# Visual Complexity of Websites and its Effects on Experiential, Psychophysiological, Visual Search Reaction Time and Recognition Responses

Master's Thesis of:


**Alexandre Nicolas Tuch, B.Sc.**
Institute of Psychology
Department of Cognitive Psychology and Methodology
University of Basel


July 2007

Thesis Supervisors:


**lic. phil. Javier Bargas-Avila**
Department of Cognitive Psychology and Methodology, University of Basel

**Prof. Dr. Frank Wilhelm**
Department of Clinical Psychology and Psychotherapy, University of Basel

**Prof. Dr. Klaus Opwis**
Department of Cognitive Psychology and Methodology, University of Basel

## Abstract

The aim of this study was to examine the effects of visual complexity on the cognitive and emotional processing of websites within the framework of aesthetic theory and psychophysiological emotion research. We hypothesized that increasing complexity of websites would have a negative impact on cognitive and emotional processing. In a passive viewing task (PVT) 36 different website screenshots differing in their degree of complexity were presented to 48 participants. Additionally, a standardized visual search task (VST) assessing reaction times and a one week delayed recognition task on these websites were conducted, and participants rated all websites for arousal and valence. In the PVT and VST, psychophysiological responses were assessed through out. Visual complexity was correlated to heart rate (HR) decrease ($r$ = .39), arousal ($r$ = .74), and valence ($r$ = -.61). In the VST, there was an increase in HR and electrodermal activity that was correlated to reaction time (RT; $r$ = .43 and $r$ = .80 respectively). Visual complexity correlated with RT ($r$ = .28) and had an influence on the recognition rate of websites ($r$ = -.40). These results demonstrate that visual complexity of a website has an impact on perceived pleasure and arousal, psychophysiological response, visual search and recognition performance and should thus be considered an important factor in website design.

# Visual Complexity of Websites and its Effects on Experiential, Psychophysiological, Visual Search Reaction Time and Recognition Responses

To satisfy the user websites must be usable as well as appealing. As an introduction to a website the startpage is especially important because it serves as a first impression of the site. This first impression is often critical for users' expectations, and may influence whether or not they continue exploring the site. Therefore, it is important for website developers to understand how to attract users, keep them on their site, and provide a memorable experience that they will want to repeat. Visual complexity may play a decisive role in the formation of the first impression, particularly the experienced pleasure and arousal. To set up a theoretical framework there will first be a description of the relationship between visual complexity and pleasure, followed by a definition of affective space and how it is related to physiological responses.

## *Visual Complexity and Pleasure*

Complexity plays a crucial role in perception of visual stimuli. According to Berlyne's (1974) aesthetic theory, viewers' pleasure is related to the arousal potential of a stimulus. This relationship is expressed in an inverted U-shaped curve for pleasure, with a linearly increasing line for arousal potential of a stimulus. Berlyne proposed that stimuli with a moderate arousal potential were pleasurable, whereas stimuli with low arousal potential were experienced as boring, and the ones with high arousal potential were experienced as unpleasant. Thus, the arousal potential of a stimulus is linked to the potency of such *collative variables* as complexity, novelty, and ambiguity—those being the most important predictors for perceived aesthetic preference. The theory predicts that by adjusting visual complexity to an optimal level viewers' pleasure to an object will increase. Stimuli of a moderate degree of visual complexity will be considered pleasant, while less complex and more complex stimuli will be unpleasant.

Several studies support the effect of complexity on the perception of visual stimuli. In a study of perception of landscape photographs Kaplan, Kaplan, & Wendt (1972) showed that desirable landscapes contain moderate degrees of complexity. Ochsner (2000) found a negative correlation between visual complexity and affective valence, and a positive correlation between visual complexity and arousal in photographs of the International Affective Picture System (IAPS; Lang, Bradley & Cuthbert, 2005).

Furthermore, studies of websites show a relationship between visual complexity and perception. For example, Geissler, Zinkhan, and Watson (2006) found that consumers respond more favorably toward websites that fall in a moderate range of perceived complexity. Two other studies have indicated that website complexity may significantly influence consumer attitudes and intentions (Bruner & Kumar 2000; Stevenson, Bruner, & Kumar 2000). In a recent study Pandir and Knight (2006) could not directly support Berlyne's (1974) theory, but found a negative correlation between complexity and pleasure in website perception. These findings, however, support a possible relationship between visual complexity and experienced pleasure. But the direction of the relation is not clear from the literature.

Stimulus complexity can be determined in two ways. First, subjective appraisals may act as indicators for the degree of complexity of an object. A second, more objective way is that stimulus complexity can be measured by the physical properties of an object (e.g. numbers of elements, dissimilarity of elements, and the degree to which several elements are responded to as a unit; Berlyne, 1960).

One possible way, and probably the easiest, to assess objective complexity of visual stimuli is digitized image compression. Although compression is not directly linked to theories of visual perception, compressed image file sizes (e.g. JPEG: Joint Photographic Experts Group; TIFF: Tagged-Image File Format; GIF: CompuServe Graphics Interchange Format) provide a way to estimate image complexity. The file size obtained after compression can be used as a measure of complexity so that a larger file size may be considered more complex. Effectively, complexity can be measured as the number of bytes preserved after compression (Riglis, 1998). In fact, there is evidence that the size of compressed image files predicts the subjective complexity of images generated by websites, marine electronic displays, and nature photographs (Donderi, 2005; Donderi, 2006). These findings suggest that compressed file size can be utilized as a reliable, valid and objective measure for visual complexity.

*Affective Space and Physiology*

Lang, Bradley, and Cuthbert (1998) propose a two-dimensional approach for the organization of emotions. The dimensions of affective valence and arousal form the so-called affective space. The valence dimension is divided into an appetitive and a defensive motivation system, and the arousal dimension reflects the intensity of activation in either the appetitive or the defensive system. This biphasic motivation can be seen as a behavioral tendency to approach (appetitive motivation) or withdraw (defensive motivation) from a stimulus. The valence dimension ranges from positive to negative affective valence (or from pleasant to unpleasant), whereas the middle of the dimension represents a neutral affect. On the other hand, the arousal dimension ranges from calm to highly aroused and reflects the intensity of the affective valence (Bradley, 2000).

The IAPS (Lang et al., 2005) provides a set of standardized emotional pictures for the use in studies of emotions. When participants were asked to rate—according to the two-dimensional approach—the pleasure (affective valence) and the arousal of these pictures, an affective space in the form of a boomerang-shaped curve resulted (Bradley, 2000; see Figure 1). According to this curve, on average affectively neutral pictures showed low levels of arousal. However, as levels of pleasantness or unpleasantness in picture ratings rose, so did levels of arousal. This interrelation of arousal and valence in affective space supports the idea of a biphasic motivation system where unpleasantly arousing pictures provoke activation in the defensive motivation system and pleasantly arousing pictures provoke activation in the appetitive motivation system.

Several studies have shown that various changes in physiological activity are integrally related to emotional responses (for an overview, see Cacioppo, Berntson, Larson, Poehlman, and Ito (2000)). The two-dimensional affective space model is also supported by physiological responses to presented stimuli: The arousal dimension appears to be primarily associated with the electrodermal system, which is innervated only by the sympathetic nervous system efferents (Bradley, 2000). Electrodermal activity (EDA) is correlated positively with the activity of the perspiratory glands. It rises with increased arousal related to emotional stimuli, independent of their emotional valence (Bradley, 2000).

The valence dimension is related to facial expression, like wrinkling the nose or frowning. Electromyography (EMG) activity of the *musculus corrugator supercilii* is sensitive to the pulling together of the brows and the wrinkling of the nose. It has been shown that the EMG activity of the corrugator is strongly associated with experienced affective valence (Bradley, 2000; Cacioppo, Petty, Losch, & Kim, 1986; Dimberg, 1990;). Thus, a high EMG reaction can be seen with negative stimuli (muscle pulling together), whereas a low reaction is the result of positive stimuli (muscle relaxing). Accordingly, neutral stimuli cause a moderate contraction of the muscle (Bradley, 2000).

The heart is dually innervated by the parasympathic and sympathic system. Consequently, heart rate (HR) can be affected by the alternation of the level of activation of either system. In picture viewing, HR shows a typical triphasic pattern with an initial deceleration, followed by an acceleratory response, and concluded by a second deceleration (e.g. Lang & Hnatiow, 1962). The measure of HR is associated with the valence dimension (pleasantness) (Winton, Putnam, & Krauss, 1984). Duly, a perception of unpleasant pictures results in an initial cardiac deceleration, whereas pleasant pictures provoke the greatest peak of acceleration. In fact, the cardiac waveform for unpleasant pictures often shows no clear acceleratory peak, but instead, a deceleration that is sustained across the picture interval (Bradley, Greenwald, & Hamm, 1993).

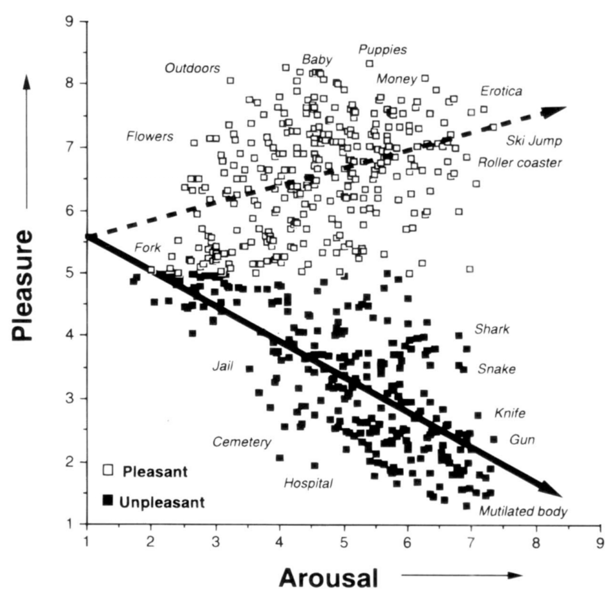


*Figure 1.* Distribution of Pictures in Affective Space (IAPS, 1998).

A factor analysis conducted on self-report, physiological and behavioral measurements showed a clear two-factor solution, supporting the two-dimensional approach for the organization of emotion (Lang, Greenwald, Bradley, & Hamm, 1993). As Table 1 illustrates, the first factor can be identified as the valence factor of appetite or aversion with high loadings for pleasantness ratings, HR change, corrugator EMG activity and zygomatic EMG activity. The second factor involves all identifiers of an arousal or intensity factor such as rated experienced arousal, interest ratings, viewing time, EDA and cortical slow waves.

Table 1.

*Factor Analysis of Emotional Picture Perception*

| Measure | Factor 1 (Valence) | Factor 2 (Arousal) |
|---|---|---|
| Valence ratings | **0.86** | -0.0 |
| Corrugator muscle | **-0.85** | 0.19 |
| Heart rate | **0.79** | -0.14 |
| Zygomatic muscle | **0.58** | 0.29 |
| Arousal ratings | 0.15 | **0.83** |
| Interest ratings | 0.45 | **0.77** |
| Viewing time | -0.27 | **0.76** |
| Electrodermal activity | -0.37 | **0.74** |

*Note*. This table is adapted from Bradley, 2000.

*Human-Computer Interaction and Psychophysiology*

In the past 10 years there has been an increasing interest in the influence of emotions in human-computer interaction (HCI). Picard (1997) introduced the field of affective computing and defined it as "computing that relates to, arises from, or deliberately influences emotions". An increasing number of studies have been published using psychophysiological measurements to identify emotional patterns in HCI. For example, pattern recognition of EDA and blood volume pressure was used to recognize user frustration in a puzzle task (Scheirer, Fernandez, Klein, &

Picard, 2002). Ward and Marsden (2003) showed that EDA response, HR response, and finger blood volume response differed depending on whether a task was conducted on a well or badly-designed version of a web site. Hazlett (2003) found that facial corrugator muscle activity during web site usage was greater for the website that was rated as more difficult and related this to experienced frustration. Another proposed physiological measure for arousal and valence was pupil size measurement (Partala, Jokiniemi, & Surakka, 2000; Partala & Surakka, 2003). There is a growing interest in the use of physiological methods in HCI to better recognize user affect, adapt to the user's affective state, generate affective behavior by the machine, model the user's affective states, or generate affective states within an agent's cognitive architecture. For an overview of the role of affect in HCI, see Hudlicka (2003).

In a review of affective computing, Ward and Marsden (2004) mention the difficulty of demonstrating statistically significant psychophysiological responses in HCI situations and of interpreting the found physiological patterns. They address this issue to the lack of rigorously controlled experimental settings and tightly controlled experimental conditions in HCI research. With these deficiencies in physiological HCI studies in mind, the present study aimed to investigate the influence of a very specific component (visual complexity) on websites under tightly controlled experimental conditions and settings as is usual for traditional psychophysiological research.

Considering the growing interest in the influence of emotions in HCI, and in accordance with established theories of emotion and aesthetics, the aim of this study was to investigate the impact of visual complexity on perception of real website startpages through physiological responses and subjective ratings. Websites were presented in a classic picture perception paradigm, as used in IAPS evaluation (Lang, 1998). The effect of visual complexity on searching performance and delayed recognition performance of websites was investigated as well.

According to the theory of Berlyne (1974), the findings of Pandir and Knight (2006) on the relationship between visual complexity and pleasure in websites, and the findings of Ochsner (2000) on correlations between complexity, affective valence, and arousal in IAPS pictures, we expected a negative correlation between visual complexity of websites and affective valence ratings (meaning that less complex sites are perceived as more pleasurable; $H_1$), and a positive correlation between visual complexity and arousal ratings (meaning that more complex websites are perceived as more arousing; $H_2$). In consideration of the relation between complexity and affective valence, we expected that increasing website complexity will lead to a higher deceleration in HR response ($H_3$). Because of the expected connection between visual complexity and affective valence, it can also be assumed that increasing website complexity will provoke higher corrugator activity in the EMG ($H_4$). Taking Berlyne's theory of stimulus arousal potential and the expected correlation between visual complexity and arousal ratings, we also expected that visual complexity would be positively related to physiological arousal measured by EDA ($H_5$).

We expected that search performance would be positively related to visual complexity meaning that more complex websites would lead to increased search time ($H_6$). It can be supposed that searching on websites will become increasingly aversive and stressful as time advances. This would lead to an increase in physiological arousal, resulting in an EDA increase ($H_7$), HR increase ($H_8$) and increased corrugator EMG activity ($H_9$).

## Method

### *Participants*

Forty-eight undergraduate psychology students at the University of Basel participated in the experiment. The sample consisted of 17 males (*M* = 25.5 years, *SD* = 7.7; range = 20-52) and 31 females (*M* = 21.2 years, *SD* = 2.2; range = 19-29).

*Stimuli*

As stimuli for the passive viewing, 36 different screenshots of existing websites (always the startpage) were used in the experiment. The reason for using real websites was to assure a high ecological validity of the experiment. The website screenshots differed on their degree of visual complexity according to the JPEG file sizes of the screenshots (*M* = 584 KB; *SD* = 158 KB; range = 272-864 KB). All screenshots were taken at the same resolution (1280 x 1024 pixels).

*Stimuli selection process.* First, 160 startpages with a wide range of visual complexity were selected from the World Wide Web according to the subjective impressions of the authors. There were several selection criteria: (1) the content of the website must be business, news, or science related; (2) written in the German or the English language; (3) not familiar to Swiss students. A full screenshot of each website was made without Internet browser elements.

Next, 108 of these 160 website screenshots were selected for validation in a web survey. The web survey was created with the web experiment tool eRes developed at the University of Basel (Schmutz, 2004). In order to reduce the length of the survey, the screenshots were divided into four sets consisting of 27 websites each; an individual participant rated one set only. Each participant was randomly assigned to one of the four sets. The website screenshots were presented to the participants one after another, and they had to rate each of them immediately on complexity, arousal, and aesthetic pleasantness by means of Likert scales ranging form 1 (*not at all*) to 7 (*extremely*). Eighty participants completed the rating task, and each of the 108 website screenshots was rated by at least 18 participants. The complexity ratings correlated significantly with the JPEG file size ($r$ = .33, $p$ < 001). On the basis of JPEG file size and subjective ratings, 36 website screenshots were selected for the experiment in order to obtain a wide range of visual complexity, including stimuli with low, middle and high complexity.

For the visual search task (VST), the same stimuli as in the passive viewing were used. As target stimulus for the VST, an asterisk was integrated on each screenshot with Adobe Photoshop CS (version 8.0.1). The asterisk was always embedded in text and adapted in size and color to its environment (see Figure 2). To avoid possible confounding, we checked if there was a correlation between the size of the asterisk and the visual complexity of the websites (JPEG file size), which was not the case.

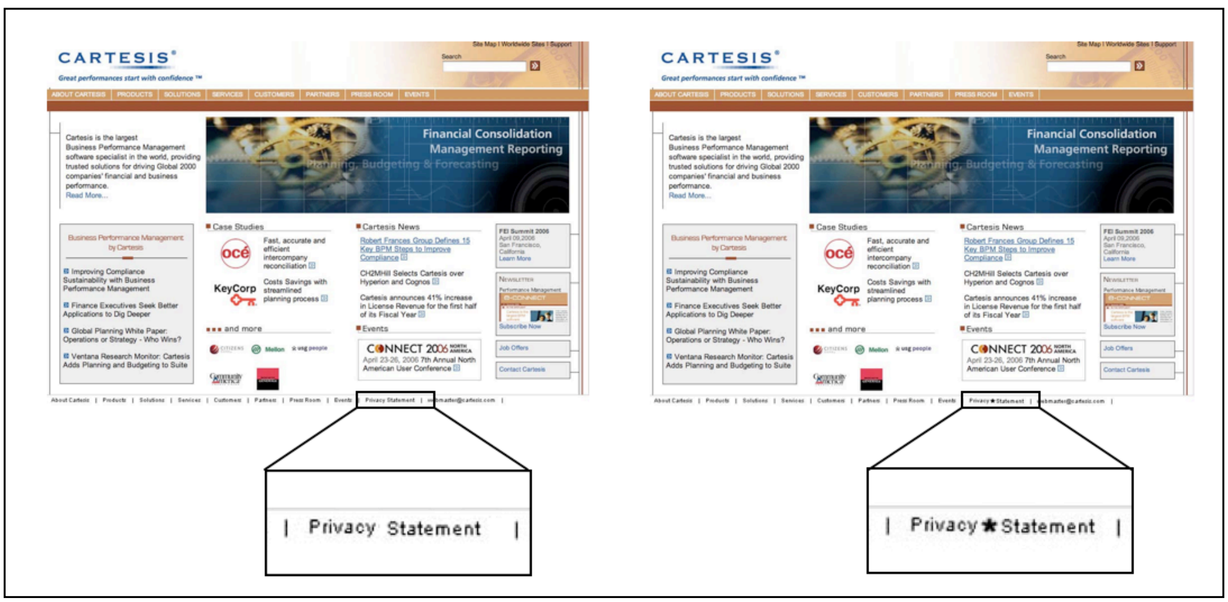


*Figure 2*. Example of a startpage used in the experiment: Original screenshot used in the passive viewing task (left) and manipulated screenshot with hidden asterisk for visual search task (right).

*Procedure*

The experiment was divided into two sessions. Only first session, included physiological recordings. It lasted for 90 minutes. One week after the first session, subjects returned to the lab to perform a recognition task based on the websites seen in the first session (for details see Figure 3).

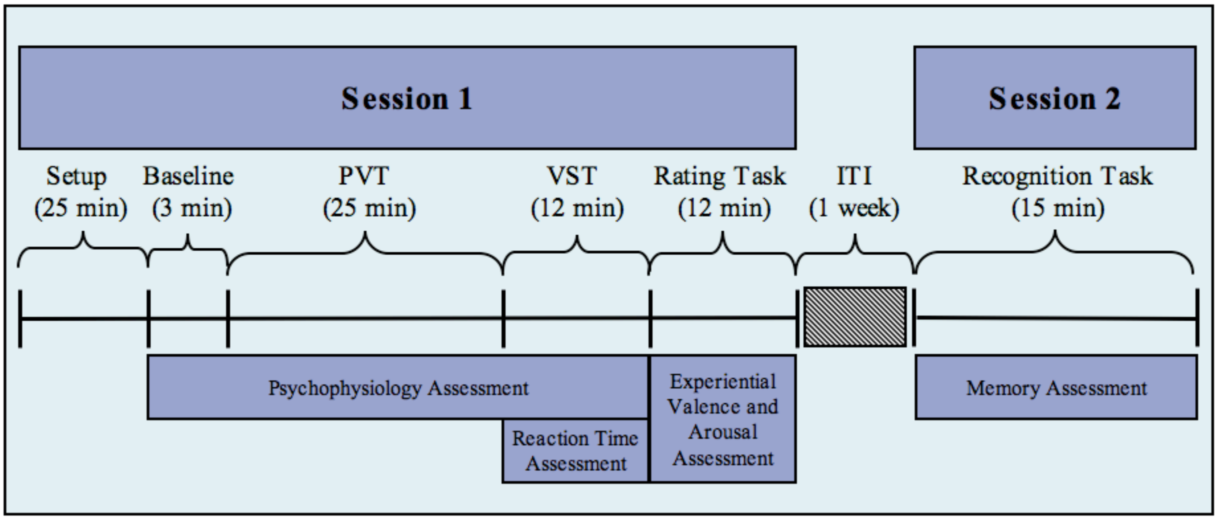


*Figure 3*. Experimental Procedure. PVT = Passive Viewing Task, VST = Visual Search Task, ITI = Inter Test Interval.

*Setting*. The experiment took place at the laboratory for Clinical Psychophysiology of the Institute of Psychology, University of Basel. The participants were seated in front of a computer monitor. With their right hand they controlled the progress of the experiment with a computer mouse. All instructions were displayed in written form on the screen. Each participant was tested separately, and room temperature was held constant between 21 and 23 degrees Celsius.

The participants were first given a short introduction to the lab setup and to the physiological measurements. After physiological sensors had been attached to the body and the left hand, the experimenter left and went to the adjacent control room where he could communicate with the participants by intercom, and observe them through a camera that was unobtrusively placed. Sensors were attached to the fingers of the left hand so that the participants would be able to use their dominant right hand to perform tasks with the computer mouse. All left handed participants ($n$ = 3) indicated that they also use their right hand to manage a computer mouse.

*Tasks.* In the beginning a three-minute baseline measurement was conducted where participants were advised to sit calm in the chair and look at the blank monitor. The experiment consisted of three tasks. The first task was a PVT: 36 screenshots of different websites which differed in their degree of complexity were presented in random order to the participants. Each screenshot was shown for 8 seconds, followed by a blank screen lasting for 8 seconds. Participants were advised to simply look at the website to get a visual overview of it (PVT).

The second task was a VST. The same 36 website screenshots, but with integrated target asterisks, were presented in a random order to the participants who were advised to scan each website for the target asterisk and to click on it with the computer mouse. Participants were advised to point at the target asterisk as fast and as precisely as possible. After clicking at the target the website disappeared and a centered fixation cross on a white background was shown for 2 seconds, after which the next website screenshot was displayed. This procedure was repeated until the participant completed the VST for all 36 websites.

Finally, the 36 website screenshots were shown again in random order and participants had to rate them on the arousal and the valence scale of the Self-Assessment Manikin (SAM). The SAM, devised by Lang (1980), is designed to assess the dimensions valence and arousal directly by means of two sets of graphical manikins. It has been extensively tested in conjunction with the IAPS (Lang et al., 2005).

One week after the first session, participants were reinvited to the lab for a computer based recognition task. The 36 websites of the previous experiment (targets) and 36 completely different websites (distractors) were presented in a random order. Participants had to judge for each screenshot if they had seen it in the previous experiment or not.

*Data Collection*

Reaction times and ratings were recorded with the software E-prime (Psychology Software Tools, 2002). Psychophysiological data were recorded during the experimental session with BIOPAC hardware and AcqKnowledge software (BIOPAC Systems, Inc., Santa Barbara, CA, USA, 2003). Physiological channels were sampled continuously at 1000 Hz. HR was measured with a standard Lead-II electrocardiogram (ECG) using three disposable electrodes. EDA was measured with two Ag/AgCl Beckman electrodes filled with an isotonic paste. Electrodes were attached to the volar surfaces of the medial index and middle fingers of the subject's left hand. A constant-voltage device was used to pass 0.5 V between electrodes. EMG was measured with Ag/AgCl miniature electrodes filled with electrolyte. The electrodes were attached bilaterally from the corrugator supercilii using the placements described by Lippold (1967) and Fridlund and Cacioppo (1986).

*Data Reduction*

Psychophysiological data were reduced using ANSLAB software (Wilhelm & Peyk, 2005). In brief, the raw ECG signal was 40 Hz lowpass-filtered and 0.5 Hz highpass-filtered to aid the automatic R-Waves identification. R-Waves were then edited for artifacts, false positives, or non-recognized R-waves, and transformed to instantaneous interbeat intervals (IBI). Note that IBI is inversely related to HR by the equation HR = 60000 / IBI.

The EDA raw signal was smoothed offline using a 1 Hz lowpass-filter, examined minute-wise and edited for artifacts.

The corrugator EMG raw signal was 48 to 52 Hz bandstop filtered to eliminate power supply interspersion, 27 Hz highpass-filtered to remove baseline drifts and low frequency components, rectified, and smoothed using a 50 ms moving average window.

For the PVT, stimulus related epochs from all three signals were extracted from 2 seconds before to 8 seconds after stimulus-onset. Next, mean values for each channel were extracted for 10 one-second-intervals. The two-second period before stimulus-onset was used as a stimulus related baseline. For the EDA signal, the difference score between the mean of seconds 3 to 4 and the baseline was calculated ($\Delta EDA_{PVT}$). For the EMG signal, the difference was computed between the mean of seconds 1 to 2 and the baseline ($\Delta EMG_{PVT}$). For the IBI signal, the difference score was computed between the mean of seconds 1 to 4 and the baseline ($\Delta IBI_{PVT}$). Those extracted epochs were selected a priori according to known physchophysiological response patterns in the picture perception paradigm.

A two-second-period before stimulus-onset and a two-second-period before stimulus-offset were extracted from all three signals in the VST. The two-second-period before stimulus-onset was used as a baseline. For EDA, IBI and EMG, the difference score between the baseline and the two-second-period before stimulus-offset was calculated ($\Delta EDA_{VST}$, $\Delta IBI_{VST}$ and $\Delta EMG_{VST}$). The difference scores represent the physiological change from the stimulus-onset to the offset.

*Statistical Analysis*

To account for between-subject variability not related to the task, all computed differences were z-transformed within individual participants. These differences were then aggregated for each website, so that every screenshot had six delta values (z-$\Delta EDA_{PVT}$, z-

$\Delta IBI_{PVT}$, z-$\Delta EMG_{PVT}$, z-$\Delta EDA_{VST}$, z-$\Delta IBI_{VST}$, and z-$\Delta EMG_{VST}$).

All data was checked for normal distribution and linearity. For a better fit to these criteria, reaction time (RT) in the VST was square root transformed ($\sqrt{RT}$).

To assure the linearity of the correlations scatter plots were carefully examined and alternative regression models (quadratic and cubic) were calculated. The alternative regression models could not provide significantly better fits, so for testing the degree of linear relationship between the variables Pearson Product Moment Correlations were calculated. An alpha level of .05 was used for all statistical tests.

## Results

### *Visual Complexity and SAM Ratings in the Passive Viewing Task*

Visual Complexity was significantly related to arousal ($r = .74$, $p < .001$; $H_1$) and valence ratings ($r = -.61$, $p < .001$; $H_2$) of the SAM scale (see Figure 4). There was also a significant correlation between the two SAM scales ($r = -.50$, $p < .003$).

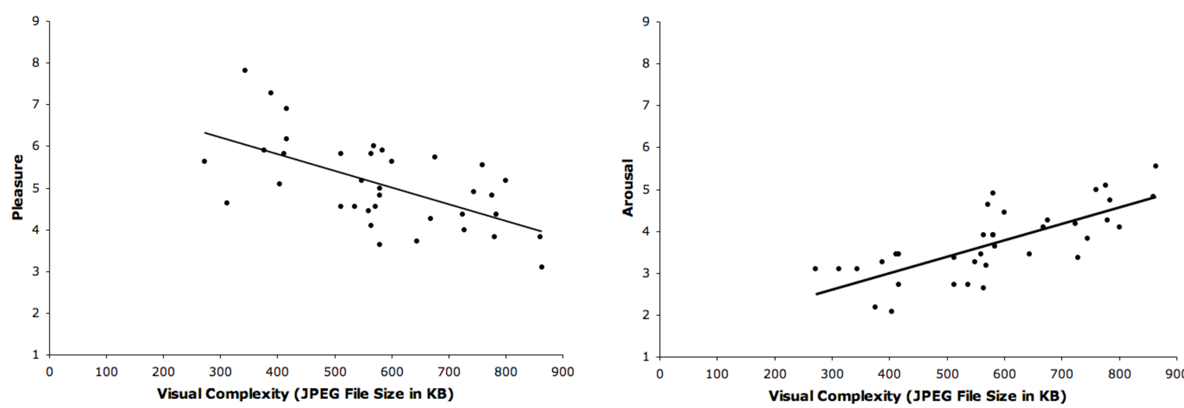


*Figure 4*. Scatter plots of the linear correlation between affective valence and visual complexity (left) and between arousal and visual complexity (right).

*Heart rate response.* There was a slight decrease in HR in response to the screenshots.

The mean increase in IBI was 1.42 ms (*SD* = 8.35 ms). There was a significant correlation between visual complexity and the z-$\Delta IBI_{PVT}$ ($r$ = .39, $p$ < .05; $H_3$; see Figure 5). Also, the arousal ratings were correlated to the z-$\Delta IBI_{PVT}$ ($r$ = .33, $p$ < .05). HR decrease was not related to valence ratings since the correlation between z-$\Delta IBI_{PVT}$ and SAM valence ratings was statistically not significant ($r$ = .05, $p$ = .396).

*EMG response.* There was a relatively weak corrugator EMG response to the stimuli (*M* = .4 µV; *SD* = .1 µS). As expected, the EMG response was related to visual complexity (see Figure 5), valence ratings and arousal ratings. The correlation between z-$\Delta EMG_{PVT}$ and visual complexity ($r$ = .28, $p$ < .05; $H_4$), the correlation between z-$\Delta EMG_{PVT}$ and the SAM valence ratings ($r$ = -.44, $p$ < .01) and the correlation between z-$\Delta EMG_{PVT}$ and SAM arousal ratings ($r$ = .38, $p$ < .05) were statistically significant.

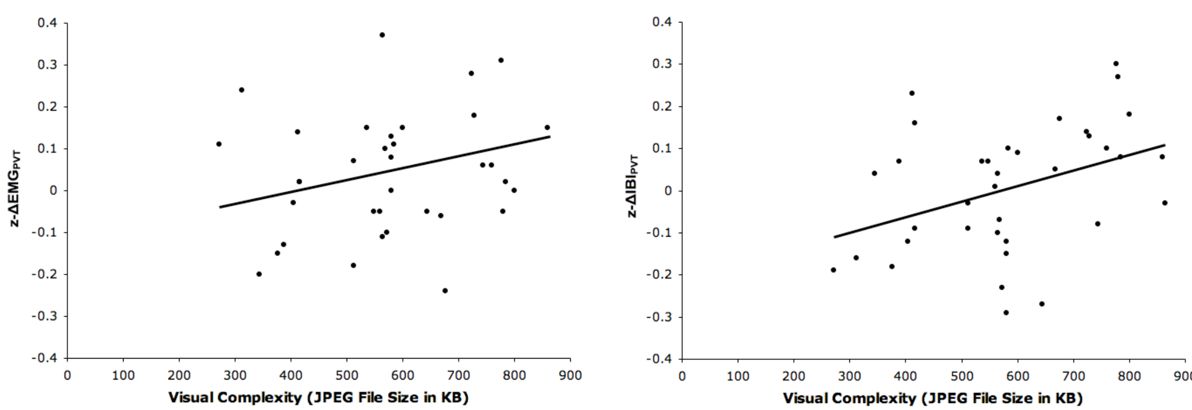


*Figure 5*. Scatter plots of the linear correlation between corrugator electromyography response and visual complexity (left) and between heart interbeat interval response and visual complexity (right) during the Passive Viewing Task.

*EDA response.* The expected EDA response to the website screenshots in the PVT did not occur (*M* = -.0051 µS; *SD* = .0006 µS). The correlation between visual complexity and z-

$\Delta EDA_{PVT}$ was not statistically significant ($r$ = -.12, $p$ = 0.307; $H_5$). Table 2 summarizes the correlations of the PVT.

Table 2.

*Correlation Coefficients for Passive Viewing Task*

| | Visual Complexity | SAM Arousal Ratings | SAM Valence Ratings |
|---|---|---|---|
| z-$\Delta EDA_{PVT}$ | -.12 | -.07 | .22 |
| z-$\Delta IBI_{PVT}$ | .39** | .33* | .05 |
| z-$\Delta EMG_{PVT}$ | .28* | .38* | -.44** |
| SAM Arousal Ratings | .74** | - | -.50** |
| SAM Valence Ratings | -.61** | - | - |

*Note*. * = $p$ < .05 (one-tailed); ** = $p$ < .01(one-tailed).

*Visual Complexity and Reaction Time in the Visual Search Task*

Mean RT to find and to point at the asterisk on the website screenshot was 8.4 s, *SD* = 4.0 s. This RT was related to visual complexity since the correlation between √RT and visual complexity reach statistical significance ($r$ = .28, $p$ < .05; $H_6$).

*EDA response*. In general, there was an increase in EDA while performing the search task on a website screenshot (*M* = .019 µS, *SD* = .015 µS). This increase was not related to visual complexity. The correlation between visual complexity and z-$\Delta EDA_{VST}$ was not statistically significant ($r$ = .13, $p$ = .22). However, EDA increase was highly correlated to RT ($r$ = 0.80, $p$ < .001), indicating increased physiological arousal as search time advanced ($H_7$).

*Heart rate response.* HR increased during the VST, this was indicated through shortening of mean IBI by 14.03 ms (*SD* = 11.62 ms). This HR increase was not related to visual complexity. The correlation between z-$\Delta IBI_{VST}$ and visual complexity was not significant ($r$ = .049, $p$ = .39). Nevertheless, as in the EDA response, there was a connection between HR increase and RT. The correlation between z-$\Delta IBI_{VST}$ and $\sqrt{RT}$ was statistically significant ($r$ = -.43, $p < .01$), meaning that HR increased with increasing RT ($H_8$).

*EMG response.* During VST, EMG response was neither related to visual complexity nor RT ($H_9$). No significant correlation was found.

*Recognition Task*

The mean recognition rate of the websites one week after the experiment was 84.1 % (*SD* = 1.2 %). The recognition rate was influenced by visual complexity since the correlation between recognition rate and visual complexity was significant ($r$ = -.40, $p < .01$), meaning that recognition was better with rather less complex websites. Recognition rate was also correlated to valence ratings ($r$ = .55, $p < .001$) and arousal ratings ($r$ = -.34, $p < .05$). For the psychophysiological data, there was a relation between HR increase during VST and recognition rate since the correlation between z-$\Delta IBI_{VST}$ and recognition rate was $r$ = -.29 ($p < .05$). Table 3 summarizes the correlations of the recognition task.

Table 3.

*Correlation Coefficients for the Recognition Task*

| | z-$\Delta EDA_{VST}$ | z-$\Delta IBI_{VST}$ | z-$\Delta EMG_{VST}$ | Visual Complexity | SAM Arousal Ratings | SAM Valence Ratings |
|---|---|---|---|---|---|---|
| Recognition Rate | .11 | -.29* | .03 | -.40** | -.34* | .55** |

*Note.* * = p < .05 (two-tailed); ** = p < .01 (two-tailed).

## Discussion

Based on these results, visual complexity of the startpage should be considered an important factor in website design. This study shows evidence that the degree of visual complexity of the startpage of a website has an impact on the perceived pleasure and arousal, physiological response and recognition performance of the participants. Startpages with low visual complexity were rated as more pleasurable and less arousing and participants achieved better search and recognition performances than on startpages with high visual complexity. The results suggest that relatively simple startpages are more favorable to users and can be better remembered (recognized) than more complex ones. Moreover, the study indicates that compressed file size (e.g. JPEG) is an easy way to obtain an index for assessing visual complexity, and provides a good predictor for users' first impression toward a website.

The findings support the hypothesized negative linear relationship between visual complexity and affective valence as well as the positive correlation between visual complexity and arousal ratings. As in the study of Panir and Knight (2006) the data do not directly support Berlyne's (1974) theory, which would suggest that startpages with a moderate degree of complexity and arousal potential would be perceived as most pleasurable. A quadratic relationship in form of an inverted U-shaped curve could neither be established between complexity and affective valence nor between affective valence and arousal. However, one could argue that the linear relation between visual complexity and affective valence is the result of an insufficiently broad range of visual complexity of the startpages used in the experiment. Depending on the location of a selected picture sample on the curve, a relation function (such as a linear correlation) can be derived from the supposed inverted U-curve, so the found negative correlation could be a result of lack of less complex stimuli in the sample (see Figure 6). This would lead to the assumption that most existing websites exceed such a low level of complexity to lie on the first half of Berlyne's inverted U-shaped curve. Geissler et al. (2006) found such an inverted curvilinear relationship between startpage complexity and users' attention and attitude toward the home page. But in contrast to this study and the study of Panir et al. (2006), Geissler et al., instead of using existing startpages, used self-designed versions of the same startpage, which varied in their degree of complexity. By designing the stimulus material himself he could probably achieve a sufficiently broad range of complexity, and was able to also include less complex startpages. When using a sample of real startpages, however, no such inverted U-shaped curve could be found.

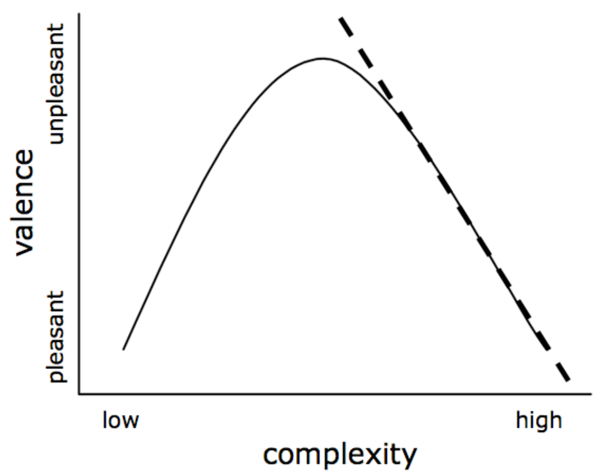


*Figure 6.* The expected quadratic relation between complexity and valence in the shape of an inversted U-curve and the empirically found linear relation between complexity and valence.

When looking at the psychophysiological responses there is no clear response pattern that can be related to visual complexity. Nevertheless, HR deceleration was positively correlated with visual complexity and corrugator EMG was negatively correlated to the affective valence ratings, while visual complexity was also correlated negatively to affective valence ratings. These results correspond in part to the findings of Lang et al. (2005) on the IAPS pictures, in that unpleasant affective pictures provoke a greater HR deceleration and higher corrugator EMG activity. But according to the affective space model of Lang et al. (1998), an increase in affective valence (negative or positive) should be accompanied by an increase in arousal ratings and physiological arousal measured by EDA. Our data only show a negative linear relationship between affective valence and arousal ratings and no systematic increase of physiological arousal (EDA) related to visual complexity or affective valence. Furthermore, there also was no connection between physiological arousal and self-rated arousal. However, when comparing response magnitudes of websites with those to highly arousing IAPS pictures, websites mostly fall into the neutral stimuli area of the affective space, which is a likely explanation for the lack of physiological arousal response toward websites. The negative correlation of valence and arousal ratings raises the suggestion that arousal in context of websites is associated more with an unpleasant experience than with a pleasant one. So, when trying to fit our websites in the affective space of the IAPS pictures, they would probably occupy a neutral to slightly unpleasant area, in accordance with the physiological pattern we found. There is a problem for the valence associated physiological response pattern, however. Despite our expectation, HR decrease was not correlated with affective valence ratings and can so hardly be considered as a measure of perceived affective valence toward startpages. This leads to the conclusion that the impact of visual complexity on HR decrease is not related to an affective response.

An alternative explanation could be that the observed HR decrease is cognitive in nature. Changes in HR upon stimulus perception can also be associated with an attention process such as the orienting reflex (OR). The OR is characterized by a decrease in HR and a simultaneous increase in arousal (Graham & Clifton, 1966; Öhman, Hamm & Hugdahl, 2000). Moreover, Fredrikson and Öhman (1979) could relate stimulus complexity to the OR by showing that more complex stimuli produce more HR deceleration overall. The data of the present study are consistent with these findings: complexity was correlated with HR decrease and arousal ratings. But there is another component of complexity that is related to an affective response pattern. The fact that visual complexity was also related to EMG response and to valence and arousal ratings raises the assumption that complexity also triggers an affective reaction in startpage perception. Increasing startpage complexity provokes a stronger EMG response, higher arousal ratings and more negative valence ratings. This pattern is in line with the findings of Lang et al. (2005) in regard to the concept of affective space. So it can be assumed that visual complexity of startpages has an impact on both components: emotional perception such as affective valence and cognitive processing such as OR.

The function of the VST and the recognition task was to investigate possible effects of visual complexity on users' performance. The results from the VST showed that visual complexity influences the search performance on the startpages. The correlation between visual complexity and RT reached a significant level, the relationship between these two variables indicates that increasing complexity leads to poorer search performance. Physiological responses during the VST did not show a clear pattern and could not be related to visual complexity or SAM ratings. But there was a relationship between RT and EDA as well as HR increase. Advanced search time on the startpages led to increased EDA response and greater HR acceleration. These responses make sense when we consider that participants were advised to accomplish the task as fast as possible and with advancing search time participants became more either more frustrated or motivated to find the asterisk, which resulted in an increase in EDA and HR acceleration. In the delayed recognition task, the recognition rate of the startpages could be related to visual complexity, valence ratings and arousal ratings. More complex startpages that were rated as less pleasant and more arousing achieved poorer recognition scores than less complex startpages that were rated as more pleasant and less arousing.

It should be noted that the presented results are from a classical picture perception task often used in traditional psychophysiological research and occurred under a strictly controlled setting and standardized experimental conditions. Considering these circumstances, it is difficult to associate the findings with how websites are perceived in real life. As the name implies, in real HCI situations (such as browsing the web) it is actual interaction, rather than passive viewing, which is most influential in shaping overall user experience. Nevertheless, when first visiting a website, initial impressions are formed by visual inspection—a process which corresponds well with the picture viewing task we employed in our study. In addition, the VST implemented a standardized prototypical model of active user-interaction with a real website.

As previously mentioned Ward and Marsden (2004) bring up the difficulty of demonstrating statistically significant psychophysiological responses in HCI situations and of interpreting the found physiological patterns. They attribute this issue to the lack of rigorously controlled experimental settings and tightly controlled experimental conditions in HCI research. To address this problem in the present study, we implemented two novel methodological features for HCI: The classic picture perception paradigm often used in traditional phsychophysiological research, and the VST as a standardized search task on websites. With the use of these methods we obtained statistically significant psychophysiological responses that could be clearly related to a specific factor such as visual complexity. These methods hold potential to be utilized productively in future psychophysiological HCI research.

*Conclusion*

This study clearly identified visual complexity as having an impact on valence and arousal judgments, HR and corrugator EMG responses and recognition memory performance of the participants and could thus establish it as an important factor for startpage perception. Further research is needed to clarify what elements in website design cause visual complexity and how complexity is related to HCI aspects like usability in websites. We believe that visual complexity provides a meaningful aspect for the evaluation of websites. Especially compressed file size (JPEG) could act as an easily assessable proxy for perceived visual complexity. However, current research on the concept of visual complexity in website design and its effects on users' perception is rudimentary and warrants broader attention.

References

Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. London: McGraw-Hill Book Company.

Berlyne, D. E. (1974). *Studies in the new experimental aesthetics—steps toward an objective psychology of aesthetic appreciation.* Washington: Hemisphere Publishing.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 602-642). New York: Cambridge University Press.

Bradley, M. M., Greenwald, M. K., & Hamm, A. O. (1993). Affective picture processing. In N. Birbaumer, & A. Öhman (Eds.), *The structure of emotion* (pp. 48–68). Seattle: Hogrefe & Huber.

Bruner, G. C., & Kumar, A. (2000). Web commercials and advertising hierarchy-of-effects. *Journal of Advertising Research, 40*, 35-42.

Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50*, 260-268.

Cacioppo, J. T., Berntson, G. G., Larson, J. T., Poehlman, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions*. (pp. 173-191). New York: Guilford.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology, 27*, 481-494.

Donderi, D.C., & McFadden, S. (2005). Compressed file length predicts search time and errors on visual displays. *Displays*, *26*, 71-78.

Donderi, D. C. (2006). An information theory analysis of visual complexity and dissimilarity. *Perception, 35*, 823-835.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23*, 567-589.

Geissler, G. L., Zinkhan, G. M., & Watson, R. T. (2006). The influence of home page complexity on consumer attention, attitudes, and purchase intent. *Journal of Advertising, 35*, 69-80.

Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin, 65*, 305-320.

Hazlett, R. (2003). Measurement of user frustration: A biologic approach. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems, 05-10 April 2003* (pp. 734-735). Ft. Lauderdale, Florida, USA: ACM Press, New York, NY.

Hudlicka, E. (2003). To feel or not to feel: The role of affect in human-computer interaction. *International Journal of Human-Computer Studies, 59*, 1-32.

Kaplan, S., Kaplan, R., & Wendt, S. (1972). Rated preference and complexity for natural and urban visual material. *Perception and Psychophysics, 12*, 354–356.

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

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion and motivation: Measuring affective perception. *Journal of Clinical Neurophysiology, 15*, 397-408.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical report A-6).

Gainesville, USA: University of Florida.

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

Lang, P. J., & Hnatiow, M. (1962). Stimulus repetition and the heart rate response. *Journal of Comparative & Physiological Psychology, 55*, 781-785.

Lippold, O. C. J. (1967). Electromyography. In P.H. Venables, & I. Martin (Eds.), *Manual of psychophysiological methods* (pp. 243-286). New York: Wiley.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129*, 242-261.

Öhman, A. (1979). The orienting response, attention, and learning: An information processing perspective. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.) *The orienting reflex in humans* (pp. 443-471). Hillsdale: Erlbaum.

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation and conditioning. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 533-575). Oxford: Oxford University Press.

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

Partala, T., Jokiniemi, M., & Surakka, V. (2000). Pupillary responses to emotionally provocative stimuli. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, 06-08 November 2000* (pp. 123-129). Palm Beach Gardens, Florida, USA: ACM Press, New York, NY.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*, 185-198.

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

Psychology Software Tools. (2002). *E-Prime (Version 1.2)* [Computer software]. Pittsburgh, PA: Author.

Riglis, E. (1998). *Modeling visual complexity in image architectures* (PhD First Year Report). Edinburgh, U.K: Heriot-Watt University.

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.

Schmutz, P. (2004). *eRes online Forschung* [eRes online research]. [Web application]. Basel: http://webserver.psycho.unibas.ch

Stevenson, J. S., Bruner, G. C., & Kumar, A. (2000). Webpage background and viewer attitudes. *Journal of Advertising Research, 40*, 29-34.

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.

Wilhelm, F.H. & Peyk, P. (2005*). Autonomic Nervous System Laboratory (Version 4.0)* [Computer software]. SPR Software Repository: http://www.sprweb.org.

Winton, W. M., Putnam, L. E., & Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology, 20*, 195-216.