Bornstein & D'Agostino, 1992).

Other studies that highlight the rapid, subconscious nature of affect are studies into affective priming. The affective priming paradigm is based on the congruency effect, that holds that when the affective prime is of the same valence as the target (e.g. both are positive) evaluation of the target (a positive or negative word) will be facilitated, and response latency in identifying the target will decrease (Fazio, 2001). Studies have shown affective primes to influence later judgements and even behavior (Winkielman, Berridge & Wilbarger, 2005). In affective priming studies using visual primes, it is typical to display either positive or negative affective images (Hermans, Spruyt, De Houwer, & Eelen, 2003) such as those from the

International Affective Pictures System (IAPS) (Bradley & Lang, 2007), or a positive or negative facial expression (Winkielman et al., 2005).

The mere exposure effect and affective priming use implicit measurements (preference ratings and response latency, respectively) of affect. Other studies have focussed on physiological and motor expression responses as indications of affect. Measurements of facial muscle activity (i.e. facial electromyography (EMG)) for instance, found that the formation of emotional facial expressions occurs within a few milliseconds after stimulus exposure (Dimberg & Thunberg, 1998; Ekman, 1992) highlighting the rapid, precognitive nature of these expressions.

Similar findings were obtained in studies into the psychophysiology of affect. It has been demonstrated that when presenting participants with affective stimuli from the IAPS displayed for only a few milliseconds (e.g. 25ms), physiological responses (e.g. skin conductance, heart rate, facial electromyography, etc.) in line with the valence of the target stimulus can be found (Codispoti, Mazzetti & Bradley, 2009; Codispoti, Bradley & Lang, 2001; Smith, Löw, Bradley & Lang, 2006). Codispoti's et al. (2009) study showed, that when a visual mask directly followed after stimulus exposure, skin conductance proved to be less discriminate, whereas facial EMG showed affective discrimination for differently valenced images with brief stimulus exposure.

Summarizing, a visceral affective mechanism allows people to make judgments of briefly presented stimuli, such as visual appeal judgements of websites. Research into emotional facial expressions shows that measuring facial muscle activity with EMG is a viable way to measure visceral affective reactions in the motor expression component of emotion (Aue & Scherer, 2008; Scherer & Grandjean, 2008). Applying facial EMG measurements as a measure of the motor expression component when viewing high and low visual appeal websites however, still rests on the assumption that visual appeal is indeed related to affect. Therefore, the next section will provide support for the relation between visual appeal and affect.

2.5 From visual appeal to affect

Visceral affective responses that are thought to be at the basis of visual appeal judgements (Lindgaard et al., 2006) are most likely related to the thalio-limbic pathway in the brain. Indeed fMRI studies have shown that when presented with a stimulus previously judged as visually appealing, the right amygdala, which is part of the limbic-system, is activated. This supports the idea that the more subjective aspect of visual appeal judgements is mediated by association processes with the observer's affective experience (Cinzia & Vittorio, 2009, p. 686).

Additional empirical evidence for the relation between affect and visual appeal was found by Zhou and Fu (2007) who took a novel approach to the affective priming paradigm. The authors used visually appealing and unappealing websites as affective primes and compared them to validated affective stimuli. They found high visual appeal websites to function similarly to positive affective images and low visual appeal websites to function similarly to negative affective images. For example when a positive target word was preceded by high visual appeal website prime, identification of the target word as positive was quicker than when the target was preceded by a low visual appeal website. The reverse was true for negative target words. This shows a relation between validated affective stimuli, and visually appealing and unappealing websites used as affective primes.

The studies discussed so far seem to demonstrate a visceral affective mechanism that allows individuals to judge stimuli based on a minimum of information. However, few studies provide a theoretical explanation of how such an affective mechanism might be related to judgements of visual appeal.

The most comprehensive account of the role of affect in visual appeal judgements is given by the categorical-motivational model (Lindgaard & Whitfield, 2004; Whitfield, 2005). It states that objects are not evaluated per se, but are judged in relation to the cognitive category accessed. For example a chair is not judged as a discrete chair, but as a member of the cognitive category 'chair'. The way people respond to objects is determined by the categories already developed for such objects, either because they are hard-wired and part of the genetic infrastructure, or through learning. In this case, all chairs a person has ever encountered.

The categorical-motivational model is bipolar. At one end are categories that are already formed and closed to further articulation (i.e. are hard-wired or existing knowledge structures), perceiving stimuli will not change these categories. At the other end are categories that are ill-formed and open to further articulation, perceiving stimuli may further define these categories (i.e. making them more closed). At the closed end, the extent to which an object is prototypical of the category determines what people feel towards the object. Objects that fit better in the category accessed, are experienced as more pleasurable. Put simply, a chair should look like a chair (Whitfield, 2005). At the open end, where categories are ill-defined, positive evaluation of stimuli comes from the further articulation of categories (i.e. creating knowledge). It is important for further articulation of categories, that stimuli are neither extremely novel, and therefore meaningless, nor too well defined, and therefore mundane. In other words, a chair should have novel and creative features (Whitfield, 2005).

What links visual appeal to affect is that objects are not only categorized based on perceived features of the object, but can also be emotionally categorized (Niedenthal, Halberstadt & Innes-Ker, 1999). This means that objects that evoke similar emotional

responses, consisting of components such as motor expression, behavior and subjective feeling, can be categorized based on the similarity of these emotional responses. Furthermore, these emotional categories are open and subject to articulation, meaning they can change over time and in different contexts.

As previously discussed theories, such as the mere exposure effect, demonstrate, people can express preference for visual stimuli without really having seen the stimulus at all (i.e. processed by the visual cortex). A direct path from the amygdala, a part of the limbic system that plays a key role in affect, to the eye, accounts for these findings (LeDoux, 1994, 1998). The amygdala, as explained earlier, also accounts for early affective responses. Visual appeal judgements, as sensory-perceptual information coming into the eye, can therefore be quickly and subconsciously categorized on the basis of affect through the amygdala. This matches, what Norman (2004) refers to as, visceral affective responses.

In this sense visual appeal judgements can be thought of as a process for acquiring sensory-perceptual knowledge (Lindgaard & Whitfield, 2004; Whitfield, 2005) through the further articulation of emotion categories (Niedenthal, et al., 1999). In essence this is an attachment of affect to cognition (LeDoux, 1994, 1998; Lindgaard & Whitfield, 2004) which enables an organism to anticipate how the effects of alternatives would 'feel'. Damasio (2000) refers to this affect-cognition connection as a 'somatic marker', which is stored knowledge about prototypical affective categories.

When presenting participants with visually appealing and unappealing websites, it is likely emotional categorization occurs when such stimuli are displayed for a brief duration (i.e. 50ms), where it is difficult to distinguish features of the object that can help normal categorization. The direct path from the eye to the amygdala (LeDoux, 1994, 1998) would allow participants to do this. In the case of novelty, categorization can occur by further articulating affective categories. In the case such categories, or somatic markers already exist, categorization can occur by anticipating the affective response stored in the somatic markers. Either way, an emotional response consisting of coordination of the five main components of emotion (Table 1) will occur.

As stated in the Introduction (Chapter 1, Section 1.2) the current study focusses on three of the five components of the Component Process Model (Scherer, 2005). Each of these three components, the motor expression component, the behavioral component, and the subjective feeling component, is thought to accompany emotional categorization. The next section outlines techniques for measuring these three components in order to gain an insight into the affective response to visually appealing and unappealing websites.

2.6 Measuring emotion

Because a full emotion requires the presence of all five components of the Component Process Model (Scherer, 2005), the current study does not provide a complete measurement of emotion. Nevertheless, measuring a limited number of emotion components still provides valuable insights into an individual's affective state.

The motor expression component will be measured using facial EMG. Facial EMG was selected as a measurement because it reveals muscle activity that occurs shortly after stimulus onset (Dimberg & Thunberg, 1998; Ekman, 1992). Moreover, studies show corrugator EMG can discriminate between briefly presented positive and negative affective stimuli (Codispoti et al., 2001; Smith et al., 2006). It has been demonstrated that corrugator EMG is one of the most discriminative methods when it comes to measuring affective physiological responses to brief stimulus exposures of visual stimuli, even when visual masks are presented after stimulus offset (Codispoti et al., 2009). However, some limitations that are specific to measurements of physiology of emotion have to be taken into account (Van den Broek, Janssen, Westerink & Healey, 2009). Some of these limitations will be discussed in the next section.

As a measure of the behavioral component of emotion, eye-tracking will be used. Both affective stimuli (Nummenaa, Hyönä & Calvo, 2006) as well as visual appeal in websites (Djamasbi, Siegel & Tullis, 2010) have been studied using eye-tracking. The fact that visual stimuli are used in the current study, makes eye-tracking a viable measurement technique to provide insight into behavioral patterns for high and low visual appeal websites.

Finally, the subjective feeling component will be studied using two methods of self-report. First of all, visual appeal rating scales, identical to those used in studies on the consistency of visual appeal judgements of briefly presented websites (Lindgaard et al., 2006; Tractinsky et al., 2006), will be used. Furthermore, LEMtool (Huisman, 2009; Huisman & Van Hout, 2010) will be employed as a non-verbal self-report measurement of subjective feeling.

The following sections will discuss each of these measurement techniques related to each of the three selected components of emotion, starting with the motor expression component, followed by the behavioral component, and, finally, the subjective feeling component. However, first the challenges of physiological measurements will be discussed.

2.6.1 *The challenges of physiological measurements*

Before discussing facial EMG as a measurement of the motor expression component of emotion, it is important to first make apparent the challenges inherent in this type of measurement of affect.

First of all, there remains some debate as to the exact nature of human physiology in relation to emotions. Frijda (1986) for instance, distinguishes three main categories of physiological responses. These include: autonomic responses related to the functioning of muscles and internal organs (such as heart rate and skin conductivity) as mediated by the autonomic nervous system; changes in hormonal secretions; and neural responses, such as changes in alpha brain-wave patterns, measurable through electroencephalography (EEG).

A thorny issue with any of these categories of physiological data is that it is hard to discern what these data relate to exactly. Consider a user working with a badly designed interface. Physiological signals, such as heart rate, blood volume pressure, and skin conductivity, may point out that there is *something* happening. However, what this 'something' is, is hard to determine. Frijda (1986) explains how several signals that have been described as being related to emotional arousal or valence are also indications of mental effort, a purely cognitive construct. It is therefore imperative to experimentally control for possible rival explanations of physiological data.

This is what Ward and Marsden (2004) hinted at when they responded to their own study on physiological signals during interface interaction (Ward & Marsden, 2003) by stating that "All that can be said is that the findings demonstrate that human physiology does respond as might be expected to events that take place during real human-computer interaction" (Ward & Marsden, 2004, p. 708). This describes the difficulty of identifying cause and effect in studies that employ physiological measurements, where signals can indicate many affective or cognitive processes occurring.

Moreover, physiological signals often show a significant delay between the stimulus presentation and the affective physiological response. Anttonen and Surakka (2005) found that heart rate could only be used to discriminate between positive and negative stimuli, six seconds after stimulus presentation. This poses serious problems for the measurement of visceral responses to stimuli presented for a short duration.

One approach that has been suggested to be more discriminative and more reliable when it comes to measuring emotions, is facial expression analysis. Facial expressions have been shown to be cross-cultural indications of specific emotions (Ekman, 1994). The study of emotional facial expressions has led to the classification of 44 facial muscles that are used to form facial expressions of emotion. This classification has been further developed into the Facial Action Coding System (FACS) (Ekman & Friesen, 1978) which can be used by trained observers to recognize emotions through facial expressions.

An issue with measuring emotions through facial expression analysis is that not all emotional experiences are accompanied by overt facial expressions (Cacioppo, Bernston, Larsen, Poehlman & Ito, 2000). Especially where weak stimuli are considered, facial

expressions may either be minimally, or not observable at all. As visual stimuli, most websites would be considered emotionally weak stimuli, especially when compared to affective images from the IAPS (Bradley & Lang, 2007), containing erotica, mutilated bodies, and threatening animals.

All this, however, does not mean no facial muscle activity is present even when moderate emotional stimuli are presented. Electromyography (EMG) can be used to measure minute, and rapidly occurring changes in electrical activity of muscles (Tassinary, Cacioppo & Vanman, 2007). Therefore, facial EMG is an appropriate measure for investigating affective responses to visual stimuli presented for a very short duration. This will be further explicated in the next section.

2.6.2 Motor expression component: Facial EMG

The power of facial EMG as a measurement of the motor expression component of emotion, lies in the fact that it is capable of detecting very small and very fast facial muscle responses, even when such responses are elicited by moderate stimuli (Winkielman & Cacioppo, 2001). This makes facial EMG a viable technique to be used to measure affective responses to visually appealing or unappealing websites, which should be considered ‘moderate’ emotional stimuli.

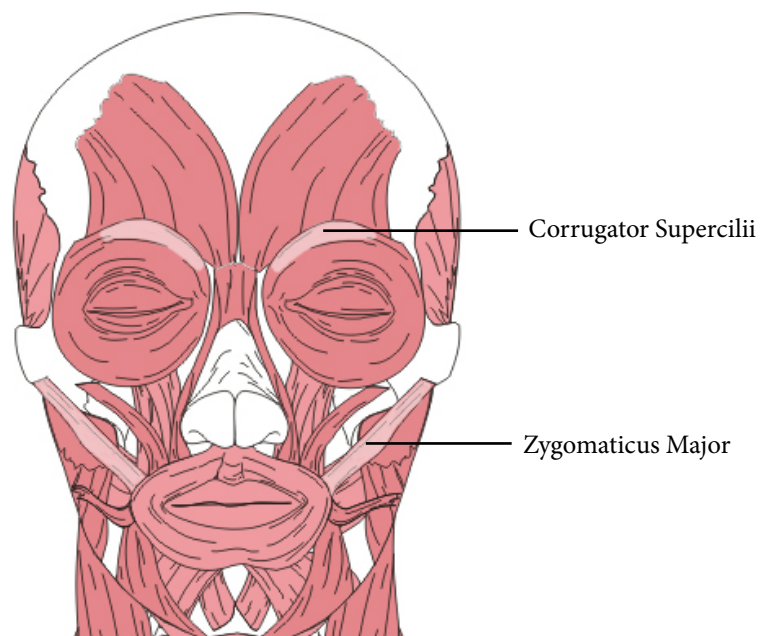


Figure 3 Basic schematic of human facial muscle structure. The corrugator supercilii (brow) and the zygomaticus major (cheek) muscle regions are highlighted.

Bradley and Lang (2007) state that measuring facial muscle region activation in the eyebrow (corrugator supercilii) and the cheek (zygomaticus major) (Figure 3) are reliable indicators of emotional valence. Activation of the corrugator muscle region, a hallmark of frowning, has been found to be strongly related to the experience of negative valence. Neutral stimuli result in moderate activation, where the experience of positive valence does not show activation of the corrugator supercilii (Cacioppo et al., 2000; Codispoti et al., 2009). Corrugator activity may also point to the intensity of the negative stimulus, as participants showed greater activation of the muscle in response to highly arousing negative images such as mutilated bodies (Bradley & Lang, 2007; Cacioppo & Petty, 1986). Conversely, activation of the zygomaticus major, which is related to smiling, is indicative of positive emotions, although the relation is less strong than that between the corrugator supercilii and negative emotions (Bradley & Lang, 2007; Cacioppo & Petty, 1986; Larsen, Norris & Cacioppo, 2003).

Measurements taken from the corrugator supercilii and the zygomaticus major muscle regions have been used to assess valenced responses in human-computer interaction. For example, Tuch et al. (2009) found that corrugator EMG activity was positively correlated with visual complexity of websites. Path analysis indicated that EMG responses were mediated through perceived valence, which would suggest that the measured EMG activity stemmed from an affective response.

Hazlett and Benedek (2007) applied facial EMG measurements of the corrugator supercilii and zygomaticus major muscle regions to study emotional valence during software use. Both in response to static stimuli in the form of screen mock-ups of an operating system, as well as in response to actual system use (i.e. using a media player), the authors found that facial muscle activity was a reliable indicator of positive and negative affect. The results from the facial EMG measurements also showed a strong relation to self-report of desirable product features in both the static and interactive experiment. However, the authors also note that during the interactive experiment, elevated facial EMG activity was most likely a sign of increased mental effort, rather than an emotional response. Most interesting for the current investigation is that in response to screen mock-ups, self-reports of visual appeal of software features corroborated EMG measurements. This indicates that facial EMG can be used to assess affective responses to visual appeal of website screenshots.

In both Tuch's et al. (2009) and Hazlett and Benedek's (2007) studies, stimuli were presented for a long duration (e.g. 8s in the study of Tuch et al. (2009)). Considering the difficulty of judging the exact cause of physiological responses, these long exposure durations are not viable for studying facial EMG responses to visual appeal. It is plausible that during longer stimulus exposures, besides a first impression of the visual design, participants take in other information such as texts, and menu structures. Presenting website stimuli for a very

brief duration (i.e. 50ms) would limit cognitive processing, due to the fact that the brief exposure duration limits the amount of information that can be processed by the participant. There is evidence that even with very brief exposures of affective visual stimuli, facial EMG measurements of the corrugator supercilii muscle region, can reliably differentiate between positive, neutral, and negative stimuli (Codispoti et al., 2001; Codispoti et al., 2009; Smith et al., 2006). When visual stimuli from the IAPS were displayed for as brief as 25ms, this distinction remained. Moreover, when a visual mask was introduced after stimulus offset, corrugator EMG measurement proved to be one of the most reliable indicators of positive or negative affect, among other physiological measures such as skin conductance (Codispoti et al., 2009).

What can be gathered from these studies using facial EMG, is that the corrugator supercilii muscle region is the most reliable indicator of emotional valence. This holds for website stimuli displayed for a longer duration, as well as for affective pictorial stimuli, even when such stimuli are displayed for a very brief duration. This makes facial EMG measurements of the corrugator supercilii muscle region a viable method for measuring affective responses to the visual appeal of website stimuli displayed for 50ms. However, corroborating these measurements with measurements of other components of emotion, such as behavior, and subjective feeling, remains essential in unearthing the reliability of facial EMG measurements in relation to visual appeal of websites.

2.6.3 Behavioral component: Eye-tracking

In most common HCI studies, where oftentimes usability is the main concept of interest, eye-tracking is used to assess which elements of an interface draw the most attention to them. This is important when, for instance, successful navigation of an interface depends on the user finding the appropriate button to click. Eye-tracking can reveal if the user saw the button or not, and might provide insight into interface elements that prevent the user from continuing. Other data that can be obtained using eye-tracking are the time spent looking at interface elements or regions, the time it takes to notice specific elements, and scan paths (Tullis & Albert, 2008).

Typical eye-tracking data consists of fixations and saccades. Fixations are points where an individual keeps his view focussed on a single spot to process information, while saccades are fast eye-movements between fixation points. Usually, fixations indicating focussed attention are defined as the eyes being stationary for 100-300ms (Djamasbi et al., 2010; Poole & Ball, 2005). It is vital to interpret these eye-tracking metrics in their context. For example, during regular website browsing, a higher fixation frequency on a particular area of the

website might indicate greater interest in that area. Conversely, during a search task, higher fixations on a particular area may indicate greater uncertainty about the target (Jacob & Karn, 2003). The duration of fixations is linked to the effort of processing the target. Longer fixations indicate more difficulty in extracting information (Poole & Ball, 2005).

Eye-tracking has also been used to assess specific visual elements of websites such as the influence of human faces (Cyr, Head, Larios, & Pan, 2009) and cross-cultural effects of color (Cyr, Head, & Larios, 2010). Furthermore, eye-tracking measurements have been used to study visual appeal of websites (Djamasbi et al., 2010). Djamasbi et al. (2010) found that websites with a large main image and a limited amount of textual information received the highest ratings. Eye-tracking confirmed that participants fixated on the main image when it was present in websites that were rated as highly appealing. Djamasbi's et al. (2010) experiment indicated that eye-tracking can be used to identify visual elements that draw the most attention during longer stimulus exposures (e.g. 10s).

Furthermore, research has shown that during picture viewing, people tend to focus more on pictures that are either positively or negatively valenced, as opposed to emotionally neutral pictures (Nummenaa et al., 2006). Studies comparing emotional positive and negative stimuli did not find a difference in attention between positive or negative images (Calvo & Lang, 2004, 2005; Lang, Greenwald, Bradley, and Hamm, 1993). This indicates an attention bias towards emotional stimuli in general. Eye-fixation metrics, such as number of fixations and fixation duration, may therefore in themselves be an indication of an affective response.

2.6.4 Subjective feeling component: Self-report

The most common way to ascertain someone's affective state is by using self-report measurements (Brave & Nass, 2008; Robinson & Clore, 2002). Self-report is considered a convenient and low-cost method that is capable of reliably capturing the subjective feeling component of emotion. Both fixed-choice scales as well as free response formats can be used to assess the conscious, subjective experience of emotions (Scherer, 2005).

Emotional self-report is not without its flaws however. Like any self-report method, the potential tendency of participants to respond in a socially desirable way cannot be eliminated easily.

Apart from this type of response bias, fixed-choice self-report of emotion may prime respondents to select options they may not have thought of otherwise. These responses may not necessarily represent the emotions that are actually experienced by the respondents. Conversely, self-report scales of emotion may lack certain response options. This forces the respondents to select the next-best option (Scherer, 2005). Moreover, this next-best option

may be interpreted differently by two separate respondents who actually experienced the same emotion.

Another issue with emotional self-report in general is that results are sometimes hard to compare cross-culturally, because of the differences that exist between cultures in the verbal expression of emotion (Desmet, 2002). This limitation has been addressed by the development of non-verbal self-report measurement instruments of emotion. The most well known, and most widely applied, is the Self Assessment Manikin (SAM)(Bradley & Lang, 1994). SAM uses a graphical figure to represent valence (from frowning to smiling), arousal (from excited to sleepy), and dominance (from small to large). The dimension of dominance is often dropped when SAM is used, because it has been found to lack discriminative power, and is highly correlated with the valence scale (Bradley & Lang, 2007).

Another non-verbal self-report instrument that has gained recognition is PrEmo (Product Emotion measurement instrument) (Desmet, 2002). PrEmo uses a set of 14 animations of a cartoon figure that combines movement of the face and body, as well as vocal cues, to depict discrete emotions. The PrEmo character displays seven positive and seven negative emotions. PrEmo is largely based on the idea that people are capable of identifying facial and bodily signals of emotion (Ekman, 1994; Ekman & Friesen, 1978; Wallbott, 1998).

An additional issue with self-report of subjective emotional feelings is that most methods are post-test measurements. Barrett (2004) argues that the time that passes between the stimulus and the reported emotion may negatively influence the accuracy of the measurement. The longer the time between the stimulus presentation and the self-report of emotion, the more the respondent will rely on memory to report his or her subjective feelings (Barrett, 2004). This is true for both verbal and non-verbal self-reports of emotion.

Attempts have been made to reduce the influence of this 'lag' between stimulus presentation and the self-report measurement. With this in mind, the LEMtool (Layered Emotion Measurement Tool) was developed (Huisman, 2009; Huisman & Van Hout, 2010).

2.6.5 Subjective feeling component: The LEMtool

The LEMtool is a non-verbal self-report instrument, that was specifically designed to measure the subjective feeling component of emotion during interaction with a visual interface (Huisman, 2009). The instrument uses a cartoon figure that expresses eight caricatured emotional facial expressions and body postures, to signal discrete emotions. In total the instrument consists of four positive and four negative emotions (see Figure 4). Like PrEmo (Desmet, 2002) the LEMtool is based on the notion that people can identify specific facial and bodily expressions of emotion (Huisman, 2009).

During interaction with a visual interface, the user can freely attach any number and combination of the eight emotions to specific parts of the interface, in order to indicate experienced emotions (see Figure 4).

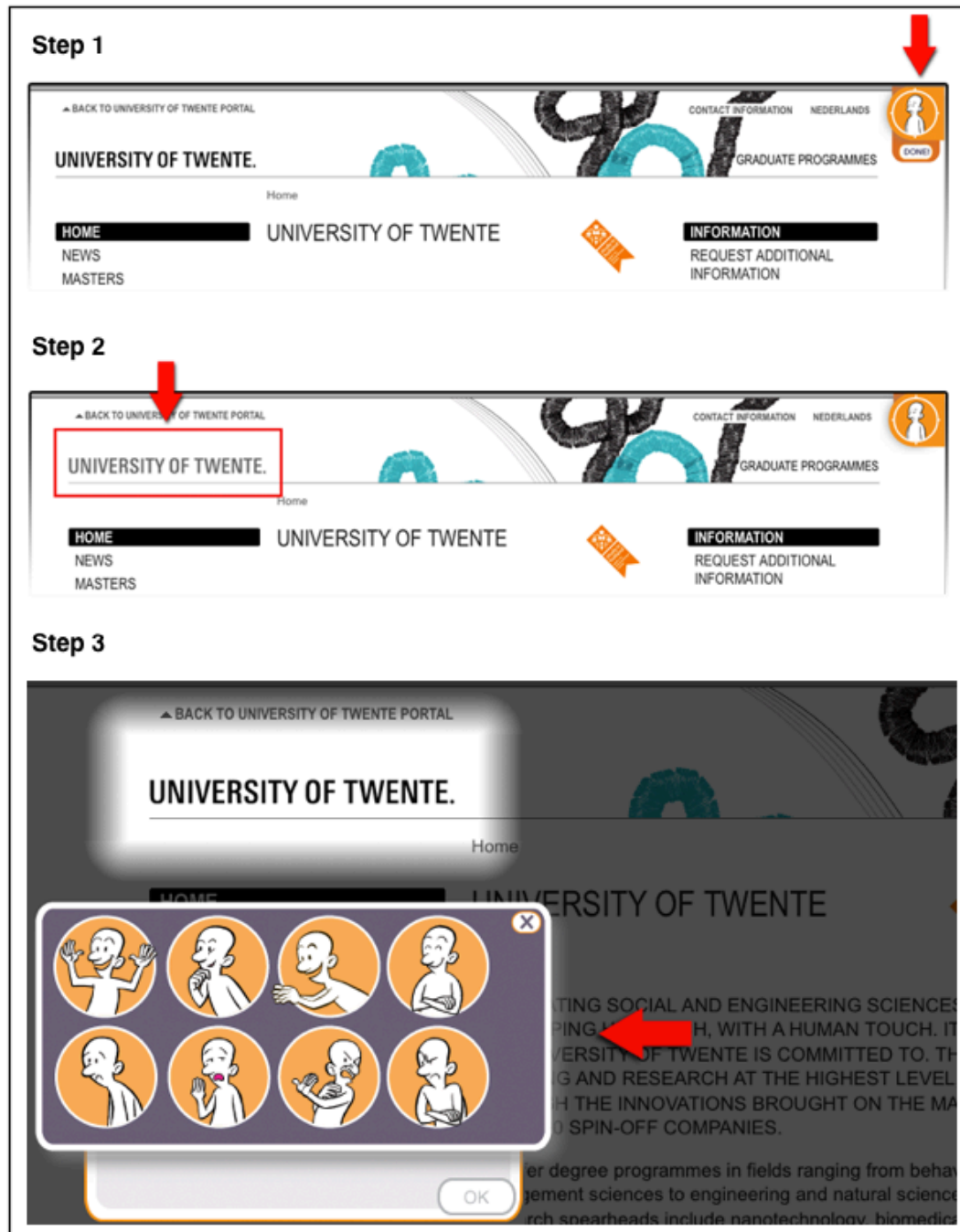


Figure 4 Functioning of the LEMtool. Step 1: activate. Step 2: select an area that elicited an emotion. Step 3: select one of the caricatures. The four figures in the upper row depict the positive emotions joy, interest, desire, and satisfaction. The bottom row depicts the negative emotions sadness, boredom, disgust, and dissatisfaction. N.B. the red arrows were added for clarification and do not appear in the LEMtool interface.

As a first validation of the LEMtool, the recognizability of the images as specific emotions was assessed. The validation experiments showed that the images were recognizable as the emotions they were intended to display, with a recognition accuracy between 80 and 100 percent. All emotions proved to be 99 to 100 percent accurate for valence, meaning positive emotions were recognized as positive and negative emotions were recognized as negative (Huisman, 2009).

Because the LEMtool is used ‘in-process’ (i.e. *during* interaction) it allows respondents to immediately report their experienced subjective feelings in relation to the interaction with a system. It could be argued that the fact that the LEMtool is deployed during interaction influences the subsequent experience of emotions. However, this has been considered an acceptable limitation in order to obtain in-process self-reports of emotion (Huisman, 2009). Moreover, the LEMtool was designed to minimize distraction, by only being present as a small icon in the top-right corner of the screen during browsing. Additionally, an acclimatization period helps participants become familiar with the LEMtool, thus minimizing its influence during the actual experimental interaction.

Preliminary applied studies have shown that users are, at the very least, able to express themselves using the LEMtool (see also Guiza Caicedo, 2009). This first indication of the working principle of the LEMtool, combined with the accuracy of the images for valence, makes it a viable self-report instrument to measure affective responses to visual appeal of websites. The intended functioning of the LEMtool, where participants select specific areas of a website, shows similarities with eye-tracking research, where participants eye-movements and fixation points are registered. Eye-tracking could prove a useful measurement technique for corroborating LEMtool measurements.

2.7 Summary

The current chapter outlined a theoretical framework, detailing the concepts of visual appeal, emotion and affect. The relation between visual appeal and affect was explained by the categorical-motivation model (Whitfield, 2005). Furthermore, the most viable measurement techniques for measuring the motor expression component of emotion (i.e. corrugator EMG), the behavioral component of emotion (i.e. eye-tracking), and the subjective feeling component of emotion (i.e. visual appeal rating scales and LEMtool) were presented. The next chapter will use the here presented theoretical framework to formulate hypotheses that serve to guide the experiments.

3

Hypotheses

In Chapter 1 two general research questions were formulated. These research questions provided the outline for the theoretical framework presented in the second chapter. In order to be able to provide a detailed answer to both general research questions, and in order to channel the design and analysis of the experiment, several hypotheses will be presented in this chapter. These hypotheses are based on the theoretical framework outlined in the previous chapter.

3.1 Motor expression component

In the current study the motor expression component of emotion will be measured using facial EMG of the corrugator supercilii muscle region. Studies have shown corrugator EMG activity to increase when negative affective stimuli are displayed (Codispoti et al., 2001; Codispoti et al., 2009; Smith et al., 2006). This effect still occurs for stimuli presented for just 25ms (Codispoti et al., 2009).

Furthermore, it has been suggested that the judgement of visual appeal of websites is related to affect, and takes place within 50ms (Lindgaard et al., 2006; Tractinsky et al., 2006). Some empirical evidence that websites can indeed be classified as affective stimuli has been presented (Zhou & Fu, 2007). In Zhou and Fu's (2007) study, high visual appeal websites showed similarities to positive affective images, while low visual appeal websites showed similarities to negative affective pictures.

Corroborating these findings by measuring the motor expression component of emotion would strengthen the notion that judgements of visual appeal are based on affect. From this the following hypothesis can be formulated:

- H1 The motor expression component of emotion will be able to differentiate between high and low visual appeal website screenshots.

3.2 Behavioral component

Previous studies have demonstrated that people are capable of distinguishing between high and low visual appeal websites when only having seen these websites for a fraction of a second (Lindgaard et al., 2006; Tractinsky et al., 2006). Considering the behavioral component of emotion, this begs the question whether these judgements are a result of differing behavioral patterns for high visual appeal and low visual appeal websites. The behavioral component of emotion will be studied in the main experiment using eye-tracking.

If judgements of visual appeal are formed on the basis of very specific content items within stimuli, it would be expected that differences in eye-tracking data would emerge. If this is not the case, it is plausible that no differences in eye-tracking data between high and low visual appeal websites will be observed. This line of reasoning is based on the notion that during browsing, people tend to focus on objects of interest (Jacob & Karn, 2003). If more favorable judgements of visual appeal for high visual appeal websites would be a result of those websites containing more objects of interest, eye-tracking data should show more, and longer fixations for high visual appeal websites than for low visual appeal websites.

However, it is expected that the visual appeal of websites is not judged according to specific objects of interest, or lack of such objects of interest, but is instead judged in relation to the affective response it elicits. This would hold that no differences will be found in eye-tracking data between high and low visual appeal websites. Moreover, such a finding would be comparable to findings from eye-tracking research using static visual affective stimuli. Studies comparing pleasant and unpleasant emotional stimuli did not find a difference in the number of fixations and fixation duration between affective stimuli (Calvo & Lang, 2004, 2005; Lang et al., 1993; Nummenmaa et al., 2006). This leads to the following hypothesis:

- H2 The behavioral component of emotion reveals similar behavioral patterns for high and low visual appeal websites.

3.3 Subjective feeling component

3.3.1 Visual appeal rating scales

It has been indicated that the rating of visual appeal of website screenshots is highly consistent between different exposure times (Lindgaard et al., 2006; Tractinsky et al., 2006). High correlations were found between visual appeal judgements after a 50ms exposure, where little visual information can be processed and visual appeal judgements of the same stimuli after

500ms, where more visual information can be processed (Lindgaard et al., 2006). Moreover, ratings proved to be equally consistent between 500ms and 10s exposures, and between 500ms and free-viewing (i.e. no time limit) (Tractinsky et al., 2006). This indicates that visual appeal judgements materialize early on in perception, and are relatively stable. Therefore, a 50ms stimulus exposure time should be sufficient for individuals to distinguish between high visual appeal and low visual appeal websites. From this, the following hypotheses can be formulated:

- H3 Visual appeal ratings are able to differentiate between high and low visual appeal websites when these websites are displayed for 50ms.
- H4 There is a high correlation between visual appeal ratings in 50ms stimulus exposure conditions and free-viewing conditions.

If these visual appeal judgements after a 50ms stimulus exposure are indeed related to affect, a significant correlation between corrugator EMG and visual appeal ratings should be found. Though small, (i.e. -0.16) such correlations have already been found between corrugator EMG activity and SAM ratings during computer interaction (Mahlke & Minge, 2008). The fifth hypothesis is formulated as follows:

- H5 There is a significant correlation between visual appeal ratings and responses in the motor expression component of emotion.

3.3.2 *The LEMtool*

It has been found that the LEMtool images are all 99 to 100 percent accurate representations of the valence they were intended to communicate (Huisman, 2009). Considering that there is evidence that visually appealing and unappealing websites can be seen as affective stimuli (Zhou & Fu, 2007), the LEMtool should be capable of distinguishing between high visual appeal and low visual appeal websites when these websites are presented for 50ms. In addition, if judgements of visual appeal are indeed stable (Lindgaard et al., 2006; Tractinsky et al., 2006) when websites are viewed without a time-limit, the LEMtool should be able to distinguish between high visual appeal and low visual appeal websites, independent of stimulus exposure time. This results in the following hypothesis:

- H6 The LEMtool is able to differentiate on valence between high and low visual appeal websites independent of stimulus exposure time.

- H7 There is a significant relation between LEMtool indications and responses in the motor expression component of emotion.

3.5 Research model

The hypotheses presented in this chapter can be visualized into a research model. Figure 5 depicts the relation between the central concepts of the present study and indicates which relation each hypothesis seeks to clarify.

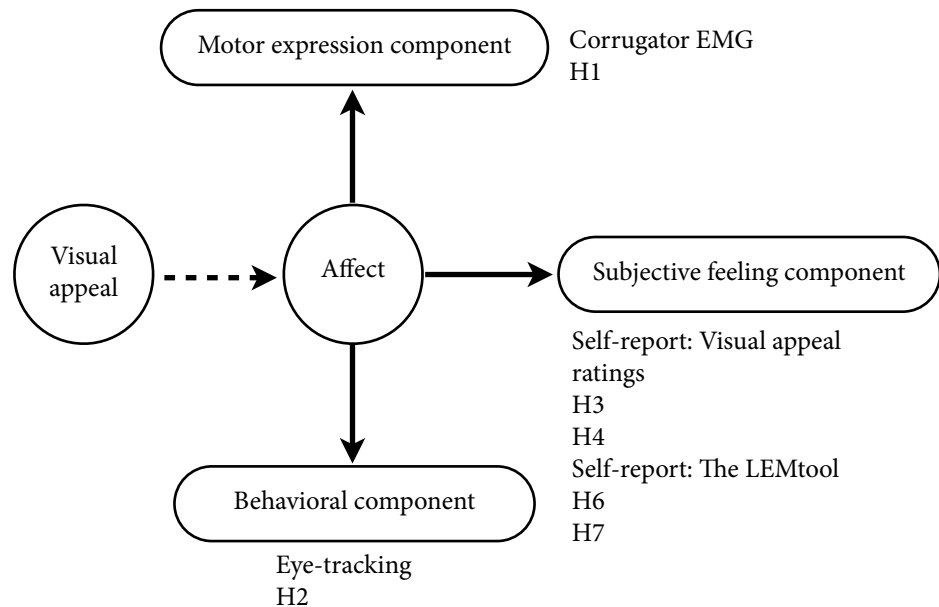


Figure 5 Research model depicting the relations between the central concepts in the present study. Hypotheses related to specific components of emotion are included.

The model in Figure 5 shows that visual appeal may elicit affect in the perceiver (Lindgaard et al., 2006; Zhou & Fu, 2007). Such an affective reaction has a number of components (Scherer, 2005), among which a motor expression component (i.e. facial muscle activity measurable through EMG), a behavioral component (i.e. direction of attention in the form of eye fixations), and a subjective feeling component (i.e. visual appeal ratings and LEMtool indications). The affective reaction can aid emotional categorization (Nummenmaa et al.,

2006; see also Whitfield, 2005) of the stimulus, eventually allowing the perceiver to reach a judgement of the visual appeal of the stimulus based on subjective feeling (Scherer, 2005).

The hypotheses and accompanying research model will be tested in an experiment (Chapter 5). Participants will be presented with high and low visual appeal websites. Varying stimulus exposure times will be used, and measurements of the motor expression, behavioral, and subjective feeling components of emotion will be taken.

First however, the following chapter will present the manipulation check that was carried out. This manipulation check served to verify that the website screenshots to be used in the main experiment, actually differed on visual appeal. Additionally, possible intervening variables were investigated.

4

Manipulation check

In the previous chapter, a number of hypotheses was formulated. These hypotheses served to structure the main experiment and, eventually, to answer the main research questions.

However, before presenting the main experiment, the current chapter discusses a manipulation check that served to validate the selected website screenshots. Before using the websites screenshots as stimuli in the main experiment, it was important to first determine that the stimulus set indeed consisted of an equal number of high visual appeal and low visual appeal websites. Furthermore, three possible intervening variables were investigated. These variables were: perceived usability, familiarity with the content of the website, and familiarity with the structure of the website.

Results of the manipulation check showed that high visual appeal websites received a significantly ($p < .001$) higher visual appeal rating than low visual appeal websites. For a limited number of websites a correlation was found between visual appeal and perceived usability.

4.1 Stimuli pre-selection

The stimulus set consisted of 24 website screenshots (Appendix A), of which twelve were high visual appeal websites and twelve were low visual appeal websites. The total number of 24 website screenshots were created in a previous study by Van Dongelen (2008). Two independent web designers created a high and a low visual appeal versions of the same website, following specific guidelines in order to keep the type, organization, and presentation of information on the website consistent. Also, special care was taken to keep the content and perceived functionality of the website the same for both high and low visual appeal websites (Van Dongelen, 2008, p.41).

These manipulations of the specific elements of each website were not considered further here. Note, that the goal of the current study was to investigate the relation between visual appeal judgements of websites and affect. The website stimuli were in essence treated as a ‘black-box’, that contains several unknown elements that contribute to a website being more or less visually appealing. This follows Gestalt theory in the sense that the website is viewed as

a whole, where the object is more than the sum of its parts (Djamasbi et al., 2010; Lavie & Tractinsky, 2004).

An added benefit of using the websites created by Van Dongelen (2008), was that they could not be considered ‘high traffic’ websites (such as amazon.com, cnn.com), limiting the potential influence of familiarity with the specific websites. Also, the websites covered a range of topics (i.e. Einstein, the isle of Rugen, and medical information about headaches), making results more generalizable.

In a pre-test Van Dongelen (2008) found that the high visual appeal version of each website was judged as significantly more visually appealing than the low appeal version. As Van Dongelen notes, this does not mean that the high visual appeal version can be considered ‘beautiful’ as such, but that it was merely judged by the participants as being more visually appealing than the low visual appeal version. This follows the definition of visual appeal used in the current investigation (see Section 2.1).

There were three reasons to carry out another manipulation check before considering the 24 websites for the main experiment. First, Van Dongelen (2008) used an exposure time of 750 milliseconds. The current manipulation check used a 500 millisecond exposure time in order to make results more comparable to Lindgaard’s et al. (2006) and Tractinsky’s et al. (2006) findings.

Second, conducting another manipulation check allowed to control for possible intervening variables that might influence visual appeal judgements. Because of the indicated relation between usability, perceived usability and visual appeal (Sonderegger & Sauer, 2010; Tractinsky et al., 2000) it was important to control for perceived usability as an intervening variable in the judgement of visual appeal of static screenshots. In addition, participants were asked to indicate how familiar they were with the presented website screenshots. Familiarity was separated into content familiarity (i.e. familiarity with texts, images, logo’s, brand’s etc.) and structure familiarity (i.e. familiarity with the websites layout, such as menu on the left, logo on top left, etc.). It was hypothesized that, similar to perceived usability, familiarity with the structure or content of the website could influence visual appeal ratings.

Third, conducting a second manipulation check made it possible to analyze the visual appeal ratings in more detail. Following Tractinsky et al. (2006), response latency (i.e. how long it took participants to give their judgement) was used as an alternative measure for the strength of preference for the website screenshots. Research suggests that the quicker a preference rating is given, the more certain an individual is of his or her judgement. Longer response latencies are an indication of increased mental effort to make the judgement (Tractinsky et al., 2006). Response latency was used as an objective measure of preference and mental effort, to obtain convergence of methods, and to allow the current stimuli set to be

compared on multiple levels to other stimulus sets used in previous studies (Tractinsky et al., 2006).

The goal of the manipulation check was to validate the difference in visual appeal between the high visual appeal websites and the low visual appeal websites. In addition, because the stimulus set consisted of a high visual appeal and a low visual appeal version of each website, detailed comparisons could be made between the high and low visual appeal version of each website. The manipulation check was considered successful if each high visual appeal version of a website received a significantly ($p < .01$) higher visual appeal rating than its low visual appeal counterpart. In addition, a limited number of correlations were expected between visual appeal ratings and perceived usability (Tractinsky et al., 2000). Furthermore, no correlations were expected between visual appeal ratings, and content and structure familiarity.

4.2 Method

4.2.1 Participants

In total 31 participants rated the 24 webpages for visual appeal. Of these 31 participants, 22 (10 male and 12 female) also completed the second phase of the pretest, where they were asked to rate perceived usability, content familiarity and structure familiarity. It was gathered from participants that completing the ratings for visual appeal and the three other constructs separately, with the stimuli being presented twice, proved strenuous. However, separation of the ratings was deemed necessary in order to obtain a reliable rating of visual appeal, uninfluenced by the ratings on the other constructs.

The age of the participants that completed both phases, ranged from 20 to 51 (Mean age 28,5). These individuals spend between 4 and 90 hours (average of 32,5 hours) a week on the Internet. The large difference in time spend online, can be explained by the fact that some might have interpreted their computer being turned on and connected to the Internet as 'being online'. Still, this data indicates that a range of Internet users, from individuals who are online occasionally, to individuals who are online often, participated in the experiment. All participants were approached through social-media networks (i.e. Twitter, Facebook, and LinkedIn).

4.2.2 Apparatus

A specific website was developed in order to present stimuli with the appropriate exposure time, present the rating-scales to participants, and record the data. This website was compatible with all popular browsers (e.g. Internet Explorer, Firefox, Chrome). The stimuli were presented using Adobe Flash. Participants used a standard computer mouse and keyboard to interact with the stimulus presentation website.

4.2.3 Materials

The 24 website screenshots, twelve designed as high visual appeal websites and twelve designed as low visual appeal websites (Van Dongelen, 2008), were used. These 24 websites included eight websites about Einstein, eight about the isle of Rugen, and eight containing medical information about headaches. Website screenshots were presented in a resolution of 1024x768 pixels. No browser elements were displayed (Appendix A).

In addition, five general websites (e.g. Amazon.com, CNN.com) were selected to serve as test websites to familiarize the participants with the procedures. These five websites were displayed in a resolution of 1024x768 pixels with no visible browser elements.

4.2.4 Procedures

All relevant instructions for the successful completion of the experiment were given through the same website that was used to present the stimuli.

Participants were first presented with a brief introduction text, explaining the general aim of the experiment. This was followed by another instruction page detailing the procedures and necessary actions for the participants to take in order to complete the ratings. This introduction was followed by a test session in which the five general websites were presented. It was stressed that the test session would not be analyzed in the results and that its sole purpose was to familiarize the participants with the brief stimulus exposure times and the response format.

After participants completed rating the five test websites, another explanation of the experimental procedure was given. It was made clear that the actual experiment was about to begin. All 24 website screenshots (12 high visual appeal and 12 low visual appeal) were presented to participants in random order. Before each screenshot would be displayed, participants had to press a 'start' button, that, when pressed, was followed by a white screen presented for 1s, after which the website screenshot was displayed automatically for 500ms.

Directly after stimulus exposure, participants were shown a screen with a ten-point visual appeal rating scale, and were asked to give a visual appeal rating. Clicking on a 'next' button presented the start button for the next stimulus. This procedure was followed for the test websites and was repeated for all 24 stimulus websites.

After participants judged all 24 website screenshots for visual appeal, the same selection of screenshots were presented in random order a second time. Participants were instructed to indicate the level of perceived usability, content familiarity and structure familiarity on a ten-point-scale for each website screenshot (see Appendix B). These three rating scales were presented separately from the visual appeal rating scale in order to ensure that they would not influence the visual appeal ratings.

After having rated all 24 website screenshots twice, once for visual appeal, and once for usability, content familiarity and structure familiarity, the experiment concluded by asking participants to indicate their age, gender, and weekly Internet usage.

4.3 Results

4.3.1 Visual appeal ratings

First, an independent samples *t*-test was calculated, comparing the visual appeal ratings for high visual appeal and low visual appeal websites. Overall, high visual appeal websites received a mean rating of 5,39 ($SD = 2,20$) while low visual appeal websites received a mean rating of 3,08 ($SD = 1,77$) on a scale of 1 to 10, where 1 was 'visually unappealing' and 10 was 'visually appealing'. This difference was significant ($t(742)=15,8, p < .001$).

To test if the high appeal version of each specific website would be judged by participants as being significantly more visually appealing than the low appeal version of this specific website, a paired samples *t*-test was calculated (Table 2, Table 3, and Table 4).

Only the 'Albert Einstein' high visual appeal website received a mean rating higher than the mid-point of the scale (i.e. 5.5). Note that, following the definition of visual appeal presented in Section 2.1, the goal here was not to determine the visual attractiveness of the websites in themselves, but to determine if there was enough of a difference in the rating of visual appeal between the high and low visual appeal version of each website.

Table 2

Differences in visual appeal ratings for the high and low visual appeal Einstein websites (scale 1-10, 1=visually unappealing and 10=visually appealing)

Website	Version	Mean rating	Standard deviation	<i>t</i> -value
Formative years	High visual appeal	4,81	1,68	$t(30)=4,93^{**}$
	Low visual appeal	3,06	1,55	
Imagination	High visual appeal	4,77	1,73	$t(30)=4,49^{**}$
	Low visual appeal	3,55	1,41	
Early years	High visual appeal	4,68	1,92	$t(30)=7,68^{**}$
	Low visual appeal	1,74	1,21	
Albert Einstein	High visual appeal	6,55	2,05	$t(30)=10,51^{**}$
	Low visual appeal	2,42	1,36	

$^{**}p < .001$

As for the Einstein websites, the high visual appeal versions of the Isle of Rügen websites were judged to be more visually appealing than the low visual appeal versions (Table 3). For three of the four websites the difference between the high and low visual appeal version was significant beyond $p < .001$ significance. For the 'Largest' website the difference between the high and low visual appeal version was significant at $p < .01$ level of significance.

Table 3 reveals that, contrary to the Einstein websites, three of the four high visual appeal Isle of Rügen websites received visual appeal ratings that were higher than the mid-point of the scale. Only the high visual appeal version of the 'Largest' website was rated lower than 5.5.

Table 3

Differences in visual appeal ratings for the high and low visual appeal Isle of Rügen websites (scale 1-10, 1=visually unappealing and 10=very appealing)

Website	Version	Mean rating	Standard deviation	<i>t</i> -value
Einfach	High visual appeal	5,74	2,39	<i>t</i> (30)=10,51**
	Low visual appeal	3,90	2,44	
100% German	High visual appeal	7,35	1,98	<i>t</i> (30)=4,91**
	Low visual appeal	3,48	2,00	
Isle	High visual appeal	6,87	1,88	<i>t</i> (30)=11,23**
	Low visual appeal	3,06	1,55	
Largest	High visual appeal	4,90	1,87	<i>t</i> (30)=3,12*
	Low visual appeal	3,58	1,82	

* $p < .01$

** $p < .001$

Table 4 shows that the high visual appeal version of three of the four headache websites were rated as more visually appeal than their low visual appeal counterparts. These differences proved to be significant beyond $p < .001$ level of significance.

For the ‘Medline’ website, the low visual appeal version received a slightly higher rating than the high visual appeal version. This difference was not significant. Previous results comparing both versions of the Medline website did find the high visual appeal version to receive significantly higher visual appeal ratings than the low visual appeal version (Van Dongelen, 2008). However, the difference between the average ratings for the high and low visual appeal version was slight (.79), in fact it was the smallest difference out of all websites tested. Corroborated by the findings of the current manipulation check this indicates that the Medline websites are minimally distinctive on visual appeal.

Table 4

Differences in visual appeal ratings for the high and low visual appeal Headaches websites (scale 1-10, 1=visually unappealing and 10=visually appealing)

Website	Version	Mean rating	Standard deviation	<i>t</i> -value
Medline	High visual appeal	3,58	1,69	$t(30)=-.12$
	Low visual appeal	3,61	1,59	
Illustrated	High visual appeal	5,77	2,08	$t(30)=7,42^{**}$
	Low visual appeal	2,65	1,62	
Health	High visual appeal	5,45	2,17	$t(30)=6,84^{**}$
	Low visual appeal	3,48	1,75	
Netdoctor	High visual appeal	4,19	1,78	$t(30)=5,23^{**}$
	Low visual appeal	2,39	1,54	

* $p < .01$

** $p < .001$

However, for the sake of keeping the current stimuli set balanced (i.e. an identical number of websites for each topic) the Medline website was not excluded from the set. Special care would be taken in analyzing the results from the main experiment for the high, and low visual appeal version of the Medline website.

4.3.2 Perceived usability, content familiarity and structure familiarity

In the second phase of the manipulation check, participants were asked to indicate perceived usability, familiarity with the content, and familiarity with the structure of the same set of websites as in the first part. These items were included to assess whether visual appeal judgements of the stimulus set were related to non-affective judgements. These three constructs, perceived usability, structure familiarity, and content familiarity, can be described as being more cognition driven than the construct of visual appeal. For instance, perceived usability would require making inferences about usability based on visual elements, whereas content, and structure familiarity would require retrieval of information from memory and relating this information to the perception of visual elements from the screenshots.

Table 5

Pearson-correlations between ratings of visual appeal and perceived usability

Website	Version	Correlation (r)
Early years	Low visual appeal	0,54**
Einfach	High visual appeal	0,68***
Einfach	Low visual appeal	0,45*
100% German	High visual appeal	0,52*
Largest	Low visual appeal	0,64***
Illustrated	High visual appeal	0,63**
Illustrated	Low visual appeal	0,64***
Health	High visual appeal	0,48*
Health	Low visual appeal	0,66***

* $p < .05$ ** $p < .01$ *** $p < .001$

For 9 of the 24 websites a significant correlation was found between perceived usability and visual appeal (Table 5). The correlation coefficients (r) ranged from .45 to .68. These findings are in line with previous research that found strong correlations between visual appeal and perceived usability (Tractinsky et al., 2000). As suggested by Lindgaard et al. (2006), the 500ms exposure time may still be long enough for participants to make inferences about the usability of a website. However, contrary to what might be expected based on Tractinsky's et al. (2000) findings, significant correlations between visual appeal and perceived usability were found for less than half of the website screenshots in the current manipulation check. This shows that, while perceived usability may be correlated with visual appeal judgements in 500ms stimulus exposure conditions, this relation was limited to a subset of stimuli. Therefore, the main factor that differentiates between the screenshots in the stimulus set, is visual appeal (see also Appendix C).

Unlike perceived usability, content and structure familiarity were found to be very weakly related to the visual appeal judgments. No significant correlations were found for structure familiarity, indicating participants were unfamiliar with the structural layout of the stimuli. Furthermore, for only 2 of the 24 websites a significant correlation was found between content familiarity and visual appeal. However, both these correlations fell below $r = .50$. Thus, participants did not seem to be familiar with the content of the vast majority website screenshots.

Out of the three intervening variables investigated, perceived usability, as expected, was the only one to have a considerable number of significant correlations with visual appeal ratings. While this did not warrant dismissal of the stimulus set, it does point to a potential intervening influence of cognitive processes in judging the visual appeal of website screenshots displayed for 500ms. Controlling for the influence of cognitive constructs in the main experiment is therefore necessary.

A more detailed analysis of the cognitive constructs was carried out. The results of this analysis can be found in Appendix C.

4.3.3 Response latency

Following Tractinsky et al. (2006), response latency was used as an unobtrusive measurement in an attempt to increase the validity of the manipulation check. Response latency was operationalized as the time between the appearance of the visual appeal rating scale on the screen, and the moment the participant confirmed the rating by clicking a button.

In total 744 (31 respondents x 24 webpages) response latencies were obtained. Following Tractinsky et al. (2006) the relation between response extremity of visual appeal ratings and response latency¹ was examined. Visual appeal ratings were treated as random factors with 5 levels based on their distance from the mid-point of the scale. For instance, ratings 5 and 6 – the scale's mid-points – were grouped under Category 1. Ratings 1 and 10 – the most extreme ratings – belonged to Category 5.

Different from Tractinsky's et al. (2006) study, the current study was carried out online. This made it impossible to control for potential interventions (e.g. notifications of any kind on one's home computer) during the rating task. These interventions may have distorted the response latency's. For this reason it was decided to exclude response latencies that were two standard deviations or more above the mean response latency, from further analysis. This way, nine ratings were excluded.

Figure 6 depicts the mean response latencies per rating category. What stands out is that the less extreme the rating (category 1), the longer the response latency. To test these differences for significance, a one-way ANOVA was performed with rating category as the independent variable and response latency as the dependent variable. There was a significant effect of rating extremity on response latency ($F(4, 734) = 17,09, p < .001$).

¹ Logarithmic transformation was used to reduce the skewness of the response latency data.

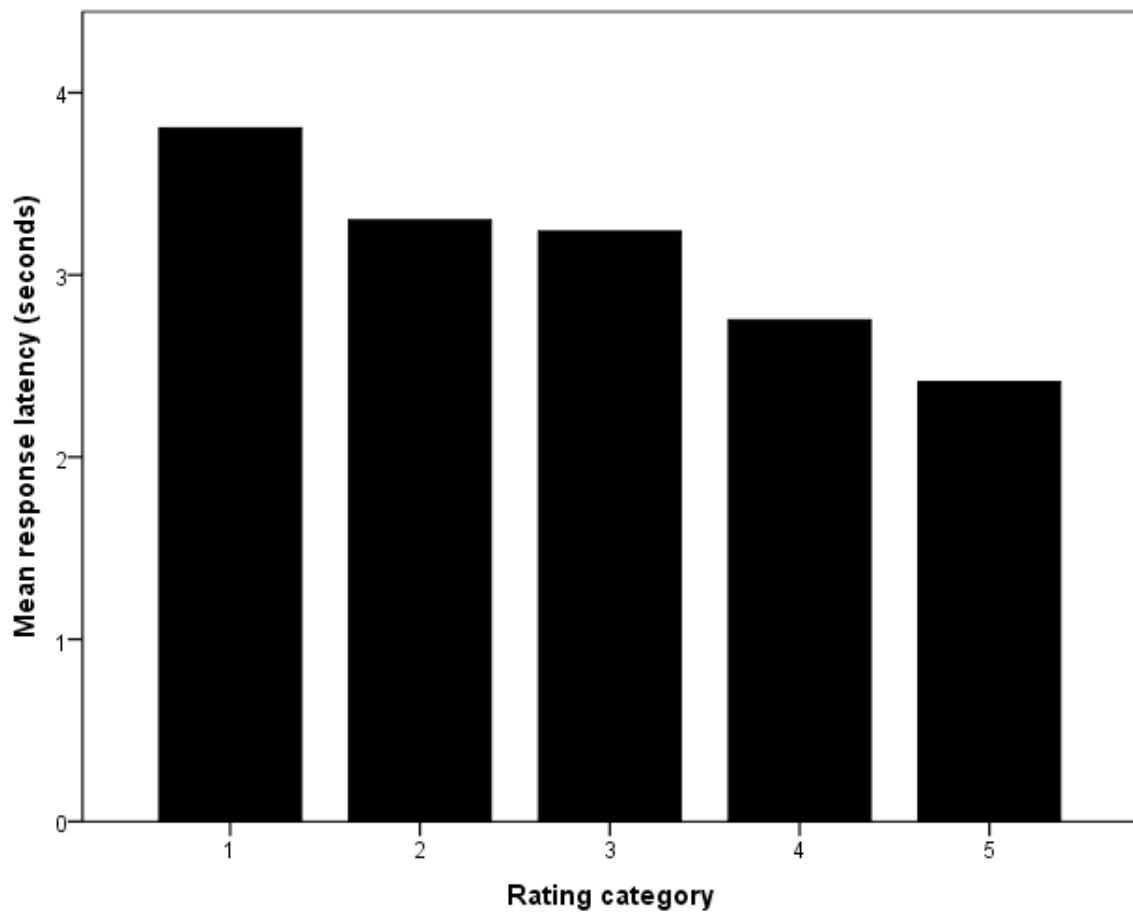


Figure 6 Mean Response Latency per rating category (Category 1 = ratings 5 and 6; Category 2 = ratings 4 and 7; Category 3 = ratings 3 and 8; Category 4 = ratings 2 and 9; Category 5 = ratings 1 and 10)

Table 6 shows the post-hoc comparisons between the five rating categories. The results support the notion that the more extreme the rating the less time is needed for a respondent to give a rating of visual appeal.

The results in Table 6 indicate that especially when extreme ratings were given, participants gave these ratings very quickly after the appearance of the rating scales. This would seem to indicate that when giving extreme ratings, less mental effort was required to make a judgement.

Table 6

Post-hoc (Tukey's HSD) comparison between response latencies of webpage visual appeal ratings

Rating Category	1 (<i>n</i> = 171)	2 (<i>n</i> = 171)	3 (<i>n</i> = 174)	4 (<i>n</i> = 124)	5 (<i>n</i> = 104)
1 (ratings 5 and 6)	-	ns	ns	$p < .001$	$p < .001$
2 (ratings 4 and 7)		-	ns	$p < .01$	$p < .001$
3 (ratings 3 and 8)			-	$p < .05$	$p < .001$
4 (ratings 2 and 9)				-	$p < .05$
5 (ratings 1 and 10)					-

In addition to general rating extremity, the difference in response latencies between extremely positive (9 or 10) and extremely negative (1 or 2) ratings was investigated. The ratings of 9 and 10 were collapsed into one category of extreme positive ratings (a total of 28 ratings) and ratings of 1 and 2 into a category of extreme negative ratings (a total of 199 ratings). Similar to Tractinsky et al. (2006) a difference between response latencies of extremely positive and extremely negative ratings was found. Latencies of positive ratings (mean = 2.89, $SD = 1.42$) were longer than latencies of negative rating (mean = 2.60, $SD = 1.04$). However, this difference was not significant ($F(1, 23) = 1.79, p = 0.18$).

The response latency results obtained in the current experiment are highly comparable to those obtained by Tractinsky et al. (2006). The only exception was that in the current experiment no significant difference was found between extremely positive and extremely negative ratings.

4.3.4 Conclusions

The manipulation check demonstrated that, overall, the high visual appeal websites were rated as significantly more visually appealing than the low visual appeal websites. Moreover, the high visual appeal counterpart of each specific website was rated as significantly more visually appealing than its low visual appeal counterpart. The only exception to this were the Medline websites, where the low visual appeal website was rated as more visually appealing than the high visual appeal website.

Results also revealed that out of the three cognitive constructs, only ratings for perceived usability showed noticeable correlation with the visual appeal rating. Some relation between visual appeal and perceived usability was to be expected (Tractinsky et al., 2000), however, in the present manipulation check, the relation between visual appeal and perceived

usability was found for less than half (9 out of 24) of the website screenshots. Thus this relation did not warrant elimination of the stimulus set.

Additionally, response latency data revealed that the website stimuli in the pretest were judged in a similar fashion as high and low visual appeal websites in comparable studies (Tractinsky et al., 2006).

All in all, the results of the manipulation check showed that the website screenshots differed from each other on visual appeal. Moreover, out of the intervening variables investigated, only perceived usability showed some correlation to visual appeal for a subset of stimuli. Finally, the stimulus set shows great similarity to stimuli used in previous experiments on visual appeal (Lindgaard et al., 2006; Tractinsky et al., 2006). It was therefore decided to use all 24 websites as stimuli in the main experiment. The results of the main experiment will be presented in the next chapter.

5

Main experiment

The manipulation check presented in the previous chapter demonstrated that the 24 website screenshots differed on visual appeal. In total twelve high visual appeal and twelve low visual appeal websites were tested. These websites were used in the main experiment, which will be discussed in this chapter.

The main experiment consisted of three separate phases. In the first phase EMG measurements of the corrugator supercilii muscle region were taken. Participants viewed a selection of twelve randomly presented website screenshots displayed for 50ms. They were asked to rate each website for visual appeal and to indicate a LEMtool emotion. During the second phase participants were presented with twelve different website screenshots displayed for 1s while eye-tracking was used to measure fixations. In the third phase the same twelve websites as in the second phase were used. There was no time-limit for viewing the websites. Participants were instructed to press a button to give a visual appeal rating for each website. Upon giving a rating, the LEMtool was presented and participants were instructed to indicate areas of each website that elicited a certain feeling.

Results of the main experiment revealed that the motor expression component of emotion differentiated between high and low visual appeal websites. The behavioral component of emotion revealed that there was no difference in general behavioral patterns between high and low visual appeal websites. Last, self-report methods also differentiated between high and low visual appeal websites.

5.1 Method

5.1.1 Participants

In total 43 (13 male, 30 female) individuals participated in the experiment. Twenty-four of these participants were first and second year psychology and communication science students who received course credits for participation. The remainder of the participants were approached by the researcher and asked to volunteer in the study. Participants age ranged from 18 to 31 (Mean age 22,4). Participants with colorblindness, and participants wearing

glasses were excluded from the experiment. This last restriction was due to limitations of the eye-tracker hardware.

5.1.2 Apparatus

In the first phase of the experiment, stimuli were presented using E-Prime 2.0 displayed on a 17 inch Samsung SyncMaster 750s CRT monitor (Samsung Electronics, Seoul, South Korea). The monitor was set to a screen resolution of 1024x768 pixels at 60 Hz. Brightness was set to 85 and contrast to 100 with color temperature set to 9300°K. Participants used a standard keyboard to indicate their responses.

Corrugator EMG activity was recorded using a ProComp Infiniti amplifier (14-bit resolution) combined with a MyoScan-Z sensor/pre-amplifier (Thought Technology, Montreal, Canada). A voltage isolator was installed between the stimulus computer and the ProComp Infiniti amplifier to allow timestamps to be added to the EMG data stream via the stimulus computers' parallel port. The ProComp Infiniti unit was connected to a data recording laptop via an optical-to-USB converter (Figure 7).

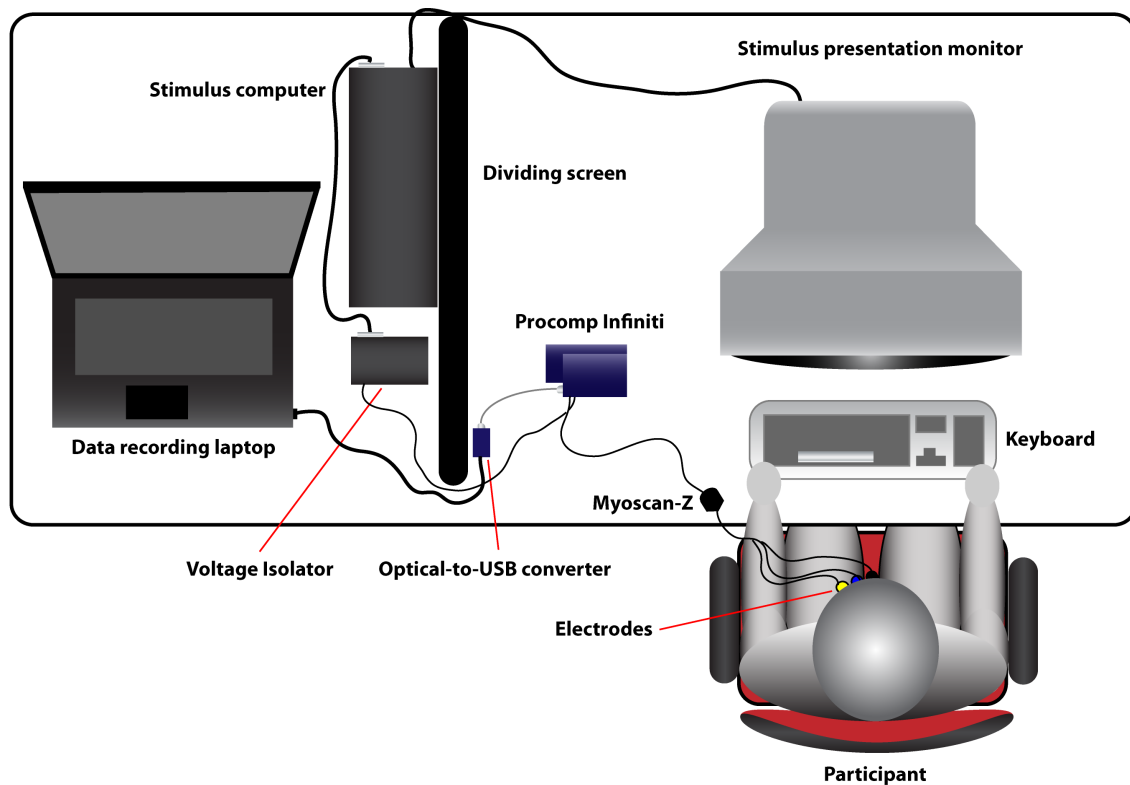


Figure 7 Detailed schematic of the set-up in the first phase of the experiment.

Two self-adhesive, 24mm pre-gelled disposable silver-silver chloride (Ag-AgCl) electrodes were placed on the corrugator supercillii muscle region above the left eye of the participant, with an identical third ground electrode placed on the centre of the forehead (Fridlund & Cacioppo, 1986). Wiring was guided behind the participants left ear to ensure no wires would obstruct the participants view.

FaceLAB version 4.5 (Seeing Machines, Canberra, Australia) was used to record eye-movements and fixations in the second phase of the experiment. Stimuli were presented using GazeTracker (Eye Response Technologies Inc., Charlottesville, VA) displayed on a HP Compaq LE1711 17 inch LCD monitor (Hewlett-Packard, Palo Alto, California). The monitor's native resolution of 1280x1024 pixels at 60 Hz was used. The monitor was set to a brightness level of 90 and contrast to 80 with color temperature set to 6500°K.

In the third phase of the experiment, stimuli were presented using a purpose build online environment running in Firefox 3.6.3.

5.1.3 Materials

The same 24 websites as in the manipulation check, consisting of twelve high visual appeal and twelve low visual appeal websites, were used. The websites included eight websites about Einstein, eight about the isle of Rugen, and eight containing medical information about headaches (Van Dongen, 2008). Website screenshots were presented in a resolution of 1024x768 pixels. No browser elements were displayed. For the first phase of the experiment a selection of 12 websites, consisting of six high visual appeal and six low visual appeal websites (Set A) was used. The remaining 12 websites (Set B) were used in the second, and third phase. Each set contained an equal number of high and low visual appeal websites from each topic. The order of Set A and Set B was switched for each participant, so that odd numbered participants would see Set A in the first phase and Set B in the second and third phase, and vice versa for even numbered participants (Appendix A).

In addition, five general websites (e.g. Amazon.com, CNN.com) were selected to serve as test websites. These test websites were displayed to familiarize the participants with the very brief exposure times and methods of rating. The five test websites were displayed in a resolution of 1024x768 pixels with no visible browser elements.

5.1.4 Procedures

Upon entering, participants were greeted by the experimenter. They were given a written explanation of the general procedures of the experiment, including the use of EMG equipment

and the required attachment of electrodes to the participants face. Participant's were told in the written explanation that the EMG equipment would be used to control for blinking during the experiment. It was made evident that this was important due to the brief exposure time of the stimuli. After reading the explanation participants were instructed to read and sign an informed consent sheet.

The experiment consisted of three separate phases, which in total took approximately 45 minutes to complete. In the first phase, participants were seated in front of the CRT monitor and the EMG electrodes were attached in accordance with the placement guidelines proposed by Fridlund and Cacioppo (1986) (Figure 8).



Figure 8 Participant with EMG electrodes attached to the corrugator supercilii muscle region, and a ground electrode attached to the forehead.

After the electrodes were attached, participants were told by the experimenter to sit still and relax, while looking at the blank monitor for two minutes. This was done in order to check the signal quality and establish a baseline. Following the relaxation period, the experiment would continue automatically with instructions for the rest of the first phase of the experiment.

After the instructions, participants were presented with the five test websites displayed in random order. When participants pressed the spacebar, a white screen was presented for 1s,

after which the test website was displayed in full-screen for 50ms. The presentation of each test website was followed by a white screen displayed for 4s, after which a ten-point rating scale appeared. Once participants indicated the perceived level of visual appeal using the keys 1 to 0 on the keyboard representing a rating of 1 (visually unappealing) to 10 (visually appealing), the eight LEMtool images appeared. Each image corresponded to a numbered key (1 to 8) on the keyboard. By pressing a single key corresponding to a single LEMtool image, participants indicated the feeling that the website elicited in them. Participants could use the 0-key to indicate that the website did not elicit any feeling.

Once the participants rated all five test websites the program would indicate the actual experiment was about to begin. A selection of twelve websites (Set A for odd numbered participants and Set B for even numbered participants) was subsequently presented in random order to the participants. The procedure was identical to that of the test websites.

The first phase of the experiment was concluded when participants rated all twelve websites. The electrodes were removed from the participant's face and participants were given the opportunity to clean off any conductive gel residue.

The second phase of the experiment required participants to take place behind the eye-tracker, situated in the same room. Once the eye-tracker was calibrated, participants would be shown twelve websites, different from those used in the first phase of the experiment, displayed in random order. Each website would be displayed for 1s. Preceding each website a black targeting cross was displayed in the centre of the screen. Participants were instructed to focus on the centre of this cross each time it appeared on the screen. Presentation of the website screenshots and the targeting cross was automatic, requiring no action by the participant. After each of the twelve websites was presented, the second phase of the experiment was concluded.

During the third phase of the experiment the same twelve websites as in the second phase were presented to participants. Instructions for the use of the LEMtool were given in the online environment used to present the stimuli. Participants were again presented with the same five test websites displayed in random order. Each website would stay on the screen until participants pressed a button 1 to 0 representing a rating of 1 (visually unappealing) to 10 (visually appealing). After giving a visual appeal rating, the LEMtool would appear in the top-right corner of the screen and participants had to select areas of the website using the computer mouse, and attach a LEMtool image to that area. Participants could give as many LEMtool indications as they liked, but were instructed to only rate elements that were related to the visual appeal of the website. They were told reading texts on the website was not required.

Once participants had rated the five test websites, the twelve stimulus websites were displayed in random order. The exact same procedure as with the test websites was followed. The third phase was concluded after participants rated all 12 websites using the 10-point rating scale and LEMtool and indicated their age, gender, and native language.

5.1.5 Data reduction

For the EMG measurements, the raw EMG signal was digitized at 2048 Hz. A bandpass filter of 20-500 Hz was applied. The raw EMG signal was rectified and integrated using a 100ms time constant. The rectified and integrated EMG signal was then re-sampled at 256Hz.

For each stimulus the EMG data over the 4s post-stimulus period was averaged. The 1s pre-stimulus period was averaged and subtracted from the post-stimulus period to obtain an average change score for each of the twelve stimuli that were presented to individual participants. Significant effects were evaluated at $p < .05$.

LEMtool data was assessed along the valence dimension. Indications for the four positive LEMtool emotion images were collapsed into a single category 'positive emotions', and indications for the four negative LEMtool emotion images were collapsed into a category 'negative emotions'.

Because websites screenshots were treated as a 'black box' (see Section 4.1), eye-tracking data was analyzed based on the number of fixations and the relation between fixation duration and non-fixation duration, per website screenshot for each participant. The minimum duration time for a fixation was set to 50ms (Cyr et al., 2009; Cyr et al, 2010). The relation between fixation duration and non-fixation duration was calculated by subtracting the total time that the website was viewed but not fixated upon (indicated by the black line in Figure 9) from the total fixation duration for that website (an accumulation of the duration of each fixation, visualized in Figure 9 by the red dots). This way a single metric was obtained that reflected the relation between fixation and non-fixation. If participants focussed on specific elements of a screenshot, the difference score would be low, whereas if participants glanced the website without fixating on specific elements, the difference score would be high. This fixation duration difference score was used for the statistical analysis of the eye-tracking data. Figure 9 illustrates these data types.

It has to be noted here that the number of fixations and the relation between total fixation duration and total non-fixation duration, are global units of analysis of eye-tracking data. These metrics do not provide insight into what participants looked at, nor do they give a detailed account of how participants looked at the websites shown (e.g. order of fixations). However, this method of analysis does indicate general behavioral patterns in eye-tracking

data. The goal here was to investigate behavioral differences between participants presented with high and low visual appeal websites. These general behavioral patterns are a valid starting point for such an investigation.

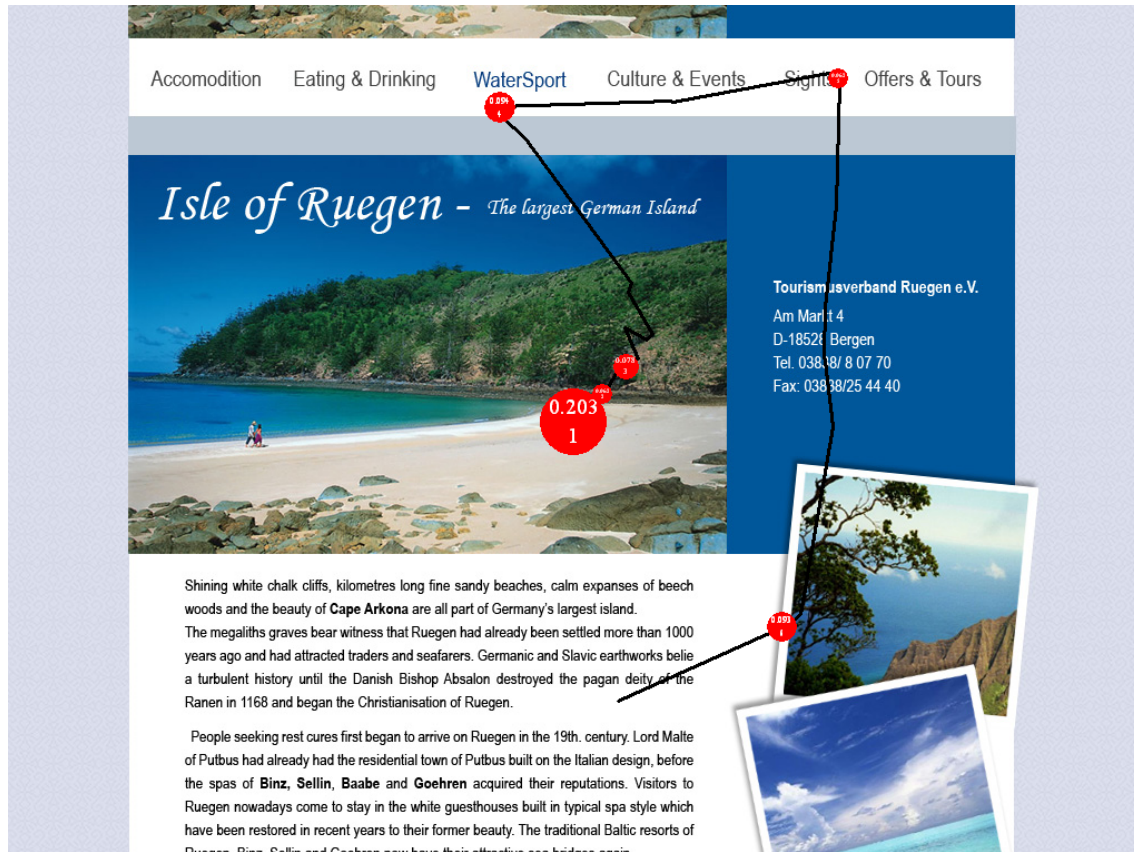


Figure 9 Example of eye tracking data after a 1s exposure time of an Isle of Rügen high visual website screenshot. The red dots indicate fixation points, with larger dots indicate a longer fixation duration. The black line represents the gazetrail, which indicates eye-movements that were not counted as a fixation.

5.2 Results

5.2.1 Motor expression component: Facial EMG

The first hypothesis presented in Chapter 3 stated that the motor expression component of emotion should be able to differentiate between high and low visual appeal websites. It was expected that low visual appeal websites displayed for 50ms would result in significantly higher corrugator EMG activity than high visual appeal websites, indicating a negative affective experience towards the low visual appeal websites.

To test this hypothesis, first, an independent samples *t*-test was calculated with stimulus set (Set A and Set B) as the independent variable and average EMG change score as the dependent variable. This demonstrated that there was no significant difference in the average EMG change score between participants that were presented with Set A and participants that were presented with Set B ($t(514) = -1.75, p > .05$). Therefore, participants presented with Set A or Set B were collapsed into a single group.

Next, an independent samples *t*-test was calculated with visual appeal (twelve high visual appeal and twelve low visual appeal websites) as the independent variable and average EMG change score as the dependent variable. No significant difference in the average EMG change score was found between high and low visual appeal websites ($t(514) = -.74, p > .10$). The average change score for high visual appeal websites ($-.13 \mu V, SD = 0.10$) did prove to be lower than the average change score for low visual appeal websites ($-.03 \mu V, SD = 1.84$), indicating that corrugator EMG activity as a measure of the motor expression component of emotion, was higher for low visual appeal websites than for high visual appeal websites.

Table 7

Post-hoc (Tukey's HSD) comparison between visual appeal rating categories (on a ten-point scale where 1 is 'visually unappealing' and 10 is 'visually appealing') for average EMG change score.

Rating Category	1 ($n = 88$)	2 ($n = 146$)	3 ($n = 172$)	4 ($n = 95$)	5 ($n = 15$)
1 (ratings 1 and 2)	-	ns	ns	ns	$p < .01$
2 (ratings 3 and 4)		-	ns	ns	$p < .05$
3 (ratings 5 and 6)			-	ns	ns
4 (ratings 7 and 8)				-	ns
5 (ratings 9 and 10)					-

To explore this issue further, visual appeal ratings were recoded into five separate categories. Category 1 contained low scores (1 and 2), category 2 moderately low scores (3 and 4), category 3 average scores (5 and 6), category 4 moderately high scores (7 and 8), and category 5 high scores (9 and 10). A one-way ANOVA was computed with the newly created visual appeal categories as the independent variable (five levels) and the average EMG change score as the dependent variable. There was a significant effect of visual appeal rating category on the average EMG change score ($F(4, 511) = 3.03, p < 0.05$). Tukey's HSD post-hoc analysis revealed that significant differences between group 1 and group 5 ($p < .01$), and between group 2 and group 5 ($p < .05$) were the main contributors (Table 7). This indicated that differences in

the motor expression component of emotion as a response to the visual appeal of websites only emerged when participants rated the websites as extremely high or low on visual appeal. This notion was examined in more detail by creating two new rating categories. One category of 'low ratings' (ratings 1, 2, and 3, $n = 153$) and one category of 'high ratings' (ratings 8, 9, and 10, $n = 48$) was created. An independent samples t -test was performed with the newly created categories as the independent variable and average EMG change score as the dependent variable. Websites that received a high visual appeal rating showed a lower average change score (mean = $-0,56 \mu V$, $SD = 1,32 \mu V$) than websites that received a low visual appeal rating (mean = $0,15 \mu V$, $SD = 2,16$). This difference was significant ($t(199) = 2,14$, $p < .05$). These findings suggest that ratings in the middle of the visual appeal scale, in this case ratings 4, 5, 6, and 7 are indications of emotional neutrality. Reactions in the motor expression component of emotion measured through facial EMG, suggest that low ratings (1 through 3) are related to negative affect, while high ratings (8 through 10) are related to positive affect.

5.2.2 Motor expression component: Influence of mental effort

As was explained in Section 2.6.1, it is often difficult to relate physiological data to specific events or causes. Moreover, the manipulation check (Chapter 4) revealed correlations between visual appeal and perceived usability, a mostly cognitive construct, for a limited number of the website stimuli used in the main experiment. To rule out the rival hypothesis that mental effort during the 4s post-stimulus period caused elevations in the corrugator EMG signal, two analysis were performed.

First, response latency data of the most extreme ratings (Figure 10), was examined. During the rating of the visual appeal of a website stimulus, higher response latencies are indications of an increase in mental effort on part of the participant (Tractinsky et al., 2006). Here, response latency was defined as the time between the appearance of the visual appeal rating scale and the button-press by the participant, indicating a visual appeal rating. Three new categories were created, one with extremely low ratings (ratings 1 and 2) one with the most average ratings (ratings 5 and 6) and one with extremely high ratings (ratings 9 and 10). Though differences for logarithmically transformed response latency data were only significant between category 1 and category 2 ($F(4) = 4,40$, $p < .05$). Figure 10 indicates that participants took longer to give average ratings compared to extremely positive or extremely negative ratings, suggesting an increase in mental effort while making average judgements.

To test if an increase in mental effort could have explained EMG elevations in the 4s post-stimulus period, a total of 516 response latencies was obtained (43 participants x 12 stimuli). Response latency data was logarithmically transformed before analysis. A Pearson

correlation was computed between the transformed response latency data and the average EMG change score. However, no significant correlation was found ($r = -0.03$, $p > .05$).

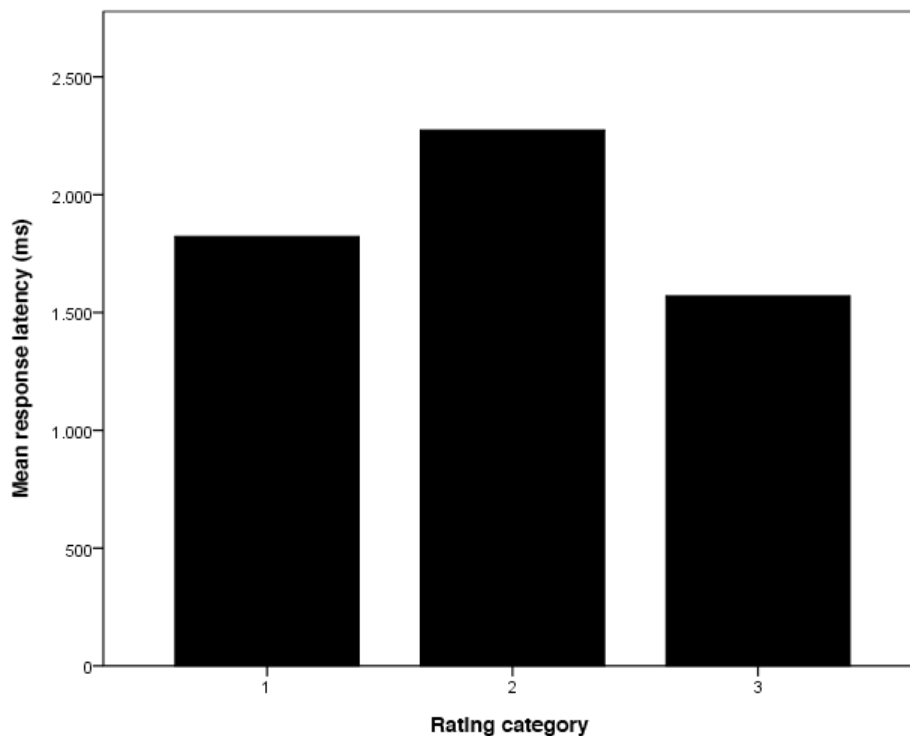


Figure 10 Mean response latency per rating category (category 1 = ratings 1 and 2; category 2 = ratings 5 and 6; category 3 = ratings 9 and 10).

Second, an increase in mental effort as a result of a high level of visual complexity of the website screenshots might have accounted for the elevation of the EMG signal (Tuch et al., 2009). It has been found that a larger JPEG file size is related to a higher level of visual complexity (Tuch et al., 2009). For each website screenshot the JPEG file size in Kilo Bytes was obtained. Next, average EMG change scores were averaged over participants resulting in a single EMG change score per stimulus. A Pearson correlation was calculated between the JPEG file size and EMG change scores. Again, no significant correlation was found ($r = 0.29$, $p > .05$).

These findings suggest that the cognitive construct of mental effort was unrelated to the elevations in the EMG signal. Therefore the corrugator EMG measurements seemed to signal a change in the motor expression component of emotion as a result of the judgement of the visual appeal of the presented website stimuli. This supports Hypothesis 1.

5.2.3 Behavioral component: Eye-tracking

The behavioral component of emotion was measured through eye-tracking. The number of fixations and the fixation duration difference score, were used to assess general behavioral patterns while viewing high visual appeal and low visual appeal websites. Hypothesis 2 stated that the behavioral component of emotion should be similar for high visual appeal and low visual appeal websites. No significant differences were expected between high and low visual appeal websites for the number of fixations and the fixation difference score.

Due to a technical error, eye-tracking data from one participant was not properly recorded. Therefore the analyses were conducted for the remaining 42 participants.

First, an independent samples *t*-test for the number of fixations was calculated with stimulus set (Set A and Set B) as the independent variable. Websites in Set A had an average of 5,06 fixations ($SD = 1,33$) and websites in Set B had an average of 4,70 fixations ($SD = 1,48$). This difference was significant ($t(502) = 2,87, p < .01$), so Set A and Set B were analyzed separately.

For both Set A and Set B, an independent samples *t*-test was calculated with visual appeal (high and low visual appeal websites) as the independent variable and number of fixations as the dependent variable. In Set A, high visual appeal websites had an average number of fixations of 4,98 ($SD = 1,29$) and low visual appeal websites had an average number of fixations of 5,14 ($SD = 1,38$). This difference was not significant ($t(250) = -0,95, p > .05$). Similar, no significant difference ($t(250) = 1,49, p > .05$) between the average number of fixations for high visual appeal (mean = 4,84, $SD = 1,53$) and low visual appeal (mean = 4,56, $SD = 1,43$) in Set B was found.

As a final check, for both Set A and Set B, actual ratings given by participants² for each website, were recoded into five categories (Category 1 (ratings 1 and 2), category 2 (ratings 3 and 4), category 3 (ratings 5 and 6), category 4 (ratings 7 and 8), and category 5 (ratings 9 and 10)). A one-way ANOVA was computed with rating category (five levels) as the independent variable and number of fixations as the dependent variable. No significant difference for number of fixations was found between the rating categories (Set A: $F(4) = 1,12, p > .05$; Set B: $F(4) = 1,09, p > .05$). These findings indicate that there was no significant difference between the number of fixations for high visual appeal and low visual appeal websites in both Set A and Set B. Participants generally fixated the same number of times on high visual and low visual appeal websites.

² Note that for one participant no rating data was recorded during the third phase of the experiment (see Section 5.2.4). Therefore this analysis was conducted with 41 participants.

The fixation duration difference score was analyzed similarly to the data for number of fixations. First, an independent samples *t*-test was calculated with stimulus set (Set A and Set B) as the independent variable and the fixation duration difference score as the dependent variable. No significant differences were found for the fixation duration difference score between Set A and Set B ($t(502) = -1,21, p > .05$), so Set A and Set B were analyzed as a single group.

Next an independent samples *t*-test was calculated with visual appeal (high and low visual appeal websites) as the independent variable and the fixation duration difference score as the dependent variable.

Identical to the analysis for the number of fixations, actual ratings given by participants were recoded into five categories. A one-way ANOVA was computed with rating category (five levels) as the independent variable and the fixation duration difference score as the dependent variable. No significant effect of rating categories on the fixation duration difference score was found ($F(4) = 0,62, p > .05$). These analyses suggest that there was no difference between high and low visual appeal websites for the fixation duration difference score. The duration that participants fixated on a specific area (indicated by the red dots in Figure 10) and the duration that they did not fixate on anything in particular (indicated by the black line in Figure 10) was similar for high and low visual appeal websites.

Taken together, the eye-tracking data supports Hypothesis 2. The behavioral component of emotion revealed similar behavioral patterns for high and low visual appeal websites.

5.2.4 Subjective feeling component: Visual appeal rating scales

Similar to previous studies into the judgement of visual appeal of briefly presented website stimuli (Lindgaard et al., 2006; Tractinsky et al., 2006), ten-point visual appeal rating scales were used to assess judgements of visual appeal. Hypothesis 3 stated that visual appeal ratings should differentiate between high and low visual appeal websites with a stimulus exposure time of 50ms. Specifically, high visual appeal websites should receive significantly higher visual appeal than low visual appeal websites.

For the 50ms exposure condition an independent samples *t*-test was calculated with stimulus set (Set A and Set B) as the independent variable. Results showed that there was no significant difference between visual appeal ratings for participants that were presented with either Set A or Set B ($t(514) = -0.63, p > .10$). Therefore, participants presented with Set A or Set B were collapsed into a single group.

Next, for the 50ms exposure condition an independent samples *t*-test was calculated with visual appeal (high visual appeal and low visual appeal websites) as the independent variable and visual appeal ratings as the dependent variable. High visual appeal websites received an average visual appeal rating of 5,60 (*SD* = 1,90) out of 10, while low visual appeal websites received an average visual appeal rating of 3,85 (*SD* = 1,94) out of 10. This difference was significant ($t(514) = 10,39, p < .001$). These findings confirmed Hypothesis 3, indicating that 50ms is enough for participants to differentiate between high visual appeal and low visual appeal websites.

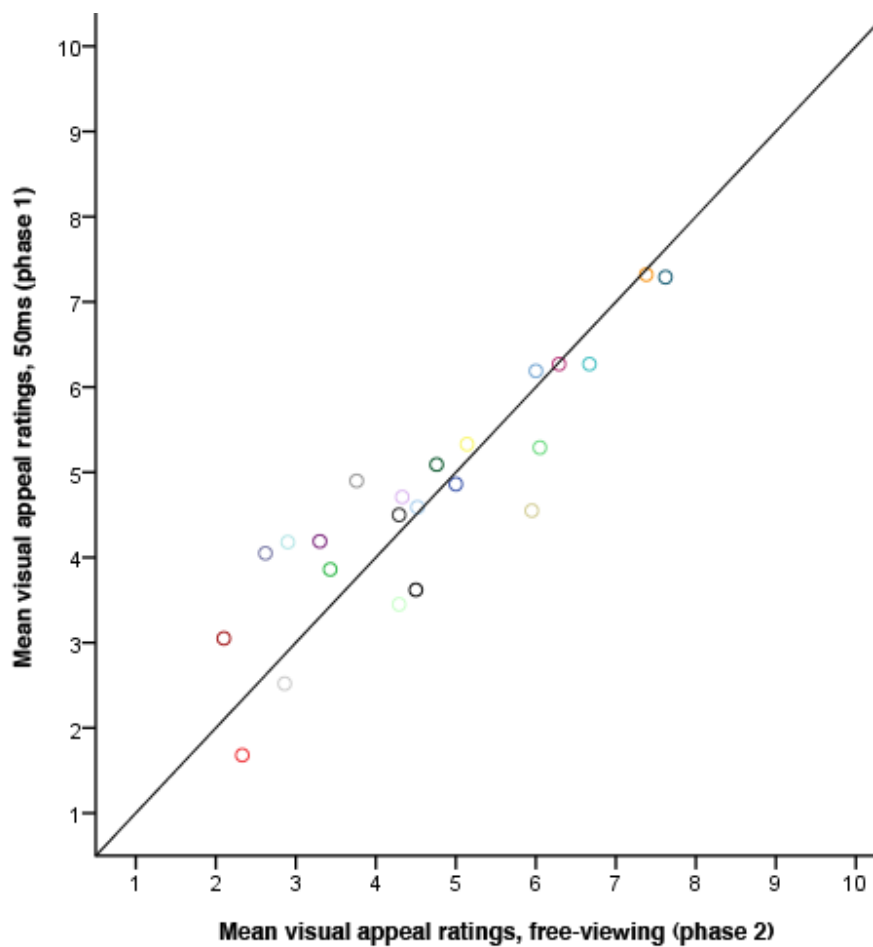


Figure 11 Scatter plot visualizing the relation between mean visual appeal ratings in the 50ms exposure condition and mean visual appeal ratings in the free-viewing condition. The dots visualize the 24 website screenshots.

The fourth hypothesis stated that there should be a high correlation between visual appeal ratings in the 50ms condition and visual appeal ratings in the free-viewing condition. This was

investigated by looking at correlations between ratings in the 50ms exposure condition and the free-viewing condition.

Unfortunately, due to a technical error in the data recording, for one participant, visual appeal ratings in the free-viewing condition were not recorded. Moreover, two ratings of two separate participants were not recorded. Analysis were carried out with the remaining ratings of the 42 participants.

First, in both the 50ms exposure condition and the free-viewing condition, visual appeal ratings were averaged for each website, resulting in 24 average ratings per condition (one for each website). Next, a scatter plot was drawn in order to visualize the relation between ratings in the 50ms condition and the free-viewing condition (Figure 11).

Figure 11 shows that average ratings tended to move towards the low end of the scale, with no average visual appeal ratings above 8 on the ten-point scale. Furthermore, mean ratings clustered relatively closely to the ideal line, with no extreme outliers. However, towards the low end of the scale, a more diffuse scatter pattern was observed.

To test the relation between the mean visual appeal ratings of the 50ms exposure condition and the free-viewing condition for significance, a Pearson correlation was computed. A high correlation ($r = .88$) was found between both conditions. This correlation was significant ($p < .001$).

These findings strongly support Hypothesis 4, showing that 50ms was enough time for participants to distinguish between high and low visual appeal websites. An increase in exposure time did not make a difference in this judgement.

Finally, Hypothesis 5 stated that a significant correlation should be found between visual appeal judgments of websites displayed for 50ms and the average facial EMG change score. Based on previous findings with other physiological measurements of components of emotion, it was expected that this correlation would be small yet significant (Mahlke & Minge, 2008). To assess the fifth hypothesis a Pearson correlation was calculated between the visual appeal ratings and the average EMG change score. A small but significant negative correlation was found ($r = -0.13$, $p < .01$). The negative correlation indicates that lower visual appeal ratings (negative judgement) would result in an increase in corrugator supercilii muscle activity, a signal of negative affect, while higher visual appeal ratings (positive judgement) would result in a decrease in corrugator muscle activity. This indicates that visual appeal ratings of website stimuli displayed for 50ms indeed seemed to contain an affective component. This supports hypothesis 5.

5.2.5 Subjective feeling component: The LEMtool

The subjective feeling component of emotion was assessed using two self-report methods. One was the visual appeal rating scales, the other was the LEMtool. Hypothesis 6 stated that the LEMtool should be able to discriminate on valence between high and low visual appeal websites independent of exposure time. Specifically, it was expected that in both the 50ms exposure condition and the free-viewing condition high visual appeal websites would receive more positive LEMtool indications than low visual appeal websites. Moreover, it was expected that for both conditions, high visual appeal websites would receive more positive than negative LEMtool indications and low visual appeal websites would receive more negative than positive LEMtool indications.

For the LEMtool indications given in the first phase of the experiment a cross-tabulation (Table 8) was computed for visual appeal (two levels, high and low visual appeal websites) x LEMtool emotion (three levels, positive, negative and no emotion). Table 8 shows that, in general, participants were capable of using the LEMtool emotion images to indicate their feelings towards the websites displayed for 50ms. Only 18 of the total of 516 (43 participants x 12 stimuli) ratings were indicated as 'no emotion'. Indications of 'no emotion' were relatively equally distributed among high (8 indications) and low (10 indications) visual appeal websites. Furthermore, participants indicated more positive LEMtool images for high visual appeal websites (156 indications) than for low visual appeal websites (67 indications). Conversely, participants indicated more negative LEMtool images for low visual appeal websites (181 indications) than for high visual appeal websites (96 indications). A Pearson chi-square test was performed to assess significance. Results demonstrated a significant difference between high visual appeal and low visual appeal website for LEMtool emotion images ($\chi^2 = 60,55$, $df = 2$, $p < .001$).

Table 8

Cross-tabulation of positive and negative LEMtool emotion images and 'no emotion' for high and low visual appeal websites with an exposure time of 50ms.

	Positive emotion image	Negative emotion image	No emotion	Total
High visual appeal websites	154	96	8	258
Low visual appeal websites	67	181	10	258
Total	221	227	18	516

To compare LEMtool indications to visual appeal ratings, for the first phase of the experiment, an independent samples *t*-test was performed with LEMtool valence (positive or negative LEMtool emotion image) as the independent variable and visual appeal rating as the dependent variable. It was found that websites that were indicated with positive LEMtool images, also received a higher visual appeal rating (mean = 6,48, *SD* = 1,46), than websites that were indicated with negative LEMtool images (mean = 3,38, *SD* = 1,47). This difference was significant ($t(496) = 23,48, p < .001$), indicating a relation between visual appeal ratings and LEMtool indications.

Finally, to assess whether LEMtool indications were related to EMG activity, for the first phase of the experiment, an independent samples *t*-test was performed with LEMtool valence (positive or negative LEMtool emotion images) as the independent variable and the average EMG change score as the dependent variable. However, no significant difference was found between websites that were indicated with positive LEMtool images and websites that were indicated with negative LEMtool images for EMG activity ($t(496) = -1,01, p > .05$). This suggests that there was no relation between LEMtool indications and EMG activity. Therefore, Hypothesis 7 was not supported.

During the third phase of the experiment, participants used the LEMtool to select areas of the website screenshots and chose a LEMtool image for each area selected. There was no time-limit during this task (i.e. free-viewing). For these data, a cross-tabulation was computed for visual appeal (two levels, high and low visual appeal websites) x LEMtool emotion image (two levels, positive and negative emotions). These data are shown in Table 9.

In total 1602 LEMtool indications were given by all participants. Table 9 reveals that these indications were relatively equally distributed among high visual appeal websites (796 indications) and low visual appeal websites (806 indications). However, more negative images (851 indications) than positive images (761 indications) were indicated.

Table 9

Cross-tabulation of positive and negative LEMtool emotion images for high and low visual appeal websites during free-viewing.

	Positive emotion image	Negative emotion image	Total
High visual appeal websites	520	276	796
Low visual appeal websites	241	565	806
Total	761	851	1602

Looking at the data of high visual appeal websites, it was found that participants indicated more positive LEMtool images (520 indications) than negative LEMtool images (276 indications). In a similar fashion, for low visual appeal websites, participants indicated more negative LEMtool images (565 indications) compared to positive LEMtool images (241 indications). Moreover, these data reveal that high visual appeal websites received more than twice as many positive LEMtool images than low visual appeal websites (520 versus 241, respectively), while low visual appeal websites received more than twice as many negative LEMtool images than high visual appeal websites (565 versus 276 respectively).

A Pearson chi-square test was performed to assess significance of the cross tabulation. Results demonstrated a significant difference between high visual appeal and low visual appeal website for LEMtool images ($\chi^2 = 201,55$, $df = 1$, $p < .001$).

Combined, these results support Hypothesis 6. The LEMtool was able to discriminate on valence between high and low visual appeal websites.

6

Conclusions and discussion

In the previous chapter the results from the main experiment were presented. Differentiation between high and low visual appeal websites was found for the motor expression component of emotion, and the subjective feeling component of emotion. As expected, behavioral patterns as a indication of the behavioral component of emotion did not differ between high and low visual appeal websites.

The current chapter will discuss the findings from the main experiment. Results from the measurements of the three components of emotion that were studied, will be combined in order to reach a general conclusion. Then findings will be explained in light of visceral affect and emotional categorization.

Finally, limitations of the present study will be discussed, and recommendations for future research will be given.

6.1 Conclusions

6.1.1 General conclusions

The present study investigated the relation between visual appeal judgements of websites and affect. In UX research it is mostly assumed that visual appeal has an impact on the affective state of the perceiver, and that this affective state influences the subsequent judgement of visual appeal (Hassenzahl, 2007). Until now, little empirical evidence existed that supported this assumption (Zhou & Fu, 2007). However, results from the present study indicate that there is indeed a relation between judgements of visual appeal and affect. Evidence for this relation was found by looking at the motor expression, behavioral, and subjective feeling components of emotion.

In the first phase of the main experiment, participants were presented with high and low visual appeal websites displayed for just 50ms. Yet, even while having seen these websites for literally a fraction of a second, participants judged the websites that were pre-designed to be visually appealing as more visually appealing than websites that had been pre-designed to be unappealing. Moreover, removing the time-limit, and having participants view the websites for

as long as they wanted, did not drastically change their judgement of the level of visual appeal of these websites. A strong correlation was found between visual appeal judgements after 50ms stimulus exposure and visual appeal judgements after stimulus exposure without a time-limit. These findings suggest that 50ms is enough for people to judge the level of visual appeal of a website. As has been suggested in previous studies (Lindgaard et al., 2006), it is plausible that visual appeal judgements are based on visceral affective responses that manifest rapidly, even after very brief stimulus exposure durations.

Indeed, evidence for this was found by looking at the motor expression component of emotion. Facial EMG measurements were taken from the corrugator supercilii muscle region. Results from the first phase of the main experiment revealed that when participants gave extreme visual appeal ratings (i.e. extremely high or extremely low), a difference in activation of the corrugator supercilii muscle region was found. Specifically, when participants gave extremely low visual appeal ratings the corrugator supercilii muscle region showed more activity than when participants gave extremely high visual appeal ratings. Furthermore, a small yet significant negative correlation was found between visual appeal ratings and corrugator supercilii activity. These findings would suggest that participants actually experience more negative affect when giving extremely low visual appeal ratings compared to when giving extremely high visual appeal ratings.

Considering that most studies that use facial EMG measurements to study affect, use relatively extreme emotional stimuli (e.g. erotica, extreme violence), it is remarkable that with emotionally weak stimuli such as the websites used in the main experiment, any significant difference at all was found for facial muscle activity. Moreover, the findings from the present study cannot easily be explained by the cognitive construct of mental effort, since correlations were neither found between visual appeal judgements and response latency of the visual appeal judgement nor between visual appeal judgements and the visual complexity of the stimuli.

Further support for the finding that visual appeal judgements are related to affect, was found by looking at the measurements of the behavioral component and measurements of the subjective feeling component using LEMtool. Results from both these measurements corroborated the findings from the facial EMG measurements and the visual appeal rating scales.

Eye-tracking data that was used to study the behavioral component of emotion did not reveal any differences in viewing behavior between high and low visual appeal websites. It could be argued that similarities in the number of fixations and the duration of fixation and non-fixation were in fact artifacts of the identical exposure time used for both high visual appeal and low visual appeal websites. Following this line of reasoning, similar behavioral

patterns would be a product of the exposure time. While this statement may hold, still, this rules out the possibility that, given identical exposure times, participants viewed high and low visual appeal websites structurally differently. If, for example, participants would focus on attractive content in the high visual appeal websites, as they would during browsing (Jacob & Karn, 2005), lack of such attractive content in low visual appeal websites would have most likely resulted in differing numbers of fixations, and differing fixation and non-fixation durations. More specifically, participants would show more, and longer, fixations for high visual appeal websites than for low visual appeal websites. Clearly, this did not happen. Similar behavioral patterns resulted in varying visual appeal judgements. Furthermore, this finding matches results from studies that used validated affective stimuli, and found no significant difference in fixations between positive and negative affective stimuli (Nummenmaa et al., 2006).

The eye-tracking data suggest that differences in visual appeal judgements are not a product of widely differing viewing behavior between high and low visual appeal websites. Considering that seeing a website at a glance (i.e. 50ms), is enough for people to judge the level of visual appeal of that website, it seems likely that these similar viewing patterns are an indication that people base their visual appeal judgements on a general perception of the website, and that a visceral affective response to that general perception contributes to the formation of their visual appeal judgement. Again, this was indicated by the relation between facial EMG and visual appeal ratings.

Finally, the LEMtool, which was used as a measurement of the subjective feeling component of emotion, also corroborated the findings from the facial EMG measurement and visual appeal rating scales. Results from the LEMtool measurements showed that participants were capable of using the LEMtool visualizations of emotion, as a way to distinguish between high and low visual appeal websites. This was the case for website screenshots displayed for 50ms as well as for website screenshots displayed without a time-limit. Participants attached significantly more positive emotion visualizations to high visual appeal websites, and significantly more negative emotion visualizations to low visual appeal websites. Considering that the LEMtool emotion images are highly recognizable for valence (Huisman, 2009), the LEMtool results in themselves are another indication of the relation between visual appeal judgements and affect. Moreover, a relation was found between LEMtool indications and visual appeal ratings. Websites that were indicated with more positive LEMtool images, also received higher visual appeal ratings, while websites that were indicated with more negative LEMtool images, received lower visual appeal ratings. This shows that the different methods used to measure the subjective feeling component of emotion corroborated each other.

Findings from the measurements of all three components of emotion that were studied in the current investigation, point in the same direction: judgements of visual appeal are related to affect. The affective component in visual appeal judgements can be measured physiologically, using facial EMG measurements of the corrugator supercilii muscle region. Furthermore, both visual appeal rating scales as well as LEMtool can be used to express visual appeal judgements. However, only visual appeal rating scales seem to capture a part of the initial affective response, as indicated by the correlation between visual appeal ratings and facial EMG activity. This will be further discussed in Section 6.2. While the eye-tracking data did corroborate the findings from the other measurements, as an isolated measurement technique, it would prove unsuccessful in studying the relation between visual appeal and affect. Suggestions for improving the use of eye-tracking in studying the relation between visual appeal and affect will be given in the discussion (Section 6.2).

The next section will take the here discussed conclusions and try to explain them in light of visceral affect and emotional categorization.

6.1.2 Visceral affect and emotional categorization

Differences for facial muscle activity, as a measurement of the motor expression component of emotion, were only found between websites that received relatively extreme ratings (ratings 1, 2, and 3 on the low end and ratings 8, 9, and 10 on the high end). No differences were found between websites that received these extreme ratings and websites that received ratings around the mid-point of the scale (i.e. ratings 4, 5, 6, and 7). This would seem to suggest that websites that received a moderate visual appeal rating, did not elicit an affective response. It seems plausible that individuals only experience an affective reaction to stimuli that they perceive as highly attractive or aversive. This fits Norman's (2004a) view of visceral affective responses as innate, and only physiologically measurable, good/bad-judgements. Findings from the current study suggest that these visceral judgements indeed are all-or-nothing judgements (i.e. high visual appeal or low visual appeal). When presented with a stimulus that is neither very attractive nor very aversive, a strong visceral affective response may not occur.

This can be further explained in the light of emotional categorization (Niedenthal et al., 1999). It is plausible that participants already have pre-formed emotional categories (i.e. somatic markers) for high and low visual appeal websites. Because high and low visual appeal websites are more easily identified as attractive or aversive, these somatic markers can be applied rapidly, and subconsciously, after a 50ms stimulus exposure time. However, websites with a moderate level of visual appeal do not give indications in either the direction of high or low visual appeal, making emotional categorization difficult. Therefore, participants had to

apply more mental effort to judge the visual appeal of average websites. This was also indicated by a higher response latency for moderate visual appeal ratings.

The judgement of visual appeal seems to contain a subconsciously active, visceral affective component that allows people to judge the visual appeal of a website screenshot within 50ms. Such a visceral affective response may be an indication of somatic markers being applied for stimuli that already have clearly formed emotional categories. Furthermore, this affective component may have a biasing effect on visual appeal judgements even after longer stimulus exposure.

While the findings from the main experiment are tantalizing, and can be explained in light of emotional categorization, some nuances have to be made. Therefore the next section will discuss limitations of the present study.

6.2 Discussion

6.2.1 Limitations

Some limitations of the present study deserve mentioning. First of all, a convenience sample of Bachelor, Master, and PhD students was used in the main experiment. Therefore generalizability of the results is limited to these groups. Additionally, participants age ranged from 18 to 31, which is an age group sometimes referred to as Generation Y (Djamasbi et al., 2010). Generation Y individuals are likely to have more extensive experience with websites and general computer usage than, for instance, more senior (50+) individuals. Though visceral affective responses should differ minimally between age groups (Izard, 1994), it is unclear what the effect of differences in computer experience would be on the judgement of visual appeal of websites.

The selection of websites used in the present study represented a range of high and low visual appeal websites. Previous studies (Lindgaard et al., 2006; Tractinsky et al., 2006) used websites that differed maximally on visual appeal, eliminating websites that would be rated as moderately visually appealing. Using websites that are more clearly distinguishable as high and low visual appeal websites might have resulted in more clearly defined results. Specifically in the case of facial EMG measurements, where differences were only found between websites that received extreme ratings.

While such an approach might have resulted in a more clear cut distinction between high and low visual appeal websites, the fact that average visual appeal websites were part of the current stimulus set, provided a more detailed insight into the relation between low, moderate, and high visual appeal ratings. For this reason, The Medline websites that were

identified in the pre-test to be undistinguishable on visual appeal, were not eliminated from the main experiment. Keeping the Medline website played to the strength of the current study in providing a range of visual appeal ratings. Moreover, the Medline websites were not eliminated in order to keep the number of ratings as high as possible.

The way eye-tracking was used in the present study did not reveal any differences in viewing behavior between high and low visual appeal websites. While this finding supported findings from the other measurements, taking a different approach might provide more insight into eye-tracking as a method to distinguish between high and low visual appeal websites. Such an approach could for instance replicate Nummenmaa's et al. (2006) experiment. Positive emotional stimuli would be substituted for high visual appeal websites and negative emotional stimuli would be substituted for low visual appeal websites. Randomly presenting pairs of high and low visual appeal websites next to each other on one screen, might provide a more detailed insight into differences or similarities in the attention biasing properties between high and low visual appeal websites.

Another finding that needs to be mentioned here is that no relation was found between LEMtool indications and facial EMG measurements. No difference was found in facial muscle activity between websites that received positive LEMtool indications and websites that received negative LEMtool indications. This finding was unexpected. If LEMtool indications were a reliable measurement of emotional valence, some relation between LEMtool indications and motor expression responses should have been found. Two explanations can be given for why this did not occur.

First, participants were only given the opportunity to give a LEMtool indication after they had already given a visual appeal rating. The time between the presentation of the stimulus and the opportunity to give a LEMtool indication, might have been too long for participants to give an indication based on their initial affective response. This might have made the LEMtool indications a less accurate representation of the initial affective response.

Second, the fact that no intensity ratings were used while indicating LEMtool emotions in the first phase of the main experiment, might have influenced the relation between LEMtool indications and facial EMG measurements. As found in the visual appeal rating scale data, a difference for facial EMG was only found between extreme ratings. It is possible that even when giving moderate visual appeal ratings, participants indicated a LEMtool emotion image. Evidence for this can be found by looking at the limited number of 'no emotion' indications, compared to the high number of moderate visual appeal ratings. Even though no strong affective response might have taken place (as indicated by the moderate visual appeal rating), participants still selected a LEMtool emotion. This might have resulted in the absence of a relation between LEMtool indications and motor expression responses. Participants might

have considered the 'no emotion' option as socially undesirable, and, in wanting to aid the study in the best way possible, selected an emotion even though no affect was experienced.

It has to be noted here that only the LEMtool images were used in the first phase of the main experiment. Normally, when using the LEMtool, participants are not forced to choose an image, but instead only activate the LEMtool when they experience a certain feeling. LEMtool was used in this way during the third phase of the main experiment. When using LEMtool images in the way they were used in the first phase, providing rating scales to indicate intensity, could prove insightful in future research.

6.2.2 Recommendations

The current study has found evidence for the relation between visual appeal judgements and affect. Furthermore, it was found that several measurement techniques can be used to gain insight into this relation. For the field of User Experience, where researchers are concerned with factors influencing a user's affective state, this means that visual appeal should be considered an important construct. In the holistic perception of a visual interface, visual appeal might be the first factor to influence the user's affective state, potentially biasing later judgements of the interface. The notion that something relatively intangible, such as visual appeal, has an impact on the user's affective state, highlights the appropriateness of studying affect in relation to UX (Hassenzahl & Tractinsky, 2006). It should be considered encouraging for researchers that even simple visual appeal rating scales seem to capture a part of the affective experience of visual appeal.

On the other side, theories related to visual appeal in HCI, should consider incorporating affect as a variable. Expressive (Lavie & Tractinsky, 2004) or hedonic (Hassenzahl, 2003) dimensions, might be more clearly defined by adopting concepts from emotion research. Components from Scherer's (2005, 2009) Component Process Model, might prove valuable in explaining affective responses to visual appeal in HCI.

For the design community, the current findings can be considered good news. Visual design of websites is not just some superfluous decoration, but has an impact on the affective state of the user. Still, findings also indicate a challenge for designers. Only websites that are considered high visual appeal or low visual appeal websites elicit an affective response. While designing websites with average visual appeal might not impact the users affective experience negatively, designing high visual appeal websites is a way to delight users, and engage positive affective responses. Clearly, this is what designers must strive for. The fact that low visual appeal websites elicit negative affective responses in users, further highlights the importance of careful design of websites.

References

- Anttonen, J. & Surakka, V. (2005) Emotions and Heart Rate while Sitting on a Chair. *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems* (pp. 491-499). New York: ACM.
- Aue, T. & Scherer, K.R. (2008). Appraisal-driven somatovisceral response patterning: Effects of intrinsic pleasantness and goal conduciveness. *Biological Psychology*, 79, 158-164.
- Barrett, L. F. (2004). Feelings or words? Understanding the content in self-report ratings of experienced emotion. *Journal of Personality and Social Psychology*, 87(2), 266-281.
- 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.
- Bornstein, R.F., & D'Agostino, P.R. (1992). Stimulus Recognition and the Mere Exposure Effect. *Journal of Personality and Social Psychology*, 63(4), 545-552.
- Bradley, M.M. & Lang, P.J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.
- Bradley, M.M. & Lang, P.J. (2007). The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In J.A. Coan and J.J.B. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment* (pp. 29- 46). New York: Oxford University Press.
- Brave, S. & Nass, C. (2008). Emotion in Human-Computer Interaction. In J. Jacko and A. Sears (Eds.), *Handbook of Human-Computer Interaction* (pp. 77-92). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cacioppo, J.T., Bernston, G.G., Larsen, J.T., Poehlman, K.M. & Ito, T.A. (2000). The psychophysiology of emotion. In M. Lewis and J.M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 173-191). New York: The Guilford Press.

- Cacioppo, J.T. & Petty, R.E. (1986). Electromyographic Activity Over Facial Muscle Regions Can Differentiate the Valence and Intensity of Affective Reactions. *Journal of Personality and Social Psychology*, 50(2), 260-268.
- Calvo, M. G., & Lang, P. J. (2004). Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion*, 28, 221–243.
- Calvo, M. G., & Lang, P. J. (2005). Parafoveal semantic processing of emotional visual scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 502–519.
- Carroll, J.M. & Thomas, J.C. (1988). Fun. *SIGCHI Bulletin*, 19, 21-24.
- Cinzia, D.D. & Vittorio, G. (2009). Neuroaesthetics: a review. *Current Opinion in Neurobiology*, 19, 682-687.
- Codispoti, M., Bradley, M.M. & Lang, P.J. (2001). Affective reactions to briefly presented pictures. *Psychophysiology*, 38, 474-478.
- Codispoti, M., Mazzetti, M. & Bradley, M.M. (2009). Unmasking emotion: Exposure duration and emotional engagement. *Psychophysiology*, 46, 731-738.
- Cyr, D., Head, M. & Larios, H. (2010). Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies*, 68, 1-21.
- Cyr, D., Head, M., Larios, H. & Pan, B. (2009). Exploring Human Images in Website Design: A Multi-Method Approach. *MIS Quarterly*, 33, 539-566.
- Damasio, A.R. (2000). A second chance for emotion. In R.D. Lane and L. Nadel (Eds.), *Cognitive Neuroscience of Emotions* (pp. 149-156). New York: Oxford University Press.
- Darwin, C. (1998). *The expression of emotions in man and animals* (Ed. P. Ekman). London: HarperCollins. (Original work published in 1872)
- Desmet, P.M.A. (2002). Designing Emotions. Unpublished doctoral dissertation, Delft University of Technology, Delft, The Netherlands.

- Desmet, P.M.A. & Hekkert, P. (2007). Framework of Product Experience. *International Journal of Design*, 1(1), 57-66.
- Dimberg, U. & Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology*, 39, 39-45.
- Djamasbi, S., Siegel, M. & Tullis, T. (2010). Generation Y, web design, and eye tracking. *International Journal of Human-Computer Studies*, 68, 307-323.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169-200.
- Ekman, P. (1994). Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin*, 115(2), 268-287.
- Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press.
- Elfenbein, H.A., Beaupré, M., Lévesque, M. & Hess, U. (2007). Toward a Dialect Theory: Cultural Differences in Expression and Recognition of Posed Facial Expressions. *Emotion*, 7(1), 131-146.
- Fazio, R.H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15(2), 115-141.
- Frijda, N.H. (1986). *The Emotions*. Cambridge. UK: Cambridge University Press.
- Frohlich, D.M. (2004). Beauty as a Design Prize. *Human Computer Interaction*, 19, 359-366.
- Guiza Caicedo, D. (2009). Developing a Tool to Assess Emotions Elicited by Services. Unpublished master's thesis, Delft University of Technology, Delft, The Netherlands.
- Hassenzahl, M. (2003). The thing and I: understanding the relationship between user and product. In M. Blythe, C. Overbeeke, A.F. Monk and P.C. Wright (Eds.), *Funology: From Usability to Enjoyment* (pp. 31-42). Dordrecht, The Netherlands: Kluwer.

- Hassenzahl, M. (2004a). Beautiful Objects as an Extension of the Self: A Reply. *Human-Computer Interaction*, 19, 377-386.
- Hassenzahl, M. (2004b). The Interplay of Beauty, Goodness and Usability in Interactive Products. *Human-Computer Interaction*, 19, 319-349.
- Hassenzahl, M. (2007). Aesthetics in interactive products: Correlates and consequences of beauty. In H. Schifferstein & P. Hekkert (Eds.), *Product Experience* (pp. 287-302). San Diego, CA: Elsevier.
- Hassenzahl, M. & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology*, 25(2), 91-97.
- Hazlett, R.L. & Benedek, J. (2007). Measuring emotional valence to understand the user's experience of software. *International Journal of Human-Computer Studies*, 65, 306-314,
- Hermans, D., Spruyt, A., De Houwer, J. & Eelen, P. (2003). Affective Priming With Subliminally Presented Pictures. *Canadian Journal of Experimental Psychology*, 57(2), 97-114.
- Huisman, G. & Van Hout, M. (2010). The development of a graphical emotion measurement instrument using caricatured expressions: the LEMtool. In C. Peter, E. Crane, M. Fabri, H. Agius & L. Axelrod (Eds.) *Emotion in HCI - Designing for People. Proceedings of the 2008 International Workshop* (pp. 5-8). Rostock, Germany: Fraunhofer.
- Huisman, G. (2009). LEMtool - Visual Emotion Measurement in Digital Media. Unpublished bachelor's thesis, University of Twente, Enschede, The Netherlands.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin*, 115, 288-299.
- Jacob, R. J. K., & Karn, K. S. (2005). Eye tracking in human- computer interaction and usability research: Ready to deliver the promises. In J. Hyona, R. Radach, and H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movements* (pp. 573-605). London: Elsevier.

- Klein, J., Moon, Y. & Picard, R.W. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14, 119-140.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. (1993). Looking at pictures: Affective, facial, visceral, and behavioural reactions. *Psychophysiology*, 30, 261-273.
- Larsen, J.T., Norris, C.J. & Cacioppo, J.T. (2003). Effects of positive and negative affect on electromyographic activity over *zygomaticus major* and *corrugator supercilii*. *Psychophysiology*, 40, 776-785.
- Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 269-298.
- Law, E.L-C., Roto, V., Hassenzahl, M., Vermeeren, A.P.O.S. & Kort, J. (2009). Understanding, Scoping and Defining User eXperience: A Survey Approach. *Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems* (pp. 719-728). New York: ACM.
- LeDoux, J. (1994). Cognitive-emotional interactions in the brain. In P. Ekman and R.J. Davidson (Eds.), *The Nature of Emotion* (pp. 73-118). Oxford: Oxford University Press.
- LeDoux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. London: Phoenix.
- Lindgaard, G. & Whitfield, T.W.A. (2004). Integrating aesthetics within an evolutionary framework. *Theoretical Issues in Ergonomics Science*, 5(1), 73-90.
- 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. & Minge, M. (2008). Consideration of Multiple Components of Emotions in Human-Technology Interaction. In C. Peter and R. Beale (Eds.), *Affect and Emotion in HCI* (pp. 51-62). Berlin: Springer.

- Mandryk, R.L., Inkpen, K.M. & Calvert, T.W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behavior & Information Technology*, 25(2), 141-158.
- Moshagen, M., Musch, J. & Göritz, A.S. (2009). A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics*, 52(10), 1311-1320.
- Nakarada-Kordich, I. & Lobb, B. (2005). Effect of Perceived attractiveness of web interface design on visual search of web sites. *Proceedings of CHINZ '05* (pp. 25–27). New York: ACM.
- Niedenthal, P.M., Halberstadt, J.B. & Innes-Ker, A.H. (1999). Emotional Response Categorization. *Psychological Review*, 106 (2), 337-361.
- Norman, D.A. (2004a). *Emotional design: Why we love (or hate) everyday things*. New York: Basic Books.
- Norman, D.A. (2004b). Introduction to this special section on beauty, goodness, and usability. *Human Computer Interaction*, 19, 311-318.
- Nummenmaa, L., Hyönä, J. & Calvo, M.G. (2006). Eye Movement Assessment of Selective Attentional Capture by Emotional Pictures. *Emotion*, 6 (2), 257-268.
- Ortony, A., Clore, G.L. & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge, MA: Cambridge University Press.
- Ortony, A. & Turner, T.J. (1990). What's basic about basic emotions? *Psychological Review*, 97 (3), 315-331.
- Partala, T. & Surakka, V. (2004). The effect of affective interventions in human-computer interaction. *Interacting with Computers*, 16, 295-309.
- Picard, R.W. (1997). *Affective Computing*. Cambridge, MA: The MIT Press.
- Picard, R.W. & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, 14, 141-169.

- Picard, R.W., Vyzas, E. & Healey, J. (2001). Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175-1191.
- Poole, A., & Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction*, (pp. 211-219). Hershey, PA: Idea Group,
- Reber, R., Schwartz, N. & Winkielman, P. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience? *Personality and Social Psychology Review*, 8(4), 364-382.
- Reeves, B. & Nass, C. (2002). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. California, Stanford: CSLI Publications.
- Reuderink, B., Nijholt, A., & Poel, M. (2009). Affective Pacman: A frustrating game for Brain-Computer Interface Experiments. *Intelligent Technologies for Interactive Entertainment*, 9, 221-227.
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128, 934-960.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
- Russell, J.A. (1994). Is There Universal Recognition of Emotion From Facial Expression? A Review of the Cross-Cultural Studies. *Psychological Bulletin*, 115(1), 102-141.
- Scheirer, J., Fernandez, R., Klein, J. & Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14, 93-118.
- Scherer, K.R. (2005). What are emotions? And how can they be measured. *Social Science Information*, 44(4), 695-729.
- Scherer, K.R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23 (7), 1307-1351.

- Scherer, K.R. & Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition and Emotion*, 22 (5), 789-801.
- Smith, J.C., Löw, A., Bradley, M.M. & Lang, P.J. (2006). Rapid Picture Presentation and Affective Engagement. *Emotion*, 6(2), 208-214.
- Sonderegger, A. & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, 41, 403-410.
- Tassinari, L. G., Cacioppo, J. T. & Vanman, E.J. (2007). The Skeletomotor System: Surface electromyography. In J. T. Cacioppo, L. G. Tassinari, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 163–199). New York: Cambridge University Press.
- Tractinsky, N. (2004). A Few Notes on the Study of Beauty in HCI. *Human-Computer Interaction*, 19, 351-357.
- Tractinsky, N., Cokhavi, A., Kirschenbaum, M. & Sharfi, T. (2006). Evaluating the consistency of immediate aesthetic perceptions of web pages. *International Journal of Human-Computer Studies*, 64, 1071-1083.
- Tractinsky, N., Katz, A.S. & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, 13, 127-145.
- Tuch, A.N., Bargas-Avila, J.A., Opwis, K. & Wilhelm, F.H. (2009). Visual complexity of websites: Effects on users' experience, physiology, performance, and memory. *International Journal of Human-Computer Studies*, 67, 703-715.
- Tullis, T. & Albert, B. (2008). *Measuring the User Experience. Collecting, Analyzing, and Presenting Usability Metrics*. Burlington, MA: Morgan Kaufmann Publishers.
- Van den Broek, E., Janssen, J.H., Westerink, J. & Healey, J.A. (2009). *Prerequisites for affective signal processing (ASP)*. Paper presented at the International Conference on Bio-inspired Systems and Signal Processing, Porto, Portugal.
- Van Dongelen, R. (2008). *What is beautiful is good and usable*. Unpublished master's thesis, University of Twente, Enschede, The Netherlands.

- Wallbott, H.G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, 28, 879-896.
- Ward, R.D. & Marsden, P.H. (2003). Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies*, 59, 199-212.
- Ward, R.D. & Marsden, P.H. (2004). Affective computing: problems, reactions and intentions. *Interacting with Computers*, 16, 707-713.
- Whitfield, T.W.A. (2005). Aesthetics as Pre-linguistic Knowledge: A Psychological Perspective. *Design Issues*, 21(1), 3-17.
- Winkielman, P., Berridge, K.C. & Wilbarger, J.L. (2005). Unconscious Affective Reactions to Masked Happy Versus Angry Faces Influence Consumption Behavior and Judgments of Value. *Personality and Social Psychology Bulletin*, 31, 121-135.
- Winkielman, P. & Cacioppo, J.T. (2001). Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence That Processing Facilitation Elicits Positive Affect. *International Journal of Personality and Social Psychology*, 81(6), 989-1000.
- Zajonc, R. (1980). Feeling and Thinking: Preferences need no Inferences. *American Psychologist*, 35, 151-175.
- Zhang, P. (2009, May). *Theorizing the Relationship between Affect and Aesthetics in the ICT Design and Use Context*. Paper presented at the International Conference on Information Resources Management, Dubai, United Arab Emirates.
- Zhou, H. & Fu, X. (2007). Understanding, Measuring, and Designing User Experience: The Causal Relationship Between the Aesthetic Quality of Products and User Affect. In Jacko, J.A. (Ed.), *Human-Computer Interaction, Part I* (pp. 340-349). Heidelberg, Germany: Springer.

Appendices

Appendix A - Stimulus set

High visual appeal	Low visual appeal
Einstein - Formative years	
Set A	Set B

Formative Years 1

Einstein's parents, Hermann and Pauline, middle-class Germans. "I was the son of entirely religious (Jewish) parents," Einstein recalled.

"There was this huge world out there, independent of us human beings and standing before us like a great, eternal riddle, at least partly accessible to our inspection and thought. The contemplation of that world beckoned like a liberation."

Our story Einstein liked to tell about his childhood was of a "wonder" he saw when he was four or five years old: a magnetic compass. The needle's invisible, unbroken swing, guided by an invisible force, profoundly impressed the child. The compass convinced him that there had to be "something behind things, something deeply hidden." Even as a small boy Einstein was self-sufficient and thoughtful. According to family legend he was a slow learner at first, needing to be reminded to concentrate and persevere with which he would build up lessons of cards in quiet drives. The boy's thought was, "I should like to be such an engineer, and try a technical school who are clever once a week at the Einstein."

"At the age of 12, I experienced a wonder in a booklet dealing with Euclidean geometry."

Einstein - Imagination	
Set B	Set A

Resources Home | Colloquial Index | Cells | Library | Gallery | Landmarks Home | Revised: October 26, 2007

Albert Einstein
Physicist
1879 - 1955

"Imagination is more important than knowledge."
—Albert Einstein

Albert Einstein was born on March 14, 1879 in Ulm, Württemberg, Germany. Einstein contributed more than any other scientist since Isaac Newton to our understanding of physics and reality.

Einstein was slow to learn to talk, not beginning to speak until sometime after his second birthday. His slow verbal development combined with a rather rebellious and skeptical attitude, and was a considerable factor in that young Albert would never amount to much.

Einstein's mother, Pauline, was a talented pianist. She introduced Albert to music as a small child, beginning his violin lessons at age six. He retained under promissory instruction until discovering the joys of Mozart's sonatas at age 13. From that point on, although he had no further lessons, his violin remained a constant companion. Einstein said later that, "There are moments in music, I say in the form of music."

When Einstein was 10, a poor student named Hans "Marmel" began doing with the Einstein family once a week. He would bring illustrated science books for Albert to study, and they would discuss what Albert learned. Hans gave him a promising book on the years before Albert was to study the subject of school. He later recalled, "From the night of his mathematical genius was so high that I could no longer follow."

In 1895, Einstein entered the Swiss Federal Polytechnic School in Zurich to be trained as a physics and mathematics professor. He

Resources Home | Colloquial Index | Cells | Library | Gallery | Landmarks Home | Revised: October 26, 2007

Albert Einstein
Physicist
1879 - 1955

"Imagination is more important than knowledge."
—Albert Einstein

Albert Einstein was born on March 14, 1879 in Ulm, Württemberg, Germany. Einstein contributed more than any other scientist since Isaac Newton to our understanding of physics and reality.

Einstein was slow to learn to talk, not beginning to speak until sometime after his second birthday. His slow verbal development combined with a rather rebellious and skeptical attitude, and was a considerable factor in that young Albert would never amount to much.

Einstein's mother, Pauline, was a talented pianist. She introduced Albert to music as a small child, beginning his violin lessons at age six. He retained under promissory instruction until discovering the joys of Mozart's sonatas at age 13. From that point on, although he had no further lessons, his violin remained a constant companion. Einstein said later that, "There are moments in music, I say in the form of music."

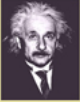
When Einstein was 10, a poor student named Hans "Marmel" began doing with the Einstein family once a week. He would bring illustrated science books for Albert to study, and they would discuss what Albert learned. Hans gave him a promising book on the years before Albert was to study the subject of school. He later recalled, "From the night of his mathematical genius was so high that I could no longer follow."

In 1895, Einstein entered the Swiss Federal Polytechnic School in Zurich to be trained as a physics and mathematics professor. He

Einstein - Albert Einstein

Set A	Set B
-------	-------

High visual appeal



Einstein 1879 - 1955

Early Years

Einstein was born on March 14, 1879, in Ulm, Württemberg, Germany. He was the son of Hermann and Pauline Koch Einstein. When Einstein was five years old, his father showed him a pocket compass. Little boy genius was deeply impressed by the mysterious behavior of the compass needle, which kept pointing in the same direction no matter which way the compass was turned. He later said he felt then that "something deeply hidden had to be behind things."

After attending elementary and secondary schools in Munich and in Aarau, Switzerland, Einstein studied maths and physics at the Polytechnic Institute in Zurich, Switzerland. He graduated in 1900. From 1902 to 1909, Albert worked as an examiner at the Swiss Patent Office in Bern. This job as patent examiner allowed him much free time, which he spent in scientific investigations. Einstein became a Swiss citizen in 1901.

The papers of 1905

At this time, Einstein made three of his greatest contributions to scientific knowledge. The year 1905 was an epoch-making one in the history of physics, because Einstein contributed three papers to *Annalen der Physik* (Annals of Physics), a German scientific periodical. Each of them became the basis of a new branch of physics.

In one of the papers, Einstein suggested that light could be thought of as a stream of tiny particles. This idea forms an important part of the quantum theory. In 1900, the German physicist Max K. E. L. Planck had proposed that the radiation of light occurred in packets of energy, called quanta. Einstein extended this idea by arguing that light itself consisted of quanta, which were later called photons.

Scientists before Einstein had discovered that a bright beam of light striking a metal caused the metal to release electrons, which could form an electric current. They called this phenomenon the photoelectric effect. But scientists could not explain the phenomenon as long as they assumed that light travelled only in waves. Using his theory of quanta, Einstein explained the photoelectric effect. He showed that when quanta of light energy strike atoms in a metal, the quanta force the atoms to release electrons.

Einstein's paper established the theoretical basis for the photoelectric cell, or "electric eye." This device made possible sound motion pictures, television, and many other inventions. Einstein received the 1921 Nobel Prize in physics for this paper on quanta.




Low visual appeal

---EARLY YEARS---

Einstein was born on March 14, 1879, in Ulm, Württemberg, Germany. He was the son of Hermann and Pauline Koch Einstein. When Einstein was five years old, his father showed him a pocket compass. Little boy genius was deeply impressed by the mysterious behavior of the compass needle, which kept pointing in the same direction no matter which way the compass was turned. He later said he felt then that "something deeply hidden had to be behind things."

After attending elementary and secondary schools in Munich and in Aarau, Switzerland, Einstein studied maths and physics at the Polytechnic Institute in Zurich, Switzerland. He graduated in 1900. From 1902 to 1909, Albert worked as an examiner at the Swiss Patent Office in Bern. This job as patent examiner allowed him much free time, which he spent in scientific investigations. Einstein became a Swiss citizen in 1901.

---THE PAPERS OF 1905---

At this time, Einstein made three of his greatest contributions to scientific knowledge. The year 1905 was an epoch-making one in the history of physics, because Einstein contributed three papers to *Annalen der Physik* (Annals of Physics), a German scientific periodical. Each of them became the basis of a new branch of physics.

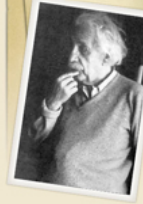
In one of the papers, Einstein suggested that light could be thought of as a stream of tiny particles. This idea forms an important part of the quantum theory. In 1900, the German physicist Max K. E. L. Planck had proposed that the radiation of light occurred in packets of energy, called quanta. Einstein extended this idea by arguing that light itself consisted of quanta, which were later called photons.

Scientists before Einstein had discovered that a bright beam of light striking a metal caused the metal to release electrons, which could form an electric current. They called this phenomenon the photoelectric effect. But scientists could not explain the phenomenon as long as they assumed that light travelled only in waves. Using his theory of quanta, Einstein explained the photoelectric effect. He showed that when quanta of light energy strike atoms in a metal, the quanta force the atoms to release electrons.

Einstein's paper established the theoretical basis for the photoelectric cell, or "electric eye." This device made possible sound motion pictures, television, and many other inventions. Einstein received the 1921 Nobel Prize in physics for this paper on quanta.

Einstein - Early years

Set B



Albert Einstein
1879 - 1955

Story Photographs School Zurich

Albert Einstein

Around 1886 Albert Einstein began his school career in Munich. As well as his violin lessons, which he had from age six to age thirteen, he also had religious education at home where he was taught Judaism. Two years later he entered the Luitpold Gymnasium and after this his religious education was given at school. He studied mathematics, in particular the calculus, beginning around 1891.

In 1894 Einstein's family moved to Milan but Einstein remained in Munich. In 1895 Einstein failed an examination that would have allowed him to study for a diploma as an electrical engineer at the Eidgenössische Technische Hochschule in Zurich. Einstein renounced German citizenship in 1906 and was to be statistics for a number of years. He did not even apply for Swiss citizenship until 1909, citizenship being granted in 1901.

Following the failing of the entrance exam to the ETH, Einstein attended secondary school at Aarau planning to use this route to enter the ETH in Zurich. While at Aarau he wrote an essay (for which was only given a little above half mark) in which he wrote of his plans for the future, see [15].

"If I were to have the good fortune to pass my examinations, I would go to Zurich. I would stay there for four years in order to study mathematics and physics. I imagine myself becoming a teacher in those branches of the natural sciences, choosing the theoretical part of them. Here are the reasons which lead me to this plan. Above all, it is my disposition for abstract and mathematical thought, and my lack of imagination and practical ability."

Set A

Albert Einstein

Born: 14 March 1879 in Ulm, Württemberg, Germany
Died: 18 April 1955 in Princeton, New Jersey, USA



Story Photographs School Zurich

[Click the picture above to see sixteen larger pictures](#)

Versions for printing

Around 1886 Albert Einstein began his school career in Munich. As well as his violin lessons, which he had from age six to age thirteen, he also had religious education at home where he was taught Judaism. Two years later he entered the Luitpold Gymnasium and after this his religious education was given at school. He studied mathematics, in particular the calculus, beginning around 1891.

In 1894 (Einstein's family moved to Milan but Einstein remained in Munich. In 1895 Einstein failed an examination that would have allowed him to study for a diploma as an electrical engineer at the Eidgenössische Technische Hochschule in Zurich. Einstein renounced German citizenship in 1906 and was to be statistics for a number of years. He did not even apply for Swiss citizenship until 1909, citizenship being granted in 1901.

Following the failing of the entrance exam to the ETH, Einstein attended secondary school at Aarau planning to use this route to enter the ETH in Zurich. While at Aarau he wrote an essay (for which was only given a little above half mark) in which he wrote of his plans for the future, see [15].

"If I were to have the good fortune to pass my examinations, I would go to Zurich. I would stay there for four years in order to study mathematics and physics. I imagine myself becoming a teacher in those branches of the natural sciences, choosing the theoretical part of them. Here are the reasons which lead me to this plan. Above all, it is my disposition for abstract and mathematical thought, and my lack of imagination and practical ability."

Rügen - Einfach

Set A

Rügen



einfach

Strand

Sale

Rent

About

Service

Ruegen

Deutsch

About Ostsee

Ruegen

Hotels

Products

Anybody who has seen the frozen waters of the bay in winter or the white carpet of wood anemones in the Granitz woods or who makes a cycling tour through the autumn foliage by boat in summer than by car. Trips to Rügen

The sea cycle and footpaths all over Rügen nowadays. You can also strike out across the water along the sea bridges, the resorts on the nearby Isle of Usedom can be reached faster by boat in summer than by car. Trips to Rügen

Set B

Ruegen



einfach

Strand

Sale

Rent

About

Service

Ruegen

Deutsch

About Ostsee

Ruegen

Hotels

Products

Anybody who has seen the frozen waters of the bay in winter or the white carpet of wood anemones in the Granitz woods or who makes a cycling tour through the autumn colours on Rügen, is bound to fall in love with this island.

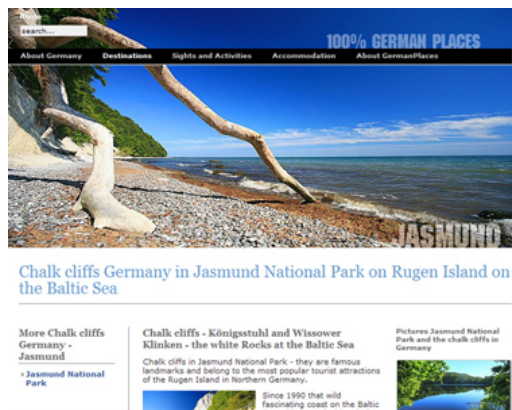
The sea cycle and footpaths all over Rügen nowadays. You can also strike out across the water along the sea bridges, the resorts on the nearby Isle of Usedom can be reached faster by boat in summer than by car. Trips to Denmark (Copenhagen, Isle of Bornholm), Sweden (Trelleborg)

Rügen - 100% German

Set B

Set A

High visual appeal



Low visual appeal



Rügen - Isle

Set A



Set B



Rügen - Largest

Set B

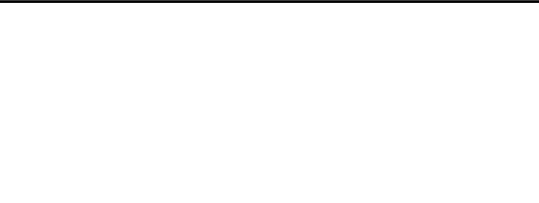


Set A

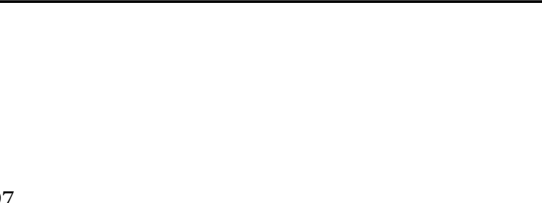


Headache - Medline

Set A

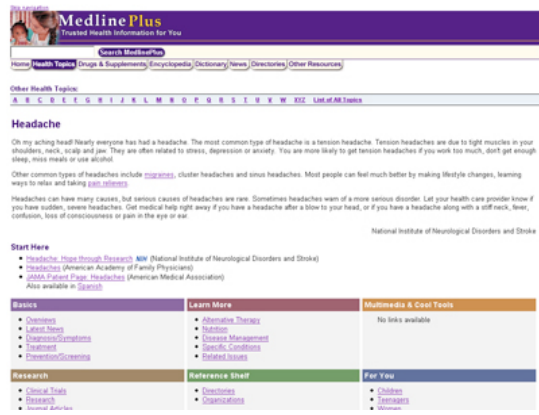


Set B



High visual appeal

Low visual appeal



Headache - Illustrated

Set B

Set A



Headache - Health

Set A

Set B



High visual appeal

Low visual appeal

Headache - Netdoctor

Set B

Set A

netdoctor

Search

All netdoctor

Log in | Register a new user

netdoctor.co.uk

Home

Become a member

Member FAQs

News and features

News

Features

Newsletters

Travel tips and health advice

Encyclopaedia

Diseases

Examinations

Medicines

Interactions

Discussion forums

Shop/Out

Test yourself

Services

Ask the doctor

Pharmacy

Directory

Health services

Find a hospital

Find a consultant

Find a birth unit

Complementary health practitioners

Support groups

Health centres

AZHC

Allergy and asthma

Cancer

Children's health

Depression

Diabetes

Headaches

Reviewed by [Dr. Rajiv Patel, MBChB, GP](#)

What is a headache?

Headaches involve mild to severe pain in one or more parts of the head as well as the back of the neck. There are many different types of headache patterns and a variety of causes, while painful and annoying, the majority of headaches do not indicate a serious disorder and, in non-chronic cases, may be relieved by medicines and/or changes in lifestyle.

What causes headaches?

There is no single cause of headaches. A number of causes have been identified which fall into two general categories.

Tension headache

This type of headache results from contraction of head and neck muscles. It is the most common form of headache and accounts for 70 per cent of headaches. It can occur in people of either sex and at any age but it is most common in adults and adolescents.

Tension headache usually occurs in isolated incidents but can become chronic for some people. Possible causes of muscle contraction associated with tension headaches include:

- [stress](#);
- [fatigue](#);
- poor posture;
- eye strain;
- [tobacco](#) and [alcohol](#) use.

• in women, hormonal changes occurring [before](#) and after a [menstrual period](#).

There is no single cause for headaches and, while they are painful and annoying, the majority of headaches do not indicate that there is a serious problem.

Sign up to Netdoctor and enjoy these benefits:

- Ask your own questions on the Netdoctor forum.
- Start your own online diary.
- Leave comments for other members.
- Receive an email every week day with the latest health news.

It's free and simple to join. So what are you waiting for?

[Review a Netdoctor member](#)

NETDOCTOR MEMBERSHIP

Sign up to Netdoctor and enjoy these benefits:

- Ask your own questions on the Netdoctor forum.
- Start your own online diary.
- Leave comments for other members.
- Receive an email every week day with the latest health news.

It's free and simple to join. So what are you waiting for?

[Review a Netdoctor member](#)

LATEST DISCUSSIONS

[Childbirth](#)

Tuesday 15:43

by [AliciaMum2008](#)

[Headaches](#)

Tuesday 15:32

by [Sally33](#)

NETDOCTOR

Search

All netdoctor

Log in | Register a new user

Headaches

Reviewed by [Dr. Rajiv Patel, MBChB, GP](#)

What is a headache?

Headaches involve mild to severe pain in one or more parts of the head as well as the back of the neck. There are many different types of headache patterns and a variety of causes, while painful and annoying, the majority of headaches do not indicate a serious disorder and, in non-chronic cases, may be relieved by medicines and/or changes in lifestyle.

What causes headaches?

There is no single cause of headaches. A number of causes have been identified which fall into two general categories.

Tension headache

This type of headache results from contraction of head and neck muscles. It is the most common form of headache and accounts for 70 per cent of headaches. It can occur in people of either sex and at any age but it is most common in adults and adolescents.

Tension headache usually occurs in isolated incidents but can become chronic for some people. Possible causes of muscle contraction associated with tension headaches include:

- [stress](#);
- [fatigue](#);
- poor posture;
- eye strain;
- [tobacco](#) and [alcohol](#) use.

• in women, hormonal changes occurring [before](#) and after a [menstrual period](#).

There is no single cause for headaches and, while they are painful and annoying, the majority of headaches do not indicate that there is a serious problem.

Sign up to Netdoctor and enjoy these benefits:

- Ask your own questions on the Netdoctor forum.
- Start your own online diary.
- Leave comments for other members.
- Receive an email every week day with the latest health news.

It's free and simple to join. So what are you waiting for?

[Review a Netdoctor member](#)

NETDOCTOR MEMBERSHIP

Sign up to Netdoctor and enjoy these benefits:

- Ask your own questions on the Netdoctor forum.
- Start your own online diary.
- Leave comments for other members.
- Receive an email every week day with the latest health news.

It's free and simple to join. So what are you waiting for?

[Review a Netdoctor member](#)

LATEST DISCUSSIONS

[Childbirth](#)

Tuesday 15:43

by [AliciaMum2008](#)

[Headaches](#)

Tuesday 15:32

by [Sally33](#)

99

Appendix B - Rating scales used in the pretest

Visual appeal, perceived usability, content familiarity, and structure familiarity rating scales used in the manipulation check. The scales are written in Dutch.

Hoe visueel aantrekkelijk vond u de getoonde website?

Zeer onaantrekkelijk 0 0 0 0 0 0 0 0 Zeer aantrekkelijk

Geef aan In hoeverre u het eens of oneens bent met de volgende stellingen.

De website leek goed *bruikbaar*.

Helemaal mee oneens 0 0 0 0 0 0 0 0 Helemaal mee eens

Ik heb, vóór deelname aan dit onderzoek, vaker websites met deze *inhoud* gezien.

Helemaal mee oneens 0 0 0 0 0 0 0 0 Helemaal mee eens

Ik heb, vóór deelname aan dit onderzoek, vaker websites met dit *ontwerp* gezien.

Helemaal mee oneens 0 0 0 0 0 0 0 0 Helemaal mee eens

Appendix C - Detailed analysis of cognitive constructs

To gain a more detailed insight into the relation between the visual appeal of the website screenshots and the cognitive constructs, paired-samples *t*-tests were calculated for each version of the websites for perceived usability, content familiarity, and structure familiarity. Previous findings (see Section 4.3) indicated that for all, but the Medline websites, the high visual appeal version was judged as more visually appealing than the low visual appeal version. Therefore, comparing the high and low visual appeal versions on the cognitive constructs, it is possible to relate these constructs to visual appeal judgements.

Table 10

Significant differences between high and low visual appeal versions of the Einstein websites for perceived usability, content familiarity, and structure familiarity (scale 1-10, 1=completely disagree and 10=completely agree)

Website	Construct	Version	Mean rating	<i>t</i> -value
Imagination	Structure Familiarity	High visual appeal	5,68	<i>t</i> (21)=-2,22*
		Low visual appeal	6,68	
Early Years	Perceived Usability	High visual appeal	6,05	<i>t</i> (21)=5,57***
		Low visual appeal	2,91	
Early Years	Content Familiarity	High visual appeal	6,32	<i>t</i> (21)=2,39*
		Low visual appeal	4,68	
Albert Einstein	Perceived Usability	High visual appeal	6,86	<i>t</i> (21)=4,76***
		Low visual appeal	4,36	

* $p < .05$

*** $p < .001$

Tables 10, 11, and 12 show significant differences between high and low visual appeal versions of each website type, on perceived usability, content familiarity and structure familiarity. In total, 17 out of the 36 (12 high-low visual appeal pairs x 3 constructs) pair-wise comparisons for all cognitive constructs were found to reveal a significant difference between the high and low visual appeal version of the websites. Of these 17 comparisons, 9 showed significant differences on perceived usability, again indicating a relation between visual appeal and perceived usability. Two out of the 17 comparisons were found to be related to content

familiarity, which reaffirmed that participants were unfamiliar with the content of the majority of website screenshots. Lastly, six of the 17 comparisons showed significant differences on structure familiarity, which can be interpreted as an indication of some familiarity with the layout structure of a number of webpages.

Table 11

Significant differences between high and low visual appeal versions of the Isle of Rügen websites for perceived usability, content familiarity, and structure familiarity (scale 1-10, 1=completely disagree and 10=completely agree)

Website	Construct	Version	Mean rating	<i>t</i> -value
Einfach	Perceived Usability	High visual appeal	6,32	<i>t</i> (21)=2,35*
		Low visual appeal	5,09	
100% German	Perceived Usability	High visual appeal	7,50	<i>t</i> (21)=5,30***
		Low visual appeal	4,50	
100% German	Structure Familiarity	High visual appeal	6,91	<i>t</i> (21)=2,49*
		Low visual appeal	5,55	
Largest	Perceived Usability	High visual appeal	7,18	<i>t</i> (21)=7,23***
		Low visual appeal	4,59	
Largest	Structure Familiarity	High visual appeal	7,27	<i>t</i> (21)=2,80*
		Low visual appeal	5,95	

* $p < .05$

*** $p < .001$

Some significant differences were found between high and low visual appeal versions of the websites for usability, content familiarity and structure familiarity. For instance the low appeal version of the Imagination Einstein website was perceived as more structurally familiar than the high visual appeal version. This indicates that visual appeal is not derived from familiarity with the structure. This notion is supported by the small number of low correlations found.

Another remarkable finding was that for the Medline website the high visual appeal version, which was previously perceived as less visually appealing than its low visual appeal counterpart, was perceived to be significantly more usable and familiar than the low visual

appeal counterpart. This indicates that visual appeal is not always an accurate predictor of perceived usability, as Tractinsky et al. (2000) found.

Table 12

Significant differences between high and low visual appeal versions of the Headache websites for perceived usability, content familiarity, and structure familiarity (scale 1-10, 1=completely disagree and 10=completely agree)

Website	Construct	Version	Mean rating	<i>t</i> -value
Medline	Perceived Usability	High visual appeal	6,14	<i>t</i> (21)=2,79*
		Low visual appeal	4,68	
Medline	Content Familiarity	High visual appeal	5,50	<i>t</i> (21)=2,94**
		Low visual appeal	4,23	
Medline	Structure Familiarity	High visual appeal	6,64	<i>t</i> (21)=4,74***
		Low visual appeal	5,00	
Illustrated	Perceived Usability	High visual appeal	6,05	<i>t</i> (21)=4,11***
		Low visual appeal	3,45	
Health	Perceived Usability	High visual appeal	6,86	<i>t</i> (21)=2,45*
		Low visual appeal	5,68	
Health	Structure Familiarity	High visual appeal	6,64	<i>t</i> (21)=2,16*
		Low visual appeal	5,41	
Netdoctor	Perceived Usability	High visual appeal	6,27	<i>t</i> (21)=3,71***
		Low visual appeal	4,27	
Netdoctor	Structure Familiarity	High visual appeal	5,68	<i>t</i> (21)=2,41*
		Low visual appeal	4,55	

* $p < .05$

** $p < .01$

*** $p < .001$

However, websites that received a high (i.e. 7 or higher on a 10-point scale) visual appeal rating were in every case considered to be more usable. It is possible that when the level of visual appeal of a website does not give individuals enough information in one direction (i.e. high visual appeal thus high usability) meaning they cannot make an accurate affective categorisation, other elements such as familiar structures are used in categorisation to make a judgement. The relatively long exposure time of 500ms would facilitate this process.

17 out of the 36 (12 high-low visual appeal pairs x 3 constructs) pair-wise comparisons for all cognitive constructs were found to reveal a significant difference between the high and low visual appeal version of the websites.