Spatial frequencies in aesthetic website evaluations – explaining how ultra-rapid evaluations are formed

Meinald T. Thielsch[1] and Gerrit Hirschfeld[2]

[1]*Department of Psychology, University of Münster, Fliednerstr. 21, D-48149 Münster, Germany*
[2] *Vodafone Foundation Institute and Chair for Children's Pain Therapy and Paediatric Palliative Care, Datteln, Germany*

Address for correspondence:
Meinald T. Thielsch, University of Münster, Department of Psychology, Fliednerstr. 21, 48149 Münster, Email: thielsch@uni-muenster.de

Abstract
The current study investigates how aesthetic website evaluations, especially those formed after very brief presentations, depend on visual information that is encoded in low or high spatial frequencies. A total of 92 participants took part in the experiment. The study used a 3x3 mixed design in which presentation time (50, 500, and 10000 ms) and spatial filtering (low-pass filtered, high-pass filtered, and unfiltered stimuli) were manipulated. First, we replicate prior results from online studies of high- and low-spatial frequencies. Second, we confirm a prediction from neurocognitive models that only low-spatial frequencies are relevant to aesthetic judgements in ultra-rapid presentation modes. Third, we demonstrate that stimulus repetitions lead to an overestimation of the importance of ultra-rapid stimulus presentations. Taken together, our results highlight the utility of neurocognitive models of visual processing to explain the rapid aesthetic evaluation of websites.

Keywords: aesthetics; first impression; spatial frequency; repetition priming; website evaluation; website perception

Practitioner summary

Using neurocognitive models we present an approach to explain how aesthetic impressions are formed. We show that ultra-rapid judgements are connected with low but not with high spatial frequencies, which are neurologically processed in different visual pathways. Furthermore we identify possible methodological problems in previous studies of ultra-rapid website perception.

This is an Author's Accepted Manuscript of an article published in Ergonomics, 55 (7), *copyright Taylor & Francis, available online at:* http://www.tandfonline.com/10.1080/00140139.2012.665496

Please cite this article as:

Thielsch, M. T. & Hirschfeld, G. (2012). Spatial frequencies in aesthetic website evaluations – explaining how ultra-rapid evaluations are formed. *Ergonomics*, 55 (7), 731-742. doi: 10.1080/00140139.2012.665496

1. Introduction

There are billions of websites from many different domains, not only entertaining but often presenting information used for private and work-related issues as well. The evaluation of websites and understanding of the processes behind those evaluations has become an important research venue in the last two decades. In doing so, it has been increasingly recognized in the field that web user's needs go beyond usability and utility, leading toward a more experiential perspective (e.g., Hassenzahl and Tractinsky 2006, International Organization for Standardization 2009, Liu 2003). This perspective takes the whole user experience into account, including user perceptions of visual aesthetics. Thus, there has been a general trend to look beyond pure instrumental factors in ergonomics (e.g., Lindgaard and Whitfield 2004, Liu 2003, Mack and Sharples 2009), that is also present within the domain of website design (e.g., Lavie and Tractinsky 2004, Moshagen and Thielsch 2010, Schmidt *et al.* 2009). Various studies have shown the importance of website aesthetics in human-computer interaction and its impact on several constructs, such as perceived usability, satisfaction or trustworthiness (for an overview see Moshagen and Thielsch 2010, p. 691). The current research suggests that aesthetic responses occur immediately at first sight (Leder *et al.* 2004) and have an important impact on first impressions (e.g., Lindgaard *et al.* 2006 & 2011, Thielsch *et al.* under review, Tractinsky *et al.* 2006). This is of high practical relevance, as users base their decision about whether a particular website is explored deeper or another one is searched for, on these first impressions. In this paper we take a neurocognitive approach to explain *how* these immediate aesthetic impressions are formed.

1.1 Processes underlying aesthetics evaluations

There is a long and partly ongoing discussion about what constitutes aesthetics (for a review, see Moshagen and Thielsch 2010). In the current research, we follow the interactionist perspective and the definition of aesthetics given by Moshagen and Thielsch (2010, p. 690) by regarding aesthetics "as an immediate pleasurable subjective experience that is directed toward an object and not mediated by intervening reasoning". However, little research is concerned with the processes that underlie aesthetic evaluations. At present theories on aesthetic perception are, in general, cognitive in nature (for a review, see Martindale 2007), and they mostly agree on the relevance of fast, unconscious processes that determine whether a stimulus is perceived as more or less aesthetically pleasant (e.g., Leder *et al.* 2004, Zajonc 1980). The *processing fluency theory* (Reber *et al.* 2004) states that the more fluently a perceiver is able to process an object, the more positive will be her or his aesthetic response. Anything can be beautiful – as long as a perceiver finds it easy to process. This integrates various factors that affect aesthetic perceptions into a common framework. For example Bauerly and Liu (2006) found that increasing the number of objects in a visual display leads to worse aesthetic ratings. As less objects are easier to process processing fluency theory provides an explanation for this fact. Processing fluency can also explain why prototypical objects are generally preferred over non-prototypical objects (e.g., Martindale and Moore 1988, Winkielman *et al.* 2006). Critically, the processing fluency theory fits very well into the interactionist perspective, stating that characteristics of an object and of the perceiver interacting with it determine the aesthetic appraisal. Thus, the influence of object properties on aesthetic evaluations is mediated by processing fluency, which also depends on certain characteristics of the perceiver, such as learning history. For example, it has often been demonstrated that a repeated exposure to stimuli results in more-favourable evaluations (e.g., Zajonc 1968). The processing fluency theory provides not only a common framework to describe the factors that impact on aesthetic appraisal, but also allows to formulate and test novel predictions about the factors that are important for website appraisal.

As the makeup of our visual system constrains the kinds of information that can be processed and the speed at which that can be accomplished (Marr 1982, Goldstein 2009), neurocognitive knowledge is also relevant to website perception. The process of visual processing is mostly a bottom-up process. When light falls onto the retina specific photoreceptors (rods and cones) the visual pattern is translated into the discharge of neurons. This neuronal discharge triggers action potentials in retinal ganglion cells that in-turn activate cells in the lateral geniculate nucleus where they are relayed and are carried to the posterior pole of the occipital cortex. This region of the brain is known as the primary visual cortex (V1) where cells encode the presence or absence of edges. From there on, higher cortical visual areas each encode specific aspects of the visual information, e.g. movement is encoded in visual area V5 (which is also known as area MT because of it's medial temporal location) or colour is processed in visual area V4 that is part of the ventral stream.

In addition to this bottom-up flow of information, the visual system uses top-down information to enhance the efficacy of the whole process. In order to do that information processing in the visual system starts by rapidly (within about 100 ms) extracting the overall gist of a visual scene (Shyns and Olivia 1994) and then using this information to facilitate the detailed analysis of individual objects (Bar *et al.* 2006). The overall layout of a scene is encoded in the low spatial-frequencies (while high spatial frequencies encode fine details about objects). Such low-spatial frequency (LF) information is neurally projected via the very fast magnocellular pathway. Magnocellular neurons are very sensitive for movements, depth, and small differences in brightness. The slower parvocellular neurons in contrast, are particularly sensitive for colours, forms, and fine details. LF information triggers top-down connections that guide the analysis of high spatial frequency (HF) information (Bar *et al.* 2006, Kveraga *et al.* 2007).

As these complex feed-forward and feedback connections are a basic feature of the visual system these relationships are similar in all human beings and can be exploited in a variety of domains. Interfering with this system, e.g. by removing LF from visually presented stimuli with a high-pass filter abolishes top-down facilitation-effects (Hirschfeld and Zwitserlood 2011). The fact that core aspects of website design like colour, contrast or animations are processed via different cellular pathways within the visual system, makes it possible to separate these two kinds of spatial information by applying spatial filters to given visual stimuli. This enables studying the effects of various parameters in isolation: By using a low-pass-filter, details that are encoded in high spatial frequencies can be removed. And by applying a high-pass-filter, information about the global layout can be removed (for an example of spatial filtered screenshots, see figure 1).

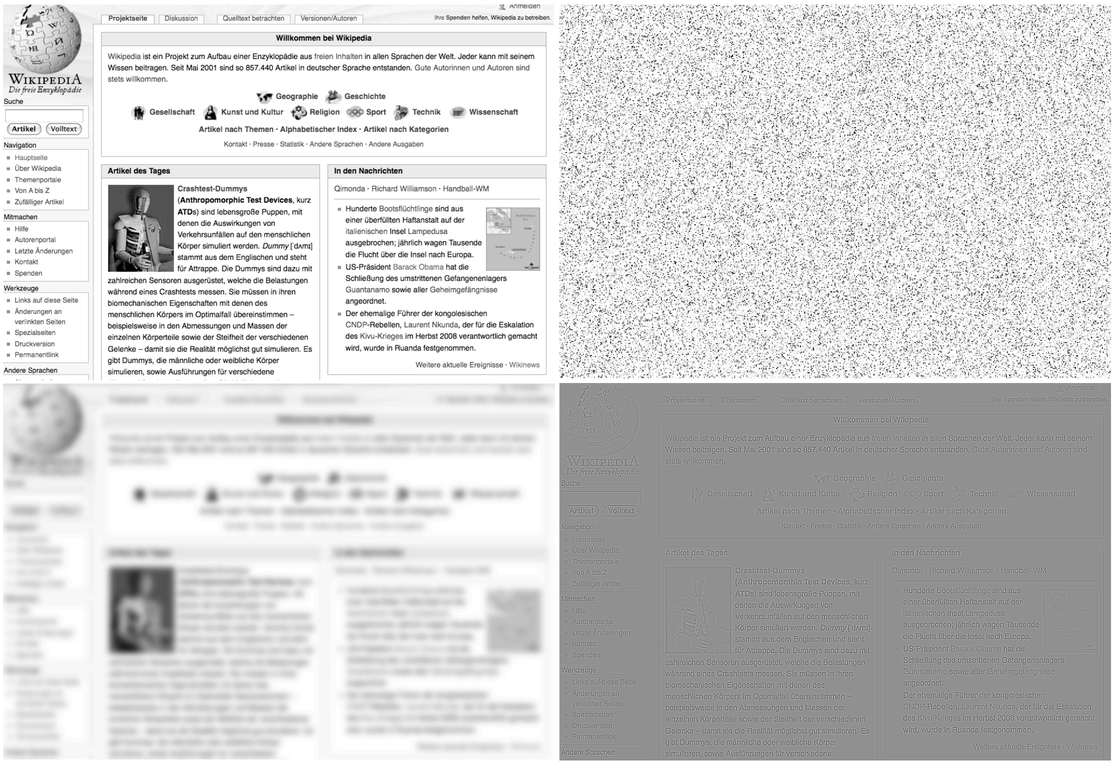

Figure 1. Examples for website stimuli used in the experiment. Top left: unfiltered screenshot; top right: mask for unfiltered screenshot; bottom left: low-pass filtered screenshot; bottom right: high-pass filtered screenshot.
Note: Contrast was partly adjusted to optimize for print.

It is known for some time that software icons, that are discernable on the basis of low spatial frequency information alone, can be interacted with more efficiently (Queen 2006). Due to the above-mentioned link between processing fluency and aesthetics these factors also matter for aesthetic appraisal. To test this hypothesis Thielsch and Hirschfeld (2010) conducted an online study to investigate the role of spatial frequencies in website evaluation. They found high correlations between ratings of high-pass or low-pass filtered screenshots and the original unfiltered versions. Because HF contain more information about the website, they found high correlations between high-pass and original websites in terms of usability and aesthetic ratings. In a regression analysis, they found a unique contribution of LF information on aesthetic evaluations. This finding confirmed a central prediction of the processing fluency theory: Low spatial frequencies are easier to process and therefore should influence aesthetic website ratings. As this was an online study, the timing could not be controlled. However, as HF information is processed much slower than LF information we could predict that HF impact would be diminished in ultra-short presentations (Thielsch and Hirschfeld 2010). To sum up, based on the processing fluency theory, spatial frequencies could provide an approach to investigate the processes behind ultra-rapid aesthetic evaluations.

1.2 Prior research using ultra-rapid presentations

Several studies have already used ultra-rapid presentations to study website perception and the impact of aesthetics on first impressions (for an overview see Tuch *et al.* under revision). This research was initiated by studies done by Lindgaard and colleagues (2006). They reported an experiment where 50 website stimuli were shown twice – one group for 50 ms and a second one for 500 ms in each presentation phase. Correlations between both conditions for each phase were remarkably high (r = .95 as well for phase 1 as for the second stimuli presentation, both ps < .01). Lindgaard *et al.* (2006) concluded that reliable decisions about website design could be made within 50 ms and that judgments of visual appeal could represent a mere exposure effect (according to Zajonc 1980). Generally, they stressed the very high importance of immediate first impressions for web designer.

Tractinsky *et al.* (2006) replicated and extended these initial results. They asked participants to evaluate 50 website screenshots and found very high correlations (*r* = .92) between ratings given after brief exposures in a first testing phase (500 ms) and ratings given after a second longer exposure phase (for 10 s). Tractinsky *et al.* (2006, p. 1080) stated that first aesthetic impressions are not solely responsible for users' attitudes toward a website but that "there is no second chance to make a first impression". These results could not be completely replicated in a study by van Schaik and Ling (2009), further investigating the stability of website aesthetics evaluations over time: They used a design similar to that of Lindgaard *et al.* (2006) and Tractinsky *et al.* (2006), in which participants evaluate websites first after being shown a screenshot for 500 ms and then after actually interacting with the website in different task contexts (action and goal modes) and forming a more deliberate aesthetic evaluation. Interestingly, they found only small correlations between aesthetic ratings in both time conditions (.04 ≤ r ≤ .32). In the second experiment (using a fictitious psychology website) correlations between the two time conditions were higher (.18 ≤ r ≤ .54), especially in an action mode (.48 ≤ r ≤ .54) but not in the range of the previously reported findings. This indicates that effects of early impressions on later judgments could be less stable than previously believed. Lindgaard and colleagues (2011) themselves found in a recent replication of their first studies a lowered correlation (of r = .73).

Furthermore the early studies of ultra-rapid website perception suffer from two methodological problems that by themselves might have lead to an overestimation of the importance of short presentation: (1) lack of masking and (2) repeated stimulus presentations. First, lack of masking, i.e. the presentation of a visual stimulus ("mask") shortly after the target is necessary to control the effective presentation time of the stimuli. Specifically, in very brief stimulus durations (≤ 50 ms) masks have to be used to prevent the formation of afterimages (e.g., Breitmeyer 2007, Enns and Di Lollo 2000). Without masking, the visibility of a stimulus is approximately 250 ms longer than intended by the researcher (Goldstein 2009). Second, repeated stimulus presentations – or within-subject designs – have been used in most of the above-mentioned studies, i.e. one participant rated the same website twice. This could lead to a high estimation of consistency, because participants have the tendency to be consistent in their own judgments and thus try to repeat or converge in their own statements. Furthermore, repetition typically improves the processing of repeated stimuli. Such repetition priming effects may be short-lived but have a huge impact on stimulus processing. Repetition priming effects manifest themselves for example in greater accuracy in identifying stimuli that are presented for very short durations (for an overview see Grill-Spector *et al.* 2006). Thus, if one is interested in the importance of first impressions one needs to estimate the effects of repetitions. We did this in our current experiment by using Latin square design in which participants are pseudo-randomly assigned to one of three different orders, in which the blocks are presented.

In doing so, both within-participant as well as between-participant analyses (by only analyzing the first experimental block) can be made.

1.3 Research question and hypotheses

In this study, we would like to explore the impact of spatial frequencies on users' first aesthetic impressions. So far it is not clear whether judgments of websites after ultra-rapid presentations are based on LF or HF information. The goal of the present study was to investigate effects of spatial filtering while adopting most of the different experimental time conditions used so far (Lindgaard *et al.* 2006 and 2011, Tractinsky *et al.* 2006) in one design and with control over possible effects of repeated stimulus presentation.

Our first goal was to replicate earlier results from a study on the role of LF and HF in free-viewing conditions (Thielsch and Hirschfeld 2010), which applied the top-down feedback model (Bar *et al.* 2006) to website perception. Specifically, we wanted to replicate the high correlations between low-pass filtered, high-pass filtered and unfiltered screenshots at long presentation durations with a different sample and under more tightly controlled conditions (hypothesis 1).

Our second goal was to test, how important the LF- and HF-route are for ultra-rapid aesthetic website evaluations. Based on the spatial frequency model, we expected systematic correlations between unfiltered websites and LF websites and no systematic correlation to HF websites for very brief presentations (50 ms) (hypothesis 2).

Our third goal was to control the effects of repeated stimulus presentation to the same participant, which occur in within-subject paradigms and could lead to inflated estimates for rapid presentations. We expected that correlations between the presentation modes are always higher in within-subject comparisons, compared to data from between-subject comparisons (hypothesis 3).

2. Method

2.1 Participants

A total of 92 volunteers (81.5 % female) participated in the study. Ages ranged from 19 to 34 years (M = 22.12; SD = 3.17). Participants completed the study on an anonymous basis and received course credits. All of the participants were students of the University of Münster; 90 of them were majoring in psychology, and 2 were minoring in the subject. All of the subjects had used the Internet before. On average, they had been using the Internet for 8.44 years (Min = 4, Max = 15, SD = 2.25) and 12.5 hours a week (Min = 3, Max = 42, SD = 7.22).

2.2 Stimulus material

We used the same stimulus set as in our prior study (Thielsch and Hirschfeld 2010), consisting out of 50 screenshots of websites from 10 different content domains. Specifically, websites were selected to represent a wide range of corporate and institutional websites in Germany, including communication and community websites, corporate sites, e-commerce, e-learning, e-recruitment, entertainment, information sites, search engines, social software, and web portals. Readers can refer to Thielsch and Hirschfeld (2010) for a more detailed description of this categorization scheme for websites and this specific set of stimuli. Screenshots of these websites always showed the index page for the site and were scaled to 1024x768 pixels (so that they subtended a maximum of 24.2 degree visual angle if seen from a distance of 70 cm and with a stimuli width of 30 cm). In the three experimental groups, different versions of the same screenshots were used. The screenshots were adapted using Adobe Photoshop CS3 Extended. Unfiltered website screenshots for the first group were transformed to greyscale to exclude the selective effects of colour (high- and low-pass filtering distorts colour). Low-pass filtering (for the second group) was performed using a Gaussian blur filter with a 6.1 pixel kernel. For the high-pass filter (used on the stimuli for the third group), we set the high-pass filter to a radius of 0.3 pixels.

Target screenshots were sandwich-masked by a scrambled version of themselves. Masks were created by decomposing the targets into pieces with a size of 2x2 pixels that were randomly rearranged using Matlab (Version 7.8).

2.3 Procedure

The experiment consisted of two parts: (1) the collection of aesthetic ratings and (2) the collection of familiarity ratings. Participants completed the first part of study in a computer lab in groups of up to six subjects. Stimuli were presented at a resolution of 1280x1024 pixels on 19 inch LCD Displays, connected to IBM-PCs (2.19 GHz ACPI Uniprozessor PC, 1 GB RAM) running Inquisit (Version 3).

We used a 3x3 mixed design: Participants were randomly assigned to one of three conditions: 30 participants into the first group (unfiltered screenshots), 32 in the second group (low-pass filtered screenshots) and 30 into the third group (high-pass filtered screenshots). Thus, spatial filtering was treated as between group factor, while presentation time (see below) as within factor. There were no significant differences between the three groups with respect to sociodemographic variables (age, gender, occupation, internet-usage). The procedure was the same for all of the three experimental groups.

After some initial information about the study, each of the participants was presented with the screenshots along with one item measuring perceived aesthetics. Each participant completed three blocks consisting of all 50 website screenshots of which the first five of each block were warm-ups, that were not analyzed further. Within each block, the presentation time was set to either 50, 500 or 10000 ms. The order in which these durations were presented to each participant were pseudo-randomly distributed across the participants according to a Latin Square. The presentation of the screenshots was randomized within each experimental block for each participant and block separately. This randomisation was used to avoid systematic errors while presenting stimuli in a fixed order (Liu and Salvendy 2009).

Each trial started with a fixation cross-presented in the middle of the screen for 500 ms, followed by the presentation of the mask for 50 ms; directly after the offset of the mask, the screenshots were presented (for either 50, 500 or 10000 ms); and directly after the offset of the screenshot, the mask appeared for 50 ms. As soon as the backward mask was removed, participants indicated their aesthetic evaluations on a seven point Likert scale ranging from "very unaesthetic", "unaesthetic ", "rather unaesthetic ", "neutral", "rather aesthetic", "aesthetic " to "very aesthetic " by clicking the label using the computer mouse (see figure 2). Participants had unlimited time to rate the websites, but extremely fast or slow ratings were excluded from the analysis (10 % of the fastest and slowest ratings). Participants in the first group (unfiltered screenshots) took on average 1720 ms (*SD* = 966) to rate a website screenshot. Participants in the second group (low-pass filtered screenshots) were slightly slower in rating the screenshots (*M* = 1721, *SD* = 1098), whereas subjects in the third group (high-pass filtered screenshots) were the fastest (*M* = 1636, *SD* = 1088). After a post-trial delay of 400 ms, the next trial started. Completing the first part of the study took approximately 20 minutes.

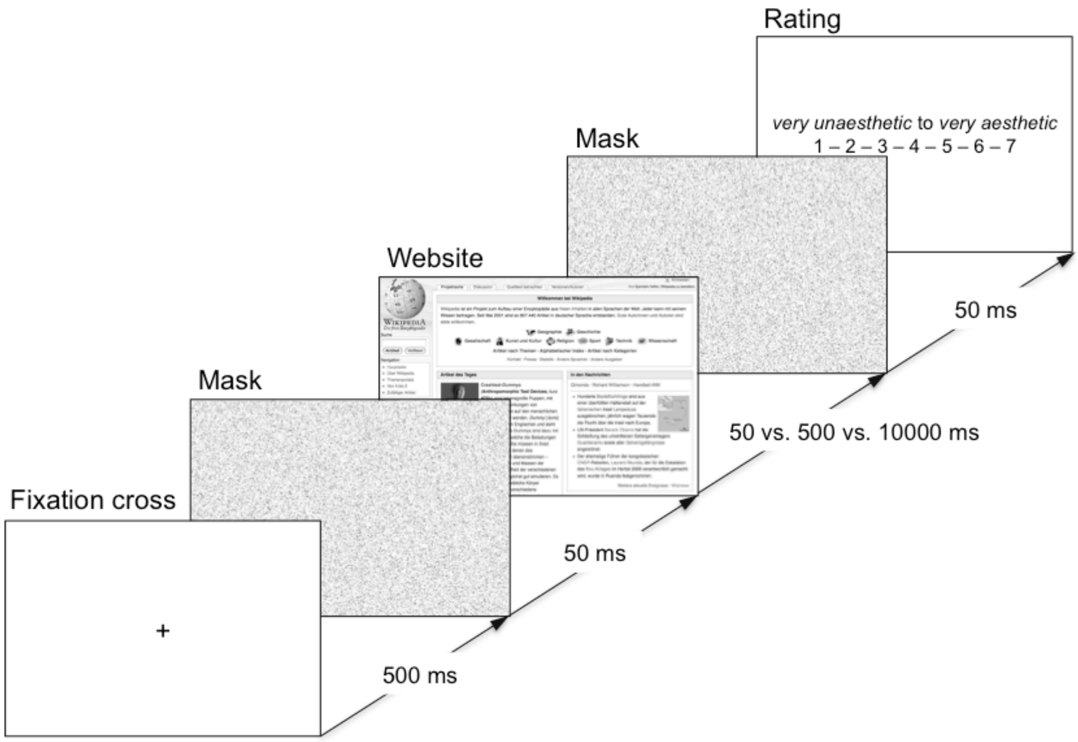

Figure 2. Procedure used in the experiment (illustration of a single trial).

After the first study part, each participant was given access to an online questionnaire built with EFS Survey (Version 7.1). The second part of the study was not conducted under controlled conditions. Each participant completed the questionnaire in his or her natural environment. The aim of this questionnaire was, on the one hand, to provide demographic background and on the other hand, to test the popularity of the screenshots used in part one of the study. For the second purpose, each of the 50 original screenshots was presented again. First, the participants had to indicate for each website whether they knew it before taking part in the study. If participants answered yes, they had to indicate on a five point Likert scale how often they visited the website (labels ranging from “less than one time per month”, “about one time per month”, “several times per month”, “several times per week” to “daily”). As the pattern of correlations was similar for known and unknown websites, all responses were used for the analysis. Completing the online questionnaire took approximately 10 minutes.

3. Results

Before starting the main analysis, we checked our data for possible bias effects: The mean ratings in the different conditions (see table 1) indicated that there were no artificial factors that reduced variance in a specific condition, e.g. ceiling effects. To check for possible bias caused by demographic properties of our sample, we used linear regression to predict the aesthetic ratings: In doing so we found no influence of participants' gender, age or an interaction of these factors. Thus, our data are very feasible for further analysis.

Table 1. Means and standard deviations (in brackets) of the raw-ratings in the experimental conditions.

	Unfiltered	*Low-pass filtered*	*High-pass filtered*
50 ms	3.94 (1.31)	3.81 (1.34)	3.62 (1.3)
500 ms	3.98 (1.38)	3.98 (1.34)	3.86 (1.35)
10000 ms	3.97 (1.49)	3.85 (1.4)	3.96 (1.37)

The main variables of interest are correlations between judgments of website under different conditions. As differences between participants are not important, correlations between different modes of presentation were calculated based on the average rating of all participants, which is the most important way to summarize this data to professionals designing websites (Monk 2004, Thielsch and Hirschfeld 2010). This resulted in a nine by nine matrix (see table 4). First, we would like to take a look at the replication of the results of Thielsch and Hirschfeld (2010) and afterwards at effects of spatial frequencies on ultra-rapid presentations.

3.1 Effects of spatial frequencies on long presentations

Our first goal was to replicate the findings by Thielsch and Hirschfeld (2010), who showed the importance of different spatial frequencies for website evaluations. Therefore, we used the data from the 10 s condition to test the impact of LF and HF on aesthetic judgments. We found a very similar pattern of correlations (see table 2, lower part). High-pass filtered screenshots correlated strongly (r = .76; p < .01) to unfiltered websites, while low-pass filtered stimuli correlated moderately with unfiltered ones (r = .53, p < .01). Overall the results resemble very closely the prior online data as expected by hypothesis 1.

Table 2. Correlations between aesthetic ratings of different screenshot versions in the 10 s condition (upper part = all experimental blocks, lower part = only first experimental block, which means a between subject comparison). In brackets correlations found by Thielsch and Hirschfeld (2010).

	Unfiltered	*Low-pass filtered*	*High-pass filtered*
Unfiltered	-	0.75**	0.81**
Low-pass filtered	0.53** (0.56**)	-	0.58**
High-pass filtered	0.76** (0.73**)	0.39** (0.33**)	-

Note. ** = p < .01

3.2. Effects of spatial frequencies on ultra-rapid presentations

Furthermore, we tested the impact of different spatial frequencies on very early aesthetic impressions of websites. For this we inspected the results from the 50 ms condition (see table 3). In the within-subject comparison, where effects of answer consistency could occur or repetition priming enhances perception processes, correlations between unfiltered and LF (r = .66, p < .01) as well as HF (r = .47, p < .01) filtered stimuli are high. In the between-subject comparison, that controls possible repetition effects in the within-subject design, only the correlation between unfiltered and LF stimuli was significant (r = .32, p < .05; see table 3), while there was no systematic relation to HF stimuli. This is in line with our hypothesis 2 and supports the idea that ultra-rapid judgements are not influenced by high spatial frequencies.

Table 3. Correlations between aesthetic ratings of different screenshot versions in the 50 ms condition (upper part = all experimental blocks, lower part = only first experimental block, which means a between subject comparison)

	Unfiltered	*Low-pass filtered*	*High-pass filtered*
Unfiltered	-	0.66**	0.47**
Low-pass filtered	0.32*	-	0.52**
High-pass filtered	0.05	0.16	-

Note. * = p < .05; ** = p < .01

3.3 Effects of stimulus repetition

For all pairs of presentation modalities, the correlations were higher when they were estimated from within-subject data (see table 4, upper half), i.e. ratings of the same stimulus by the same participant, compared to the analysis based on between-subject data. This was especially for the low-pass filtered stimuli and short presentations, where estimates for the correlations were twice as high as the corresponding estimate from between-subject data. As mentioned before, this points to effects of repeatedly presenting the same stimuli to the participants, which could lead to the implicit desire to give consistent responses and repetition priming. This result is in line with our third hypothesis.

3.4 Additional analysis

In addition, we inspected the overall pattern of correlations between all different time conditions and the spatial frequency-conditions in our study (see table 4). Two aspects are noteworthy:

1. Even though we found very low correlations with ultra-rapid presentations (50 ms), correlations between brief (500 ms) and long (10 s) presentations of website screenshots are always quite high, even in a between-subject condition when neither repetition priming nor consistency effects are possible. This is in line with prior research (Tractinsky *et al.* 2006) stressing the importance of first impressions made within the first 500 ms viewing a website.
2. High-pass filtered stimuli showed mostly higher correlations to unfiltered stimuli when the presentation time is longer than 500 ms. Furthermore, in the within-subject condition, one could find such high correlations even within the 50 ms condition. This can be interpreted as evidence for the increasing importance of high spatial frequencies when websites are shown repeatedly or for more than 50 ms. It is important to note that the latter point is a post-hoc interpretation that we did not have any a-priori hypothesis about, so this finding needs to be interpreted with great care.

Table 4. Correlations between aesthetic ratings in different time conditions and with different spatial filtering (upper part = all experimental blocks, lower part = only first experimental block, which means a between subject comparison).

		50ms			*500ms*			*10000ms*		
		Unfiltered	Low-pass filtered	High-pass filtered	Unfiltered	Low-pass filtered	High-pass filtered	Unfiltered	Low-pass filtered	High-pass filtered
50ms	Unfiltered	-	0.66**	0.47**	0.66**	0.50**	0.63**	0.60**	0.61**	0.57**
	Low-pass filtered	0.32*	-	0.52**	0.57**	0.62**	0.60**	0.57**	0.60**	0.60**
	High-pass filtered	0.05	0.16	-	0.60**	0.40**	0.80**	0.61**	0.52**	0.70**
500ms	Unfiltered	0.12	0.22	0.10	-	0.60**	0.70**	0.83**	0.64**	0.71**
	Low-pass filtered	0.39**	0.19	0.34*	0.29*	-	0.48**	0.50**	0.74**	0.46**
	High-pass filtered	0.14	0.26	0.43**	0.31*	0.41**	-	0.71**	0.59**	0.84**
10000ms	Unfiltered	-0.03	0.18	0.39**	0.53**	0.38**	0.55**	-	0.75**	0.81**
	Low-pass filtered	0.17	0.23	0.32*	0.48**	0.30*	0.25	0.53**	-	0.58**
	High-pass filtered	0.13	0.30*	0.45**	0.41**	0.31*	0.65**	0.76**	0.39**	-

Note. * = p < .05; ** = p < .01.; green = results for hypothesis 1, conditions as in Thielsch and Hirschfeld (2010); blue = results for hypothesis 2

4. Discussion

In this study we combined spatial frequency manipulation with presentation time variations. Our study revealed three main findings and confirmed our hypotheses: First, we replicated the previous results of Thielsch and Hirschfeld (2010) and found high correlations between aesthetic responses to low-pass filtered, high-pass filtered and unfiltered presented website screenshots. Second, we demonstrate a mediocre but robust effect of LF when stimuli are presented only once and very briefly for 50 ms. Third, we showed that within-subject designs could systematically overestimate the correlations between first impression ratings.

4.1 Low and high spatial frequencies

The human visual system constrains on the way with which visual information is being processed. As all humans use the same visual system, there may be a set of fundamental relationships that are true in any population under study, e.g. an increasing number of visual objects make an interface harder to understand (Bauerly and Liu 2006). The starting point for the present study was that, low and high spatial frequency information are processed differently. Specifically, LF convey only very global information, and are processed much faster than HF information. The latter convey very detailed information and is processed via a different nerve-tract (Goldstein 2009). Thus, with ultra-short presentation durations and masked presentation the visual system cannot extract any information from images from which the low-spatial frequencies are removed. In line with this, we did not find any correlation between unfiltered ratings and those based on HF in the 50 ms condition when looking at the first ratings only (as expected in hypothesis 2). HF screenshots were only correlated to unfiltered screenshots, when looking at the within-subject data. We believe the reason for this is that repetition enhances the visibility of shortly presented stimuli, essentially counteracting the effect of the masking (Grill-Spector *et al.* 2006). Thus, it might be that by the repeated 50 ms presentation of high-pass filtered screenshots, coarse attributes of the stimuli are used by participants to rate the stimuli. While this aspect needs to be investigated further, for the situation of a typical user visiting a website for the first time our findings suggests, his or her general rating after a very brief presentation like 50 ms seems to be less important than claimed in some previous studies (Lindgaard *et al.* 2006 and 2011).

However, the speed-advantage of LF information not only impacts on ultra-rapid presentations, but also influences perception in more natural conditions. Several studies have reported that low-spatial frequency-information can be used to guide the following more in depth analysis of the high-spatial frequency information in a visual scene (Bar et al, 2006, Shyns and Olivia, 1994). Accordingly removing low-spatial frequency information from visual stimuli reduces facilitative effects that are triggered by LF (Kverga *et al.* 2007, Hirschfeld and Zwitserlood 2011). As speed of processing is tightly related to aesthetic appeal (Reber *et al.* 2004), LF information is also critical for the judgment of aesthetic appeal.

HF information in contrast is only relevant when there is sufficient time for the visual system to access this kind of information. In our experiment the 500 ms condition made it possible to recognize HF information like rough edges and, most important, to read text on screenshots. As there are high correlations between content and aesthetic ratings of websites (de Wulf *et al.* 2006, Moshagen and Thielsch 2010) this might explain the high impact of HF information. Furthermore, we controlled for prior knowledge of stimuli, but there occurred no significant differences in the result pattern.

4.2 Repeated stimulus presentations

As expected in hypothesis 3, correlations in between-subject comparisons are always lower than the corresponding within-subject comparisons. The latter ones were very similar to previous studies mostly using within-subject methodology (Lindgaard et al. 2006 and 2011, Tractinsky et al. 2006). This may be explained by the fact that if participants are repeatedly confronted with a stimulus, they try to give similar responses. However, such effects of answer consistency alone are not sufficient to explain the effects in the 50 ms condition. Given the constraints of the visual system, such high correlations for masked stimuli are implausible. However, in this condition, repetition priming effects, which lead to more efficient stimulus processing, could have occurred. Such effects manifest themselves for example in greater accuracy in identifying stimuli that are presented for very short durations (Grill-Spector *et al.* 2006). Furthermore, repetition results in mere-exposure effects (Zajonc 1980). Thus, within-subject design with repeated stimulus presentation could result not only in consistency-effects but also in repetition priming and mere-exposure effects that influence the final judgment of a given visual stimuli.

4.3 Practical implications

What is the practical use of our findings? A designer creating a website would prefer to get practical hints which colour is best in the given context or which kind of forms are well fitting. But, from our point of view a detailed understanding of the processes underlying website ratings is very useful: The analysis of spatial frequencies, which are the base in visual perception of all other design variables, could lead us to general design principles without a laborious testing of the multitude of shapes, connections and interactions of typical web design variables. We believe that applied researchers and practitioners in the field can use spatial filtering when assessing the aesthetic appeal of newly developed websites and analysing first impressions and the important stage of users first contact with a new graphical interface. Of course, it is possible to just ask people about their first impression of a website or a website prototype, but such evaluations could be influenced by other aspects like content (Hartmann *et al.* 2008, Thielsch *et al.*, under review) and it is difficult to assess very early ultra-rapid impressions in this way. Furthermore, practitioners and most applied researcher usually do not have the needed resources in terms of lab facilities and adequate experimental software to test a website for immediate first impressions in the critical time phase of the first 500 ms a website is seen. Thus, it might be a solution just to present a low spatial filtered screenshot, which conveys only the early processed visual information and could so presented without time limitations. In doing so, one will get an idea of the very first visual impression of a website. Even when working on a new design, one could get an rough impression of it in terms of low spatial frequencies by using a “squint test” as described by Queen (2006), what means to squint your eyes to obstruct sharp focus and rely mostly on dark and light values. But at this point we agreed with Queen (2006), that it would be more practical to filter a screenshot with appropriate software. Such filtered screenshots, sketches or prototypes of a design could easily be made in common graphic editing programs and afterwards shown to test persons. Probably they will be of most use in a comparison task with other competing prototypes or websites.

4.4 Limitations and future research

Some limitations should be considered while interpreting the results of our present study. First, we used the same set of stimuli as in our prior study to spatial frequencies (Thielsch and Hirschfeld 2010). We found the same results in the lab as those before in a natural testing environment while asking a different sample, but one might still argue that there is a need for further cross-validation with a different set of stimuli. As we used a large set of websites from very different domains, we considered this aspect as a small potential source of bias. As shown in table 1, there are no ceiling effects within the ratings. In fact, the mean ratings were close to the middle of the answer scale and indicates a well-balanced selection of website stimuli.

Second, we used a single-item measurement to assess visual website aesthetics, which was well done in prior research on immediate aesthetic responses to website stimuli and is generally very common in user experience research (e.g., Hassenzahl 2004, Sonderegger and Sauer 2010, Tractinsky *et al.* 2000). Nevertheless, single-item measurement in this area is rightly criticized for reliability problems and concerns of adequate construct assessment (Moshagen and Thielsch 2010, p. 692). However, we averaged ratings across participants and used a large sample of stimuli in the hopes of reducing measurement error in a sufficient manner. Moshagen and Thielsch (2010) found a general aesthetic factor in website evaluation consisting of four sub-facets; thus, we tried to assess the general factor and the core of the construct as well as possible while using a Likert scale labelled from "very unaesthetic" to "very aesthetic". The overall clear pattern of our results across different conditions and subsamples indicates that this procedure was appropriate and successful. Furthermore, it might cause validity problems to let participants rate a website screenshot presented for only 50 ms with multi-item questionnaires like the one proposed by Lavie and Tractinsky (2004) or the VisAWI (Moshagen and Thielsch 2010). Such an assessment is prone to bias by recognition problems of aesthetic facets, halo effects or the poor motivation of participants who are not willing to answer a large amount of questions about a stimulus they have barely consciously recognized. Still, comparisons between such short assessments of aesthetics like those performed in the current study and more differentiated assessments of stimuli (presented in longer time conditions) are an important avenue for future research. Generally, it would be of much interest to further explore the relation between immediate and deliberate aesthetic judgments. Although a deliberate judgment is clearly impossible within 50 ms, high correlations between 500 ms and 10 s presentations leads to the suggestion that reflective processes of aesthetic responses start even within 500 ms. Future research should try to differentiate immediate aesthetic responses, reflective impressions and deliberate judgments. In this context, one should also compare free viewing to actual use conditions (as done by van Schaik and Ling 2009).

Third, the used website screenshots were presented only in greyscale. Given the high importance of colour to website aesthetics (e.g., Cyr *et al.* 2010, Kim *et al.* 2003, Moshagen *et al.* 2009, Moshagen and Thielsch 2010), this is a restriction of the external validity of the given results. However, the lack of colour does not influence the current findings' internal validity nor does it influence them with respect to the effects of spatial filtering or of repeated stimulus presentation. As colour is mainly processed via the parvocellular pathway (Goldstein 2009), one would predict that colour similarities exert an influence only within longer presentation durations.

Forth, all of the tested participants and the stimuli used shared the same cultural background. There is some empirical evidence for the influence of cultural and ethnic background in aesthetic design (Tractinsky 1997), especially for website aspects like colour and images (Cyr *et al.* 2009; Cyr *et al.* 2010) and compositional elements (Bi *et al.* 2011). The extent to which our findings are prone to cultural differences should be analyzed by a cross-cultural approach. Due to the conceptual closeness of our findings to mere neurophysiological processes, we would expect rather small cultural differences – especially as colour was already controlled in our study and therefore could not have influenced our results.

Furthermore, one might think that the sample used in this study is relatively small. But research focusing on such universal basic processes tend to be based on small samples - because until now no study has reported systematic differences between individuals for example in the density of visual pathway feedback connections. So far no individual person has been reported in the literature, which lacks of these feedback connections. Additionally we checked for systematic effects based on demographic properties without finding any hint to such bias. Nevertheless, a cross-validation of our results with another sample is valuable.

4.4 Summary and conclusion

To sum up, we confirmed the importance of spatial frequencies for website evaluations. We were able to analyse the impact of spatial frequencies on first impressions and showed that there are mediocre but robust effects that support the assumed special role of low-spatial frequencies for ultra-rapid aesthetic judgments. In doing so, we found evidence for the validity of the processing fluency theory. Furthermore, we showed effects of repeated stimuli presentation. Thus, control over effects like answer consistency and repetition priming is important for further research dealing with first impressions of websites. In general we confirmed in our data prior results that aesthetic judgments are formed at least within 500 ms and partly stable over time.

In the visual processing of a website the perception of first impressions seems to start with low spatial frequencies which are neurologically quickly processed and acted like a door opener. From this starting point top-down processes are triggered and deeper perception of a website, based on high spatial frequencies, is enhanced. Thus our results are not only highly relevant for basic research, but also for applicable for website testing procedures. Knowing of the importance of low spatial frequencies gives applied researches the opportunity to test website prototypes in terms of spatial frequencies using short presentation modes. In doing so one is able to differentiate analyse the first aesthetic impression of a website. Surely, this could be only one small aspect amongst others in website testing. But as aesthetics is highly relevant for first impressions and the website of a competitor is only one click away, such a differentiated analysis of a website in question may be worthwhile.

Acknowledgements
The authors thank Maren Hulisz for constructive suggestions and for her help testing participants as well as two anonymous reviewers for their very valuable comments to prior versions of this manuscript.

References

Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmidt, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., and Halgren, E., 2006. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Science*, 103 (2), 449-454.

Bauerly, M., and Liu, Y., 2006. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Studies*, *64*(8), 670-682.

Bi, L., Fan, X., and Liu, Y., 2011. Effects of Symmetry and Number of Compositional Elements on Chinese Users' Aesthetic Ratings of Interfaces: Experimental and Modeling Investigations. *International Journal of Human-Computer Interaction*, 27(3), 245-259.

Breitmeyer, B.G., 2007. Visual masking: past accomplishments, present status, future developments. *Advances in cognitive psychology*, 3(1-2), 9-20.

Cyr, D., Head, M., Larios, H. and Pan, B., 2009. Exploring human images in website design: A multi-method approach. *MIS Quarterly*, 33, 539–566.

Cyr, D., Head, M., and Larios, H., 2010. Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies* 68 (1), 1-21.

De Wulf, K., Schillewaert, N., Muylle, S., and Rangarajan, D., 2006. The role of pleasure in web site success. *Information & Management*, 43, 434-446.

Enns, J., and Di Lollo V., 2000. What's new in visual masking? *Trends in cognitive sciences*, 4(9), 345-352.

Goldstein, E.B., 2009. *Sensation and Perception*. 8th ed. Belmont: Wadsworth Publishing.

Grill-Spector, K., Henson, R., and Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences,* 10 (1), 14-23.

Hartmann, J., Sutcliffe, A., and De Angeli, A., 2008. Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction*, 15(4), Article 15.

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

Hassenzahl, M., and Tractinsky, N., 2006. User experience - a research agenda. *Behaviour & Information Technology* 25, 91-97.

Hirschfeld, G., and Zwitserlood, P., 2011. How vision is shaped by language comprehension - top-down feedback based on low-spatial frequencies. *Brain Research,* 1377, 78-83.

International Organization for Standardization, 2009. *ISO FDIS 9241-210: Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems.* International Organization for Standardization, Geneva.

Kim, J., Lee, J., and Choi, D., 2003. Designing emotionally evocative homepages: An empirical study of the quantitative relations between design factors and emotional dimensions. *International Journal of Human-Computer Studies* 59, 899–940.

Kveraga, K., Boshyan, J., and Bar, M., 2007. Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27 (48), 13232-13240.
Lavie, T., and Tractinsky, N., 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, 60, 269-298.
Leder, H., Belke, B., Oeberst, A., and Augustin, D., 2004. A model of aesthetic appreciation and aesthetic judgements. *British Journal of Psychology* 95, 489-508.
Lindgaard, G., Dudek, C., Sen, D., Sumegi, L., and Noonan, P., 2011. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. *ACM Transactions on Computer-Human Interaction*, *18*(1), 1-30.
Lindgaard, G., Fernandes, G., Dudek, C., and Brown, J., 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & Information Technology*, 25 (2), 115-126.
Lindgaard, G., and Whitfield, T. W. A., 2004. Integrating aesthetics within an evolutionary and psychological framework. *Theoretical Issues in Ergonomics Science*, 5 (1), 73-90.
Liu, Y., 2003. Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology. *Ergonomics*, 46 (13-14), 1273-1292.
Liu, Y., and Salvendy, G., 2009. Effects of measurement errors on psychometric measurements in ergonomics studies: Implications for correlations, anova, linear regression, factor analysis, and linear discriminant analysis. *Ergonomics*, 52(5), 499-511.
Mack, Z., and Sharples, S., 2009. The importance of usability in product choice: A mobile phone case study. *Ergonomics*, 52 (12), 1514-1528.
Marr, D., 1982. *Vision*. New York: Freeman and Company.
Martindale, C., 2007. Recent trends in the psychological study of aesthetics, creativity, and the arts. *Empirical Studies of the Arts*, 25 (2), 121-141.
Martindale, C., and Moore, K., 1988. Priming, prototypicality, and preference. *Journal of Experimental Psychology: Human Perception and Performance,* 14, 661-670.
Monk, A., 2004. The product as fixed effect fallacy. *Human-Computer Interaction*, 19 (4), 371-375.
Moshagen, M., Musch, J., and Göritz, A.S., 2009. A blessing, not a curse: Experimental evidence for beneficial effects of visual aesthetics on performance. *Ergonomics*, 52 (10)*,* 1311-1320.
Moshagen, M., and Thielsch, M.T., 2010. Facets of visual aesthetics. *International Journal of Human-Computer Studies,* 68, 689-709.
Queen, M., 2006. Icon analysis: Evaluating low spatial frequency compositions. *Boxes and Arrows* [online journal]. Available from: http://www.boxesandarrows.com/view/icon_analysis [Accessed 22 Aug 2011].
Reber, R., Schwarz, N., and 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.
Schmidt, K.E., Liu, Y.L., and Sridharan, S., 2009. Webpage aesthetics, performance and usability: Design variables and their effects. *Ergonomics*, 52 (6), 631-643.

Shyns, P.G., and Olivia, A., 1994. From blobs to bondary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5 (4), 195-200.
Sonderegger, A., and Sauer, J., 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics,* 41 (3), 403-410.
Thielsch, M. T., Blotenberg, I. & Jaron, R., in press. User evaluation of websites: From first impression to recommendation. *Interacting with Computers.*
Thielsch, M.T., and Hirschfeld, G., 2010. High and low spatial frequencies in website evaluations. *Ergonomics,* 53 (8), 972-978.
Tractinsky, N., 1997. Aesthetics and apparent usability: empirically assessing cultural and methodologicalissues. CHI 97 Conference Proceedings, Atlanta, 22-27 March, ACM, New York, pp. 115-122.
Tractinsky, N., Cokhovi, A., Kirschenbaum, M., and Sharfi, T., 2006. Evaluating the consistency of immediate asthetic perceptions of web pages. *Journal of Human Computer Studies*, 64, 1071-1083.
Tractinsky, N., Katz, A.S., and Ikar, D., 2000. What is beautiful is usable. *Interacting with Computers*, 13, 127-145.
Tuch, A.N., Presslaber, E.E., Stöcklin, M., Opwis, K., and Bargas-Avila, J.A., under revision. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments.
Van Schaik, P., and Ling, J., 2009. The role of context in perceptions of the aesthetics of web pages over time. *International Journal of Human-Computer Studies*, 67(1), 79-89.
Winkielman, P., Halberstadt, J., Fazendeiro, T., and Catty, S., 2006. Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17 (9), 799-806.
Zajonc, R.B., 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1-27.
Zajonc, R.B., 1980. Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.