

Nicholas A. Johnson. An investigation of the aesthetic fidelity of live websites. A Master's Paper for the M.S. in I.S degree. April, 2008. 42 pages. Advisor: Barbara Wildemuth.

The idea of "aesthetic fidelity" was developed to measure the degree to which affective messages and emotions intended by interface designers are communicated to users through the aesthetics of an interface. Until now, however, aesthetic fidelity research has been confined to highly artificial experiments. This study investigates the aesthetic fidelity of live, "real-world" interfaces. Two web designers were interviewed to discover the processes they use when designing websites and to compile a list of affective messages they tried to send in specific sites they created. A small user study was then conducted to find out how strongly these affective messages were received by users. Results from both phases of the research are used to draw conclusions on how the idea of aesthetic fidelity can be applied to real-world interfaces, how future aesthetic fidelity studies can be conducted and what tools designers can use to increase the aesthetic fidelity of their interfaces.

Headings:

Web sites/Design

Web sites/Psychological aspects

Use studies/Internet

AN INVESTIGATION OF THE AESTHETIC FIDELITY OF LIVE WEBSITES

by
Nicholas A. Johnson

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2008

Approved by

Barbara Wildemuth

Table of Contents

Introduction.....	2
Literature review.....	4
Methods.....	12
Phase 1: designer interviews.....	13
Phase 2: User study.....	15
Results of phase 1: designer interviews.....	17
Results of phase 2: user study.....	21
Discussion.....	25
Conclusion.....	32
Acknowledgements.....	33
References.....	34
Appendices.....	37
Appendix A: Rating questionnaire.....	37
Appendix B: User study directions.....	38
Appendix C: User-provided affective messages.....	39

Introduction

Until recently, human-computer interaction (HCI) research has not dealt much with interface aesthetics, a subject broadly defined as the visual design of a user interface and the affective messages communicated to users by that design. Early interest in the field of interface aesthetics came mostly from graphic designers thinking about how to apply their skills to a new, electronic medium (Faison, 1996). With a few exceptions, HCI researchers were generally happy to leave aesthetics to the designers in favor of studying other concepts like “usability” and “functionality” (Bertelsen & Pold, 2004).

Interest from the HCI community in interface aesthetics began to increase in the mid-1990s when a pair of researchers showed users’ perceptions of the usability of an interface were influenced by the interface’s visual design (Kurosu & Kashimura, 1995). That interest changed to outright enthusiasm in the late 1990s when Tractinsky (1997) not only replicated Kurosu and Kashimura’s results but found the link between an interface’s visual design and users’ perceptions even stronger than he expected.

Since Tractinsky presented his results in 1997, investigation into interface aesthetics by HCI researchers has become substantial. Study after study has proven the link between the visual design of an interface and the feelings the interface can evoke in users (Hoffmann & Krauss, 2004). Interface aesthetics have been shown to influence everything from system acceptability (Kurosu & Kashimura, 1995) to user

satisfaction, perceptions of system quality, ease of use and usefulness (Ben-Bassat, Meyer, & Tractinsky, 2006).

Not only does the visual design of a user interface play a significant role in how users perceive an interface, but it affects their performance as well. Norman (2004) nicely sums up the findings in the title of the first chapter of his book *Emotional Design*: “attractive things work better” (p. 17). The crux of Norman’s argument is that any object (including a user interface) that strongly appeals to users on three different levels—visceral, behavioral and reflective—will be more engaging to the user and thus help them perform better. Studies in interface aesthetics have shown that an interface’s visual design can especially appeal to users on the visceral and reflective levels and subsequently increase their performance. Cawthon and Moere (2007), for instance, found that interfaces used to complete data visualization tasks that were consistently rated by users as more aesthetic (i.e., as having more visceral appeal) also had low rates of task abandonment and high rates of correct responses. Similarly, Jansen, Zhang and Zhang (2007) found that information retrieval interfaces that sported branding elements from web search engines (like Yahoo! or Google) that users were familiar with (i.e., that had high reflective appeal) performed better than interfaces carrying unknown brands—even though the information presented in all the interfaces was identical.

Despite the preponderance of evidence for the importance of interface aesthetics, little has been done to help interface designers put the research into practice; practitioners have been largely left to their own devices to “abstract generalizable laws and interpret them in a design context” (Sutcliffe, 2002, p. 183).

This lack of a bridge between aesthetics research and practice has left designers in a quandary: they understand the importance of interface aesthetics but do not have a measure of what exactly makes for “good” aesthetics.

HCI researchers have recently taken steps to help close the gap between aesthetics research and practice. Park, Choi and Kim’s (2004) approach to closing that gap is particularly intriguing: they developed the measure of “aesthetic fidelity” to represent the degree to which affective messages and emotions intended by the designer are communicated to users through the aesthetics of an interface. However, their development of aesthetic fidelity has so far been only experimental, and research done on the topic has only been carried out in controlled, laboratory settings.

This study aims to explore the idea of aesthetic fidelity in a more practical domain by applying methods from previous research to real-world interfaces. Specifically, this exploratory study focuses on the most ubiquitous kind of interfaces, websites, and seeks to answer the following research questions:

1. Are designers of real-world websites successful in creating sites with aesthetic fidelity?
2. How well can the idea of aesthetic fidelity be applied to real-world, non-experimental interfaces?
3. What methods should be used in future studies of the aesthetic fidelity of live interfaces?
4. What methods can designers of live interfaces use to increase the aesthetic fidelity of their designs?

Literature review

To answer these research questions and explore the use of aesthetic fidelity measures on real-world websites, it is first important to more fully understand what

is meant by aesthetic fidelity, what kinds of affective messages might be measured by the concept and what tools designers might use to craft these messages.

Aesthetic fidelity

The concept of aesthetic fidelity was first developed in a pair of papers (Kim, Lee, & Choi, 2003; Park et al., 2004) stemming from the same multipart study. In these papers, the authors define “aesthetic fidelity” as “the degree of consistency between designers’ intentions and users’ impressions” (Park et al., 2004, p. 354) and begin to probe the idea through three experiments.

In the first of the three experiments, the authors met with web designers to brainstorm adjectives that could be used to describe a sample of 12 different homepages. The adjectives suggested in this brainstorming session (such as “dependable”) were used to create a survey of semantic differentials for each of the 12 homepages (seven-point scales ranging from “dependable” to “not dependable,” for example). The survey was then administered to over 400 web users who were asked to rate four of the 12 pages on each differential scale after looking (but not interacting with) the page for three minutes. The authors used cluster analysis on the survey data to create a list of 13 “emotional dimensions” that could be used to generally describe the affective content of a website: bright, tense, strong, static, deluxe, popular, adorable, colorful, simple, classical, futuristic, mystic and hopeful.

In the second part of the study the authors recruited 36 professional web designers. Designers were asked to choose one of the 13 emotional dimensions from the first part of the study and create a homepage that clearly communicated that dimension. Designers were first given two hours to create a prototype sketch of their

homepage. They were next asked to take three days to reflect on their prototype and to gather more ideas that could be used to refine the design. Finally, designers were given two hours to create a computerized mockup of their design using the prototype sketch and the other ideas they had gathered. When the prototype was complete, the authors used think-aloud data gathered during the design sessions to create a list of visual design factors (color, shape) that designers had employed to communicate the chosen emotional dimension.

In the final phase of the experiment, the authors repeated the survey from the first phase. This survey, however, used the homepages developed in the second phase as the test objects and the 13 emotional dimensions defined in the first phase as the survey questions. In essence, then, this final experiment tested how well designers could communicate specific emotional dimensions to users and what visual factors they used to communicate those dimensions. The data from this phase of the study showed a high degree of correlation between intended and received emotional dimensions. The authors were further able to use the list of visual design factors developed in the previous phase to examine which factors were most effective at communicating which emotional dimensions (survey respondents tended to rate a page high on the “bright” dimension when the shape of a menu bar was thick and had low saturation, for example).

While this study did show quite convincingly that aesthetic fidelity could be manipulated and that certain design factors were strongly associated with particular emotional dimensions, it suffered from one major shortcoming: the study was entirely laboratory based. Although the authors did work with practicing web

designers in parts of their study, the practitioners were artificially constrained in their design process. Further, the designers had to create homepages that communicated only one emotional dimension rather than develop a more true-to-life site that must convey many affective messages at once.

In order to assess aesthetic fidelity in a less artificial setting, then, it is important to first understand two areas of the problem: how to gauge affective messages in a live website and what methods web designers might have used to develop and refine these messages.

Determining affective messages

The kinds of affective messages that can be communicated by a live website are practically limitless: sites can communicate everything from the generic emotional dimensions developed in the original aesthetic fidelity study to messages specific to each site (“Google-ness,” for example). A number of studies have worked to operationalize and narrow the scope of affective messages to a manageable number.

In the studies mentioned above, lists of emotional dimensions were generated by working with experienced web designers (Kim et al., 2003; Park et al., 2004). Researchers used a two-hour long brainstorming session to elicit all the adjectives the designers could think of to describe any of the 12 homepages in the first phase of their study; this session yielded a list of 445 adjectives. The authors then worked to refine the brainstormed list by combining very similar adjectives under a more general heading and eliminating adjectives that were inappropriate for their study.

Finally, the authors met with another group of designers to confirm the validity of their refined list and make further adjustments.

Lavie and Tractinsky (2004) took a similar approach in their study to compile a list of “perceived visual aesthetics of web sites.” The authors first generated a list of 41 possible aesthetic dimensions through a combination of a literature review and meetings with various design professionals. Several user studies were then performed where users were asked to rate a sample of websites on each of the dimensions. The authors used the data from each round of user studies to refine the list for the subsequent study until they ended with a list of five apparently independent aesthetic categories: classic aesthetics, expressive aesthetics, usability, pleasurable interaction and service quality.

Park, Kim and Kim (2004) demonstrated another iterative technique for defining affective messages in their study on creating visual personalities on blogs. Their method, however, started with a huge number of possibilities: 1520 “basic personality adjectives from psychology, design and marketing literature.” They then met with a group of design professionals to trim the list to 1200 words, which were subsequently condensed into 181 personality adjectives using a survey. Finally, the authors grouped those 181 adjectives into 20 “distinctive cyber-personality dimensions.”

All of these studies take a common, iterative approach to identifying what affective messages websites are trying to communicate. First, a brainstormed list of messages is compiled. This list is then refined one or more times using either user studies or the judgments of design researchers and professionals. The refined list is

then checked for validity and further refined if needed. In this way, all of the above studies narrowed the nearly infinite “affective message space” into a list of messages that could be explored in a reasonable HCI study.

Crafting aesthetic fidelity

The final area that must be understood to investigate this study’s research questions is what methods web designers are likely to use to hone the aesthetic fidelity of their sites. There are presently no standardized, popular tools designed specifically to facilitate the aesthetic design process. There are, however, a number of visual design principles and common user-centered design tools that could conceivably be used by designers to help increase the aesthetic fidelity of the interfaces they create.

Graphic design literature provides guidance on which elements of visual design might be most important in creating and refining affective messages to be communicated through a website. Textbooks like Frascara (2004) and Williams (2004) run through a wide array of design methods and best practices, but most can be condensed under four headings: color, structure, imagery and typography. Each area is discussed in depth in graphic design texts, often with examples that, although not using the term specifically, speak to how the principles can be applied to communicate affective messages (a highly structured page can convey a feeling of “order,” for instance).

Several experimental studies confirm that these four areas of visual design—color, structure, imagery and typography—hold much sway on users’ perceptions of a website. Design factors identified as important to aesthetic fidelity by Kim, Lee, and

Choi (2003) dealt with all four of these categories. Bauerly and Liu (2006) and Ngo, Teo, and Byrne (2003) found structure of a webpage to be an important aesthetic factor (with the latter set of authors going so far as to design a set of equations to quantitatively measure the structure of a page). Zviran, Te'eni, and Gross (2006) found that simply changing the colors used in an email can change how a user interprets that email. Finally, Krauss (2004) used all four of these areas to heuristically evaluate the aesthetics and affective messages of a live website.

Just as the graphic design literature provides a guide as to which visual design principles are likely to have an effect on aesthetic fidelity, the HCI literature provides the same kind of guidance for user-centered design tools. Although no study has yet investigated the link between the use of user-centered design tools and aesthetic fidelity, many studies have established a list of the most commonly used tools by design professionals.

Vredenburg, Mao, Smith, and Carey (2002) established just such a list by surveying over 100 practitioners of user-centered design. The study found a number of tools were both commonly used and thought highly effective by the respondents: field studies, user requirements analysis, iterative design, usability evaluation, task analysis, focus groups, formal heuristic evaluation, user interviews, prototyping, surveys, informal expert review, card sorting and participatory design. Gruen, Rauch, Redpath, and Ruettinger (2002) and Head (2003) suggest two other specific tools that can be used to augment these methods: scenarios and personas, respectively. Several other studies have confirmed that this list of 15 user-centered design tools

represents the most used and most respected methods in the field (Mao, Vredenburg, Smith, & Carey, 2005; Venturi & Troost, 2004).

Summary

The literature reviewed above provides insight on how the concept of aesthetic fidelity might function in real-world interfaces. The original studies in aesthetic fidelity (Kim et al., 2003; Park et al., 2004) give persuasive evidence that the concept is a valid measurement tool that can be used to gauge a user's response to affective messages communicated by an interface. In the real world, the affective messages that can be communicated through an interface are nearly infinite, but the "affective message space" can be successfully narrowed down through careful consultation with experts in interface design and a process of iterative refinements. These insights from previous research suggest that this study's first two research questions are valid and that the aesthetic fidelity of real-world interfaces should be measurable and meaningful. What is not known, however, is how such measurements should be interpreted when they are used to gauge how effectively complex, specific, "real-world" affective messages are communicated—this is precisely the area this study's first two research questions seek to probe.

Similarly, the literature reviewed here suggests methods that might be helpful in aesthetics fidelity research. Clearly there are two necessary phases to a study of aesthetic fidelity: gathering a list of affective messages to be communicated and testing how strongly users receive these messages. Previous studies in interface aesthetics suggest various methods for conducting each phase, such as using in-depth interviews with designers to understand their thinking and asking users to rate

interfaces based only on their visual appearance. This study's third research question asks whether these methods will hold up in research that focuses on interfaces created for the real world rather than the information science laboratory.

Finally, usability research notes that there are a number of common principles and tools used by designers to create usable interfaces. These studies, however, have largely focused on tools designers use to refine the functionality rather than the aesthetic appeal of their designs. This study's final research question asks whether any of these same tools can also be used to help designers increase the aesthetic fidelity of an interface.

Methods

This two-phase study explored the applicability of aesthetic fidelity to live websites. The first, qualitative phase focused on interviews with website designers while the second, quantitative phase comprised a small user study. Two web designers were selected from a pool of candidates and interviewed about the processes they use when creating websites and what affective messages they tried to communicate in selected sites. Results of these interviews were used to create a user study where subjects viewed the sites discussed with the designers and rated the sites on how strongly each communicated the emotions and ideas elicited from the designers. Data from both phases was then analyzed to determine how well the concept of aesthetic fidelity can be used to measure the "affective effectiveness" of real websites and what methods designers might be able to use to increase aesthetic fidelity. The methods for both sections are described in more detail in the following sections.

Phase 1: designer interviews

The first phase of the study began by identifying candidate websites. Since this study used live sites, careful consideration was given to what kinds of sites should be included. Unlike in Kim et al. (2003) and Park et al. (2004), the sites used in this study have not been specially crafted to communicate specific affective messages and emotional dimensions to users—such terms may not have even entered the discussion when the sites used in this study were designed. Even so, for this research to work, the sites used in it must be able to provoke some affective response in users. For that reason, candidate websites were limited to those promoting a singer, band or other musical group. This class of sites was chosen because of the high degree of visual crafting that goes into many such sites to make the site mirror the affective messages of the singer or group. In order for a site to make the candidate list, the site's designers must also have been identifiable and contactable.

An initial list of 30 candidate sites was constructed by performing web searches using queries like “musician website” and “band website design.” Results from these searches that met the above criteria were recorded and designers of the identified sites were contacted by email to solicit their participation in the study. Although over one-third of the designers contacted indicated their willingness to be interviewed, the small, exploratory nature of this study prevented them all from participating. In the end, two designers (and their two corresponding websites) were chosen for inclusion in the study based both on the designers' ability to work within the time constraints of this study and the sites' high level of visual crafting. Designers

were promised anonymity for the study and are thus referred to here as “Designer A” and “Designer B.”

With the list of designers finalized, designer interviews began. These interviews had three goals: to discuss the web design process used by each of the designers, to learn more about the kinds of tools used by the designers during the creation of the studied sites and to elicit a list of affective messages or ideas the designers tried to communicate through each site. Interviews were conducted over the telephone (in the case of Designer A) and via email (for Designer B) depending on the designers’ preferences. Discussions with designers were based around an interview guide that prompted the designer to discuss their web design background, the processes and tools they use when designing sites (with special attention paid to the principles and tools discussed in the literature review above) and the kinds of messages they were hoping to convey with the sites being studied.

After each designer interview, the data from the initial discussion was used to compile two lists: a list of visual and user-centered design methods used during site creation and a list of affective messages designers hoped their sites would convey to users. Once the analysis was complete, designers were given a chance to review the lists for validity and suggest changes (following the iterative process for identifying affective dimensions discussed above). Although designers had the opportunity to refine the lists as many times as necessary, both were satisfied with the results after the first follow-up.

With the final lists in hand, the second phase of this research, a small user study, could begin.

Phase 2: User study

Using the list of affective messages compiled for each site from designer interviews, a simple user study was designed to measure the aesthetic fidelity of each site. The study was based on the methodology for assessing aesthetic fidelity in Kim et al. (2003) and Park et al. (2004). Namely, study subjects looked at a screenshot of the homepage of each site on a computer monitor directly in front of them and rated the site using a questionnaire comprised of seven-point rating scales (presented in a pre-randomized order) asking about the strength with which each affective message was conveyed. Each site was rated on *all* affective messages identified by both designers in the first phase of the research. The questionnaire (Appendix A) also gave subjects the chance to fill in up to three emotions or ideas they felt each site communicated that were not identified by the designers.

Users were recruited for the study from the student population of the University of North Carolina at Chapel Hill. Subjects who took part in the study were compensated \$5 for the 15-30 minutes they spent on the study. All user studies were done in groups, with 8-10 subjects present for each of the three study sessions held.

The following procedures were used:

1. The experimenter met the subjects in the study room and gave them a brief description of what they would be asked to do along with a packet of study materials.
2. Subjects read and signed an informed consent form.
3. Subjects filled out a simple demographic questionnaire that included questions on gender, age, major/profession and experience with using the web.
4. Subjects read the written direction sheet (Appendix B).

5. Subjects rated each site using the provided screenshots and the rating questionnaire (users with odd randomly assigned subject numbers rated Site A first and then Site B, while users with even subject numbers rated the sites in the opposite order).
6. When the subjects completed all of their ratings, they were thanked and given \$5 for their participation.

When the final user study session was complete, data collected during the sessions was entered into a statistical analysis package and data analysis begun.

Data analysis

Data gathered during both phases of the study was used to explore the three research questions. First, interview data was analyzed for a number of qualitative factors, including:

1. Background information about the designers (looking especially for patterns that might be useful to know about in larger studies of aesthetic fidelity)
2. The processes used by designers when building a website (and especially when trying to convey affective messages through a site)
3. The designer's thoughts on how the concept of aesthetic fidelity could be applied to real-world websites (even though the term "aesthetic fidelity" never came up during the interviews)
4. The kinds of affective messages designers hoped to convey in their sites

Data from the user study was analyzed quantitatively using the methods in Kim et al. (2003) and Park et al. (2004) to compute aesthetic fidelity: simple means on the rating scales were determined for all of the identified affective messages for both sites in a number of combinations (these combinations are discussed further in the "Results" section). User-provided affective messages were also analyzed for patterns that could inform future aesthetic fidelity studies.

Results of phase I: designer interviews

Interview with Designer A

Designer A has been creating websites for approximately two years, working one year as a freelance designer and one as the owner of a small design firm. She has no formal training in user interface design (she first learned how to make websites from a friend), but she has had some post-secondary education in art and design. She said she believes that good design is a combination of art and marketing. Currently, Designer A focuses on helping bands and other small organizations create a web presence on a limited budget, and she estimates that she has created 50 websites in her career.

Designer A got involved with Site A through her friendship with members of the band the site promotes. She said the band had a previous website but hired her to revamp their web presence with a “nice but simple site.” Designer A is no longer heavily involved with updating Site A, although she noted that she has remained involved with the long-term maintenance of other band websites she has created.

There were several goals Designer A had in mind when creating Site A. First and foremost, she said, it was important to make a site where “people can find what they’re looking for.” At the same time, she also wanted to make a site that conveyed the band’s sound and personality with a cohesive theme. The band, she said, is hard working and interested in advancing their musical career while at the same time being a fun group that enjoys making music. Designer A summed up this contrast as a “juxtaposition of moods,” a feeling she said she tried to echo with the design of Site A. She said she tried to convey this juxtaposition through the use of contrasts in her

design, combining elements meant to evoke feelings of “evil” vs. “comic,” “lightheartedness” vs. “sci-fi.”

With a background in art, Designer A used several visual design principles to send these affective messages. She combined “evil” colors like red and black with lighthearted imagery. She contrasted a balanced structure with jagged, “informal” type in the site’s header. She pointed out that she would have continued to play with contrasts using other typographical elements on the page but was limited by the constraints of typography on the web (i.e. being constrained largely to common typefaces that are guaranteed to render in multiple operating systems and web browsers).

Designer A also used a number of user-centered design tools to help refine the look and feel of Site A. She said she made heavy use of informal user testing, recruiting her friends and family to look over prototypes of the site (a process she said she repeats on every site she designs). She also put the site through an informal heuristic evaluation, making sure the final version passed certain rules of thumb (working links, good readability) before making the site live. Finally, Designer A went through a process familiar to many user-centered designers: finding a balance between a perfect site and a site that is finished on time and on budget.

Four affective messages were chosen from the interview with Designer A to test in the user study of Site A:

- “a contrast between evil and comic”
- “a juxtaposition of moods”
- sci-fi
- lightheartedness

Interview with Designer B

Like Designer A, Designer B has no formal education in web design or computer or information science and describes himself as “an artist at heart.” He got his start in web design eight years ago, building a site for a band he was in at the time. He currently works as an independent designer and estimates he has designed roughly 30 sites in his career.

Also like Designer A, Designer B became involved with Site B through a friendship with the band the site promotes. He designed the site three years ago as a natural extension of his previous collaborations with the band, including creating album art and other band memorabilia. According to Designer B, because the band members are not overly concerned with maintaining an up-to-date web presence, he has had minimal involvement with the site since creating it.

Designer B said he had several goals in mind when creating the site, with his primary aim being to design a site that matched the style of his previous collaborations with the band. Specifically, he said:

“The goal [the band] and I always try to convey when we collaborate is an organic and ‘analog’ look. I try as much as possible to use non-digital art at the base of all the work I do for them: using clip art and entirely digital illustration just doesn’t seem to fit the band at all. Basing a website design around drawings is certainly not common, so I decided to put something together to fit as much real art into every page.”

Although Designer B had a particular look he was trying to convey with the site, he was quick to point out that he does not necessarily “design sites to communicate a specific emotion or idea, but I design sites that fit whom they’re representing in every way. So it’s not so much that I’m trying to communicate an organic feel, but that I’m trying to create a unique online representation of an entity.” In the case of

the band being represented by Site B, Designer B said he tried to capture the band's "marriage of structured music to improvisation," a message he hoped to communicate through the illustration at the top of Site B, described as "a marriage of fairly precise fine art and chaotic digital illustration." Like Designer A, Designer B stressed that he wanted to keep the site simple and usable as well.

Designer B said he used several of the principles and tools asked about in the interview to help him design the site. As an "artist at heart," Designer B put much thought into how to use the four visual design principles (color, structure, imagery, typography) to communicate affective messages through the site. He chose both the color scheme and imagery to complement the cover art for the band's latest album, noting that the dominant image on the homepage plays on the idea of the "combination of nature and the technology of music" that echoes both his and the band's aesthetic preferences. His primary concern with the structure of the page was to create a "unique, simple and user-friendly" interface that stressed usability. Typography, he said, was less of a concern because of the typographical limits inherent in website design. Designer B also made use of three common user centered design techniques: prototyping (testing several iterations of the site before settling on a final version), task analysis (by informally consulting friends on how they would use the site) and an informal user study (again using friends as test subjects and joking that because the band he was designing for "are so technology inept" that they made for excellent testers to discover flaws an expert user might overlook).

In the end, four affective messages whose aesthetic fidelity could be tested were gleaned from the interview with Designer B (with the caveat that these

messages were separate parts of a whole aesthetic designed to echo the feeling of the band the site was designed to promote):

- the idea of “organic”
- the idea of “analog”
- the idea of a marriage between structure and improvisation
- usability

Implications for user study

The results of the designer interviews had a direct effect on the design of the next phase, the user study. The eight affective messages identified between the two designers were used to construct the user study questionnaire (Appendix A). The questionnaire was constructed by first associating each message with a seven-point rating scale. The order of the messages on the questionnaire was then randomized to prevent the messages given by Designer A or Designer B from being clumped together and possibly influencing users’ opinions of the site. Finally, discussions with designers made it clear that there were numerous other affective messages each site could communicate to users that were not among the eight on the questionnaire, so space was left to give user study subjects a chance to provide up to three of their own emotions or ideas they felt each site communicated.

Results of phase 2: user study

A total of 24 subjects took part in the user study phase of the research. Subjects were between the ages of 18 and 31 years old (mean = 21.1 years, standard deviation = 3.5) and were all either undergraduate (N = 20) or graduate (N = 4) students at the University of North Carolina at Chapel Hill. Subjects had a variety of major fields of study in the humanities and social and natural sciences. Tables 1 and 2

show that most subjects also made significant use of the web in their daily lives and had at least some experience visiting sites like those they worked with in the study (those that promote bands or musicians).

Table 1: Responses to the question, “on average, how much time do you spend on the web?”

Average time spent using web	Number of subjects
More than 4 hours per day	8 (33%)
3–4 hours per day	9 (38%)
1-2 hours per day	6 (25%)
Less than 1 hour per day	1 (4%)

Table 2: Responses to the question, “how often do you visit websites that promote a band or musician?”

Frequency of visits to band/musician websites	Number of subjects
Daily	2 (8%)
Weekly	8 (33%)
Monthly	5 (21%)
Less than monthly	9 (38%)

As described in the methods section, subjects in the user study were shown screenshots of Site A and Site B and asked to rate each on a seven-point rating scale. Table 3 shows the mean rating score for each of the eight affective messages on the user study questionnaire (standard deviations appear in parentheses after each mean). The letter after each message indicates the website it is associated with; that is, messages marked as “A” were suggested by Designer A to describe Site A and those marked as “B” were elicited from Designer B to describe Site B.

Table 3: Mean ratings (and standard deviations) for sites A and B for the eight affective messages. The letter in parentheses after each message indicates with which site the message is associated.

Affective message	Site A	Site B
“a contrast between evil and comic” (A)	4.71 (2.02)	1.92 (1.81)
“a juxtaposition of moods” (A)	2.83 (1.75)	4.17 (1.74)
sci-fi (A)	4.50 (1.46)	2.46 (1.50)
lightheartedness (A)	1.79 (1.17)	4.63 (1.01)
the idea of “organic” (B)	2.17 (1.05)	5.42 (1.01)
the idea of “analog” (B)	3.38 (1.28)	3.75 (1.33)
the idea of a marriage between structure and improvisation (B)	3.17 (1.62)	5.04 (1.42)
usability (B)	5.13 (1.22)	6.17 (0.90)

Table 4 presents another way to look at the questionnaire data. In it, each of the eight affective messages are given along with the site they are associated with and from which site users felt each message most strongly. Results from a two-tailed, paired t-test are also given for each message to indicate whether the difference in mean ratings between sites is statistically significant. High aesthetic fidelity, then, would occur when users received the affective message significantly more strongly from the site with which it is associated. The five messages that appear in bold in Table 4 meet the condition for high aesthetic fidelity. For one message (“the idea of ‘analog’”), there was no statistically significant difference in the strength with which it was conveyed across the two sites. In two cases, the site intending to convey the message was less successful in conveying it than the other site; these messages are marked with an asterisk in Table 4.

Table 4: Assessment of which messages were had high aesthetic fidelity. Messages in bold have high aesthetic fidelity; messages marked with an asterisk (*) were conveyed more strongly by the site with which it was *not* associated.

Affective message	Associated with:	Most highly rated in:	Significance
“a contrast between evil and comic”	Site A	Site A	0.000
“a juxtaposition of moods”*	Site A	Site B*	0.012
sci-fi	Site A	Site A	0.000
lightheartedness*	Site A	Site B*	0.000
the idea of “organic”	Site B	Site B	0.000
the idea of “analog”	Site B	No difference	0.322
the idea of a marriage between structure and improvisation	Site B	Site B	0.000
usability	Site B	Site B	0.004

Subjects in the user study were also able to provide up to three affective messages per site that each screenshot evoked in them; they then rated these messages using the same scale as above. Users provided a total of 57 of their own messages between the two sites: 29 for Site A and 28 for Site B (Appendix C). In general, users felt the messages they provided were portrayed very strongly in the sites. Subjects rated 17 of the 29 messages (58.6%) they identified for Site A at six or seven on the rating scale and 19 of 28 (67.9%) for Site B. Many of the user-provided messages for Site A centered on negative emotions. The three most highly ranked messages were “intimidation,” “bloody” and “excessive/extreme,” and three separate users indicated they felt Site A transmitted a feeling of “anger.” User-provided messages for Site B were generally associated with more positive emotions or were

focused on musical topics: top-rated messages for Site B included “appealing,” “energetic,” “creative” and “jam-bandishness.”

Discussion

The results from both phases of this study can be used to shed light on the study’s four research questions and increase understanding of how aesthetic fidelity functions in real-world interfaces.

Are designers of real-world websites successful in creating sites with aesthetic fidelity?

Based on Tables 3 and 4, it would appear that Designer B was somewhat more successful than Designer A in achieving aesthetic fidelity. Table 4 shows that Site B achieved high aesthetic fidelity for three of the four messages identified by Designer B: “the idea of organic,” “the idea of a marriage between structure and improvisation” and “usability.” The fourth message identified by Designer B, “the idea of analog,” may only have been kept from high aesthetic fidelity status by the frequent confusion expressed by subjects in the user study on what exactly “analog” meant. Site A, on the other hand, had only two messages with high aesthetic fidelity: “a contrast between evil and comic” and “sci-fi.” Further, two messages identified by Designer A as goals for Site A were actually rated significantly higher for Site B.

Analysis of user-provided affective messages confirms that Site B was apparently more effective than Site A at communicating its intended affective messages to users. As seen in Appendix C, the majority of user-provided messages for Site A focused on the “evil” half of Designer A’s goal of creating a site that evoked “a contrast between evil and comic.” Users frequently identified words like “bloody,”

“evil,” “intimidating” and “anger” for the site and often indicated they felt these emotions and ideas very strongly. Strongly felt user-identified emotions for Site B, on the other hand, more closely matched Designer B’s intentions. The site evoked feelings like “rock music,” “appealing,” “creative” and “intrigue” in users.

Taken at face value, these data indicate that Site B had better aesthetic fidelity than Site A and is thus somehow more effective overall. This assertion is, however, overly simplistic. Aesthetic fidelity, born as a simple and compact concept for a series of highly regulated experiments, seems to require a much broader interpretation when applied to real-world interfaces. In the original experiments (Kim et al., 2003; Park et al., 2004), aesthetic fidelity measured the transmission of only a core set of emotional dimensions, but in this study it was pressed into service to measure how strongly a potentially infinite set of messages are communicated to users. Saying that Site B is somehow more effective at emotional communication than Site A is thus inaccurate—such an assertion is like comparing apples to oranges. Instead, when applied to real-world interfaces, aesthetic fidelity seems best suited as an internal guide that designers can use to refine their interfaces rather than as an externally valid measure of how “affectively effective” a site is.

With this newly expanded definition of aesthetic fidelity in mind, initial answers to this study’s other three research questions can be posited.

How well can the idea of aesthetic fidelity be applied to real-world, non-experimental interfaces?

At a very basic level, the idea of aesthetic fidelity can be transferred from the laboratory to the real world. Designers were able to successfully generate a number

of affective messages that they hoped their sites could communicate, these affective messages could be rated by typical users and aesthetic fidelity scores could be computed. However, as discussed above, it is necessary to change the scope of aesthetic fidelity when working with live interfaces. The concept worked so cleanly in the original aesthetic fidelity experiments because it only needed to measure how strongly users felt on 13 general emotional dimensions. For real-world interfaces, however, designers are less interested in measuring how emotionally “bright” an interface is than measuring how well it conveys specific ideas like “a marriage between structure and improvisation.”

“Real-world” aesthetic fidelity, then, must be expanded in scope to account for the infinite range of emotions it can measure in live interfaces. Further, it works best not as a method to compare the communicative effectiveness of different interfaces but rather as an internal measure of how well a single interface is transmitting specific affective messages to users.

What methods should be used in future studies of the aesthetic fidelity of live interfaces?

Both phases of the research suggest methods that can be used to great effect in future research into the aesthetic fidelity of live interfaces. Results from the designer interviews add further support to the assertion that the affective messages measured in aesthetic fidelity studies should not be reduced to a core set of “emotional dimensions” as in Kim et al. (2003) and Park et al. (2004). Results from the user study phase made clear that future studies could benefit from more

qualitative, in-depth methodologies to more fully understand the reasoning behind users' ratings of the sites being investigated.

As discussed above, the original studies into aesthetic fidelity were so successful because the researchers were able to reduce a vast array of emotions and ideas into a small set of “emotional dimensions.” While this method worked well in tightly controlled experimental studies of aesthetic fidelity, the designers interviewed in this study made clear that such a strategy would be too simplistic to be meaningful when doing research with “real-world” interfaces. The emotional dimensions identified by Designers A and B were nuanced and subtle. Instead of seeking to convey the general emotional dimension of “bright,” Designer A wanted to convey a message of “lightheartedness.” Rather than trying to convey the idea of “structure” to users, Designer B sought to evoke the idea of “a marriage between structure and improvisation.” Reducing these specific affective messages into general emotional dimensions may make future aesthetic fidelity studies easier, but it would also make them less valid: if a study found Site B only conveyed “structure” to users, for instance, one could not be certain that it was sending the exact affective message intended by Designer B. In order to truly understand the aesthetic fidelity of a “real-world” interface, complex, subtle and often messy “real-world” emotional dimensions must be used.

The user study phase of future aesthetic fidelity research could benefit from more fully embracing the “messiness” inherent in evaluating real-world interfaces as well. True to the spirit of the original investigations into aesthetic fidelity, this study asked users to look at screenshots of various websites and translate their qualitative

feelings about them into a quantitative questionnaire. While this method was able to produce interesting results, it also produced results stripped of the nuances that go hand-in-hand with aesthetic judgments. For instance, the user study results show that, on a purely quantitative level, Site A had less aesthetic fidelity than Site B. What these data cannot possibly begin to probe, though, is why this is the case. Is the difference attributable to some difference in the designers of each site? Is it because the background of user study subjects affected how strongly they felt the various emotions each site attempted to convey? During the course of the user study, several subjects asked for clarification on some of the affective messages on the questionnaire: perhaps confusion over the way Site A's messages were phrased led to the difference in scores between the two sites. Without more in-depth interactions with user study subjects, it is impossible to understand why they give interfaces the ratings they do, and thus very difficult to determine what factors affect the aesthetic fidelity of real-world interfaces.

What methods can designers of live interfaces use to increase the aesthetic fidelity of their designs?

Designer interviews revealed several interesting features of the interface design process that could have an effect on the aesthetic fidelity of live websites. Specifically, it was observed that aesthetic fidelity was crafted in two phases, that the designers used direct contact with potential users as a primary tool, and that there is unlikely to be a direct correlation between the use of a certain tool or tools and an increase in aesthetic fidelity.

During their interviews, both designers described taking a two-phased approach when creating their sites. The first phase could be described as “crafting,” where they considered the kinds of affective messages they wanted their sites to send, thought about what elements of visual design would be most appropriate to send those messages and then mocked up a design based on their thoughts. The second phase is best described as “refining,” where the designers informally bounced their initial ideas off of potential users to elicit feedback on how the site could be improved. This process is, of course, very similar to the standard systems development life cycle, although the process described by the designers focuses more explicitly on “feelings” rather than “functionality.”

Both designers used different toolsets in the two phases of their design process. In the first phase, they relied mostly on their knowledge of visual design principles to craft affective appeal, using their artistic backgrounds to create designs that could send the right messages to users. In the second phase they relied more on traditional user-centered design tools, like informal user studies, to tweak their designs. This pattern suggests that designers wanting to increase the aesthetic fidelity of their interfaces might do well to increase their understanding of visual design principles to help them *craft* affective messages and to rely on user-centered design tools to help *refine* the interfaces.

While both designers had substantial artistic background to rely on when they were crafting their sites, both still chose to consult with users (at least informally) to verify that their initial ideas for each site were being expressed as intended. Although data from two designers are insufficient to support broad

generalizations, the designers' reliance on methods like informal user studies suggests that an interface's aesthetic fidelity can benefit through even minimal user testing. As seen in the results from the user study phases of this and other research (Pandir & Knight [2006], for example), the affective messages users receive from a website are highly subjective and can vary greatly from user to user. With such a wide variation, interface designers—even those with extensive design backgrounds and experience—can only guess at how their designs will be perceived by users. Direct contact with either real or potential users is thus the surest way to determine whether affective messages are being received as intended. Thus, it seems likely user-centered design tools that involve direct contact with users (user studies or focus groups, for example) will be more beneficial to increasing the aesthetic fidelity of a site than those that do not rely on contact (heuristic evaluations, for instance, which instead rely on experts trying to assume the role of typical users).

That said, both phases of this study show that there is unlikely to be a direct correlation between the use of certain tools and an increase in aesthetic fidelity. If there is one thing that this research made abundantly clear, it is that the aesthetic fidelity cannot be achieved solely with increased effort by the designer. The interview with Designer A, for instance, revealed that she put extensive thought into the kinds of messages she wanted users to receive from Site A, and she put effort into developing these messages. Yet two of Site A's affective messages did not have high aesthetic fidelity. This lack of aesthetic fidelity is likely not due to a lack of knowledge or effort on the part of Designer A, but rather attributable to the highly subjective nature of the entire concept of aesthetic fidelity. Thus, while future

research into how designers might improve the aesthetic fidelity of their interfaces may be able to make broad suggestions of processes for crafting and refining affective messages, there is unlikely to be one “correct” way of ensuring an interface’s aesthetic fidelity (just as there is no one “correct” way to create a “beautiful” painting).

Conclusion

By expanding the idea of aesthetic fidelity to account for the “messiness” of real-world affective messages and embracing its power as an internally valid tool to refine the messages sent by a single site, aesthetic fidelity could become a useful lens through which to view subsequent studies in interface aesthetics. Based on this exploratory study, it appears future research into aesthetic fidelity could take a number of potential paths:

- A very in-depth study could be constructed where a single designer is followed by researchers throughout the interface design process. Aesthetic fidelity measures of prototypes and mockups could be taken at various points in the design process to observe how decisions in both the “crafting” and “refining” phases affect the aesthetic fidelity of the interface.
- The basic framework of the current study could be augmented with a much more in-depth user study. Rather than just asking users to rate interfaces on a simple rating scale, users could be interviewed to understand what elements of the interface most affect how strongly users feel various affective messages and the reasons users assign the ratings they do.

- A kind of “bottom-up” aesthetic fidelity experiment could be designed where users are first asked to come up with a number of their own affective messages for an interface and these messages are then compared to emotional dimensions identified by designers. In this case, aesthetic fidelity would be measured by the degree to which user-provided affective messages match designer-provided ones.

These research ideas represent only three of many possible studies that could be performed in the future to more fully refine the idea of aesthetic fidelity with respect to “real-world” interfaces. With further definition, the concept of aesthetic fidelity can give both interface designers and researchers another method to explore the difficult but rewarding area of interface aesthetics.

Acknowledgements

This study received funding from the Chatman Award from the University of North Carolina at Chapel Hill’s School of Information and Library Science’s Alumni Association, for which the author is extremely grateful. Special thanks also go to Prof. Barbara Wildemuth, the author’s research advisor, for her invaluable insights into the study.

References

- Bauerly, M., & Liu, Y. (2006). Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computers Studies*, 64(8), 670-682.
- Ben-Bassat, T., Meyer, J., & Tractinsky, N. (2006). Economic and subjective measures of the perceived value of aesthetics and usability. *ACM Trans. Comput.-Hum. Interact.*, 13(2), 210-234.
- Bertelsen, O. W. & Pold, S. (2004). Criticism as an approach to interface aesthetics. *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*, Tampere, Finland. 23-32.
- Cawthon, N. & Moere, A.V. (2007). The effect of aesthetic on the usability of data visualization. *Information Visualization 2007: Proceedings of the 11th IEEE international conference on information visualization*, Zurich, Switzerland. 637-648.
- Faison, B. S. (1996). Graphic designers in transition: from print communications to interactive media design. *interactions*, 3(1), 39-57.
- Frascara, J. (2004). *Communication design :principles, methods, and practice*. New York: Allworth Press.
- Gruen, D., Rauch, T., Redpath, S., & Ruettinger, S. (2002). The use of stories in user experience design. *International Journal of Human-Computer Interaction*, 14(3&4), 503-534.
- Head, A. J. (2003). Personas: setting the stage for building usable information sites. *Online*, 27(4), 14-21.
- Hoffmann, R., & Krauss, K. (2004). A critical evaluation of literature on visual aesthetics for the web. *SAICSIT '04: Proceedings of the 2004 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, Stellenbosch, Western Cape, South Africa. 205-209.
- Jansen, B. J., Zhang, M., & Zhang, Y. (2007). The effect of brand awareness on the evaluation of search engine results. *CHI '07: CHI '07 extended abstracts on human factors in computing systems*, San Jose, CA, USA. 2471-2476.

- Kim, J., Lee, J., & 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(6), 899-940.
- Krauss, K. (2004). Visual aesthetics and its effect on communication intent: a theoretical study and website evaluation. *Proceedings of the Southern African Computer Lecturers Association (SACLA)*, University of KwaZulu-Natal, Durban, South Africa. 451-469.
- Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. *CHI '95: Conference companion on Human factors in computing systems*, Denver, Colorado, United States. 292-293.
- Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computers Studies*, 60(3), 269-298.
- Mao, J., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, 48(3), 105-109.
- Ngo, D. C. L., Teo, L. S., & Byrne, J. G. (2003). Modelling interface aesthetics. *Information Sciences*, 152(1), 25-46.
- Norman, D. A. (2004). *Emotional Design: Why we love (or hate) everyday things*. New York: Basic Books.
- Pandir, M., & Knight, J. (2006). Homepage aesthetics: The search for preference factors and the challenges of subjectivity. *Interacting with Computers*, 18(6), 1351-1370.
- Park, S., Kim, H., & Kim, J. (2004). How to visually create clear personalities with blogs? *SIGGRAPH '04: ACM SIGGRAPH 2004 Posters*, Los Angeles, California. 8.
- Park, S., Choi, D., & Kim, J. (2004). Critical factors for the aesthetic fidelity of web pages: empirical studies with professional web designers and users. *Interacting with Computers*, 16(2), 351-376.
- Sutcliffe, A. G. (2002). Heuristic evaluation of website attractiveness and usability. In C. W. Johnson (Ed.), *Proceedings 8th Workshop on Design, Specification and Verification of Interactive Systems (Glasgow)* (pp. 188-199). Berlin: Springer-Verlag.
- Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, Atlanta, Georgia, United States. 115-122.

- Venturi, G., & Troost, J. (2004). Survey on the UCD integration in the industry. *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*, Tampere, Finland. 449-452.
- Vredenburg, K., Mao, J., Smith, P. W., & Carey, T. (2002). A survey of user-centered design practice. *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, Minneapolis, Minnesota, USA. 471-478.
- Williams, R. (2004). *The non-designers design book :design and typographic principles for the visual novice* (2nd ed.). Berkeley, Calif.: Peachpit Press.
- Zviran, M., Te'eni, D., & Gross, Y. (2006). Does color in email make a difference? *Communications of the ACM*, 49(4), 94-99.

Appendix A: Rating questionnaire

How strongly does this site convey:

	Not at all				Very strongly		
lightheartedness:	1	2	3	4	5	6	7
the idea of “organic”:	1	2	3	4	5	6	7
a juxtaposition of moods:	1	2	3	4	5	6	7
a contrast between evil and comic:	1	2	3	4	5	6	7
a feeling of “sci-fi”:	1	2	3	4	5	6	7
the idea of “analog”:	1	2	3	4	5	6	7
the idea of a marriage between structure and improvisation:	1	2	3	4	5	6	7
usability:	1	2	3	4	5	6	7

Does this site convey any other feelings or ideas to you? If so, please list them here and rate how strongly you feel each is conveyed.

	1	2	3	4	5	6	7
	1	2	3	4	5	6	7
	1	2	3	4	5	6	7

Appendix B: User study directions

You will now be shown screenshots of two websites. For each site, you will fill out a questionnaire with a list of emotions or ideas and rating scales. Please look at each screenshot for as long as you need and rate (by circling one rating for every emotion or idea) how strongly the site conveys each emotion or idea to you. You will also have a chance to provide your own opinions of what emotions and ideas each site communicates. There are no right or wrong answers—the rating you give should reflect your personal feelings.

When you are done rating the first site, please click the “Next” link to see the second site and rate that as well.

Please ask the researcher if you have any questions and let the researcher know when you are finished.

Appendix C: User-provided affective messages

User-provided affective message for Site A	User rating
intimidation	7
bloody	7
excessive/extreme	7
violent	6
originality	6
dark	6
evil	6
ambushed	6
independence/"outside the box"	6
intimidating	6
darkness or evil	6
artistry	6
fear	6
anger	6
threatening	6
creative	6
tackiness	6
a kind of knowing, not quite smugness, but ironic superiority	5
danger	5
pretension	5
anger	5
uninterested	5
hardcore	5
hate	5
lack of originality	5
loud	5
anger	3
interesting/appeal	3
boxy	(no rating given)

User-provided affective message for Site B	User rating
seriousness, as in they take themselves seriously	7
rock music	7
juxtaposition of hardcore and soft	7
jam-bandishness	7
appealing	7
energetic	7
creative	7
pleasant	6
originality	6
emotional openness	6
balance b/t movement & focus	6
mellowness	6
intrigue	6
loudness	6
hippiness	6
hippie	6
interesting/appeal	6
artistry	6
contemporary	6
happy	5
curious	5
simplicity	5
creativity	5
authenticity	5
artistry	5
freedom	4
mixture	4
young	4