This paper has been accepted for publication in *Ergonomics*

Webpage aesthetics, performance, and usability:

Design variables and their effects

KRISTI E. SCHMIDT

Department of Industrial and Operations Engineering, University of Michigan

1205 Beal Avenue, Ann Arbor, MI, 48109-2117, USA

Telephone: (734) 904-1855, Fax: (734) 764-3451

krischmi@umich.edu

YILI LIU

Department of Industrial and Operations Engineering, University of Michigan

1205 Beal Avenue, Ann Arbor, MI, 48109-2117, USA

Telephone: (734) 763-0464, Fax: (734) 764-3451

yililiu@umich.edu

SRIVATSAN SRIDHARAN

Marsh Inc.

5908 Scanlan Avenue, Saint Louis, MO, 63139, USA

Telephone: (734) 717-8327, Fax: (314) 342-2600

vatsa@umich.edu

Keywords: webpage usability; webpage performance; webpage aesthetics; human performance

Abstract

The primary objectives of this research are to identify the underlying clusters of design variables affecting the perceived usability of a webpage; and to examine the effects of webpage design variables on webpage performance. Fifty-seven design variables and ten underlying clusters that conceptualize the structure of user webpage judgement are identified through content analysis on literature and structured interviews, balanced incomplete block user survey administration, and cluster analysis. Five clusters are selected to conduct three experiments that quantify the change in user aesthetic preference, perceived ease of interaction, and interaction speed as a function of loading speed, image colour, image size, font size, link style, and column width. Results show that user performance alone is not a good indicator of aesthetic judgement and overall effectiveness of a webpage.

Relevance of the findings for ergonomics practice

The value of integrating global construct analysis processes and local controlled experimentation processes in ergonomic interface research is illustrated. Fifty-seven webpage design variables are defined, ranked, and clustered according to perceived importance and overall preference. Experimental results illustrate that both technical performance and aesthetic factors are important webpage design considerations.

1. Introduction

1.1 *Background*

The Internet has grown in user population and breadth to a great extent since its creation. There were over 530 million global home Internet users who each averaged 34 surfing sessions during the month of January 2008, spending a mean of over 58 minutes per surfing session, for a total of over 33 hours active Internet surfing for the month (Nielsen//NetRatings 2008). The breadth, volume, and accessibility of the Internet have made it popular for individuals and organizations to create and maintain webpages that go beyond communication to collaboration, ecommerce, and entertainment. The surge in Internet presence was not, however, paired with widespread design sophistication or consideration for usability. The increased complexity of the Internet and the sheer volume of Internet users make the Internet a very complex and often competitive environment. If users are unable to find what they need from a given webpage due to the lack of information or the complexity of navigation, they will become frustrated and move on to another site. On average, users spend 46 seconds viewing each page (Nielsen//NetRatings 2008). This relatively short amount of time demands rapid communication of critical information and may impose high information processing load on the users.

Variables affecting user judgements of a webpage have often been intuitively defined and communicated by designers through instructional manuals. Oliver (2002) defines four principles of webpage interface design and development: 1) usability—how intuitively or easily the media item is navigated and processed; 2) visualization—creation of visually interesting and aesthetically pleasing media items while avoiding potentially distracting or unnecessary features;

3) functionality—features of the media item and how useful they are for supporting a given task; and 4) accessibility—tools that help users access the site in alternative formats and provide increased functionality. Burstein (2002) emphasizes accessibility and groups webpage design variables into fifteen design elements: links, colour issues, images, image maps, animated images, spacing, tables, frames, style sheets, cookies, JavaScript, Java, plug-ins, screen size, and file distribution. These online manuals are often based upon designer intuition and qualitative evaluation of existing webpages.

Others have developed methodologies that evaluate a webpage quantitatively as well as providing insight by critically examining several design aspects. Turner (2002) invented a webpage usability methodology called Heuristic Evaluation by Proxy (HEP) as a guide for site design and creation. The HEP is a set of questions aimed toward individual webpages that address seven categories of usability heuristics: navigation, page design, content, accessibility, media use, interactivity, and consistency. A HEP grading system establishes a benchmark that the usability of a webpage can be measured against.

Cox and Dale (2002) developed a model for webpages that is based on six key quality factors (KQFs) that contribute to user satisfaction: 1) clarity of purpose; 2) design; 3) accessibility and speed; 4) content; 5) customer service; and 6) customer relationships. Design was further broken down into five issues: 1) links; 2) consistency, menus and site maps; 3) pages, text and clicks; 4) communication and feedback; and 5) search and fill-in forms. The KQFs were evaluated by assessing metrics to a sample set of webpages, then comparing the scores of the webpages to the customer base, error rates, and financial performance. Financial performance was hypothesized as a reflection of the quality of a webpage, yet this hypothesis may be difficult to generalize beyond ecommerce webpages.

Two recent studies paid special attention to the emotional and aesthetic aspects of webpage design. Kim et al. (2003) used surveys and controlled design sessions to examine the quantitative relations between webpage design factors and human emotional dimensions. Although the focus of their study was on the emotional aspect and they did not examine the relationship between emotional response and usability, they use the term 'emotional usability' to imply that emotion is part of usability.

Lavie and Travinsky (2004) conducted a series of survey-based studies as their exploratory approach to the development of a measurement instrument of perceived webpage aesthetics. Using factor analysis of the survey results they found that user's aesthetic perceptions consist of two dimensions, which they call 'classical aesthetics' and 'expressive aesthetics.' Further they found a positive correlation between aesthetic perception and usability, confirming similar earlier findings (e.g., Kurosu and Kashimura 1995; Tractinsky 1997).

In addition to the survey-based research reviewed above, some controlled experimental work has recently emerged that investigates tradeoffs among design variables regarding webpage performance, readability, human performance, and aesthetic preference. For example, Dennerlein and Johnson (2006) explored the consequences of interface design and task on the physical aspects of the user interaction. Bauerly and Liu (2006, 2008) investigated the role of symmetry, grouping, and interface composition on webpage aesthetic judgements. The relationship between position of navigational menu frames and selection of hypertext link colour was investigated by Pearson and van Schaik (2003) using an interactive search task to measure performance. Bernard et al. (2003) investigated the readability, typeface legibility, sharpness, ease of reading, and general preference of several font size, font type, and format combinations. It should be noted

that some of the survey-based research used the term ‘experiments’ to describe their work, but the described studies are design sessions or interface survey sessions, not controlled experiments.

1.2 *Current study*

The current studies integrate the survey based and experimental approaches employed in previous studies. They explicitly investigate the relationship between usability, aesthetic judgements, and performance. It should be noted here that there does not exist a consensus on the definition of usability (Hassenzahl & Tractinsky 2006, Thuring & Mahlke 2007). One of the purposes of the present study is to investigate how users define usability; particularly whether it includes only performance-related factors or if it includes aesthetic or emotional factors as well.

The definition of aesthetics or aesthetic judgements is itself a hot topic of debate and an area of theoretical research. Since the focus of the current paper is on the specific topic of webpage usability, it is not the purpose of the current study to debate the definition of aesthetics or to review the vast body of related literature and theories. One of the co-authors of the current paper has published (Liu 2003a) a comprehensive review of the major schools of aesthetic thoughts and theories, such as the philosophical theories, psychophysical theories, cognitive theories, social and historical theories, and sexual and natural selection theories of aesthetics. Liu (2003b) further discussed the benefits of considering the aesthetic and the ethical issues simultaneously with traditional human factors issues in design. Since the focus of the current paper is on the specific topic of webpage usability, we will not repeat those discussions here. It suffices to say that in the current study, we adopt an operational definition of aesthetics: What is aesthetic is what a research participant feels and reports pleasing or appealing in appearance.

This operational approach to defining aesthetics is used by the large majority, if not all, of the application-oriented researchers, either explicitly or implicitly.

The specific research questions addressed in this study are: 1) what are the design variables and their relationships to the user's judgement of webpage usability; 2) what is the role of aesthetic and performance variables in usability; and 3) how do some of the design variables affect aesthetic judgement or performance? A dual-process research and evaluation methodology (Liu 2003a) was used to investigate these issues. This dual-process methodology utilizes two parallel but closely related types of research methods that are aimed at achieving a comprehensive, rigorous, and quantitative understanding of user response, in this case with respect to webpage design. The two types of research methods are multidimensional construct analysis and controlled experimentation.

Multidimensional construct analysis is a global top-down methodology that quantitatively answers questions involving the conceptual structure of hypothetical constructs such as usability or aesthetic judgement, the identification of the major psychological and physical dimensions involved, and the establishment of the relative importance and relationship of these dimensions.

Controlled experimentation is a local bottom-up process that uses well-controlled experiments to establish a quantitative view of how user psychological responses change as a function of specific design variables identified in the multidimensional construct analysis. Specifically, user ability to perceive and judge values, changes and variations in design parameters, and corresponding preferences of the levels of values of design variables are of interest. In this study, three experiments were conducted to quantitatively investigate the effects of several design variables identified in the multidimensional construct analysis on user overall preference, aesthetic preference, perceived ease of interaction, and performance.

2. Multidimensional construct analysis

The first task is to rigorously identify design variables that may affect user judgement of webpages and to identify the relationships among these design variables. Multidimensional construct analysis first uses content analysis on relevant text and structured interviews to identify variables of importance with respect to webpage design. This process is similar to Lavie and Tractinsky's item generation (2004). A survey approach is then utilized to rank those identified variables according to their perceived contribution to webpage usability. The survey data are then analyzed using multivariate statistical data reduction methods to cluster the variables into related groups.

2.1. *Method*

2.1.1. *Participants*: A total of 20 undergraduate students (16 males and four females) with the age range 18-22 years participated in both the Structured interview and the Survey. All participants had at least five years of prior Internet experience and accessed the Internet on average at least two hours a day. The participants were compensated $10 for approximately one hour of their time.

2.1.2. *Structured interviews*: Each participant was specifically instructed to list variables that they thought were important in affecting usability while browsing webpages at their leisure. The participants were encouraged to list as many items as they could think of, as to not limit their thought and creativity.

2.1.3. *Content analysis*: The data collected from the structured interview was combined with text from publications that addressed webpage usability and/or webpage design guidelines (Burstein 2002, Nielsen 2004, De Graaff 2004, Gibbs & Szentivanyi 1997, Hom 1998, Ericsson 1999, Perlman 2004, Marion 2003, Instone 2003). A content analysis was performed on these texts to obtain a list of variables that describe or may affect webpage usability or webpage design guidelines by extracting all meaningful words. Meaningful words are those words that are not transition words and proposition words (i.e. not *a*, *and*, *then*, *with*, or *to*). Similar words or synonyms were consolidated at the end of this process.

2.1.4. *Survey:* Fifty-seven variables were extracted from the content analysis to include in a survey to rank the variables. Fifty-seven variables were used because these words appeared most frequently in the content analysis and seemed to be most representative of the whole range of words used. They also keep the survey at a reasonable length without causing potential fatigue or confusion on the part of the participants. The survey used a Balanced Incomplete Block (BIB) design (Dunn-Rankin 1983), in which a large set of ranking items are broken up into smaller groups. Using fifty-seven variables allowed us to use a reasonable-length BIB design, which groups the full list of fifty-seven items (impossible to rank at once) to fifty-seven groups of eight variables each. This design reduces the cognitive load on the participants by asking them to rank a group of 8 variables at a time, and the fifty-seven groups can be completed in about half an hour. The fifty-seven groups compared each variable to the rest of the variables twice.

Of the 20 undergraduate students who participated in the structured interview, each had at least five years of Internet experience, and each logged onto the Internet at least once daily.

Participants were asked to rank each of 8 variables within a group according to its perceived importance in webpage design and to its contribution to webpage preference.

2.2. *Results and discussion*

2.2.1. *Rank*: Statistical analysis of the BIB survey profile data yields an overall rank for the fifty-seven variables for each participant as well as an overall rank considering all participants' responses. The most important variables (most important first) are: information layout, server response time, time to load, download time, and speed. The five variables that least affect webpage preference (least important first) are: advertisements, songs, sudden pop-up windows, movies, and coordinated audio and video.

2.2.2. *Cluster analysis*: Cluster analysis identifies clusters, or groups, of similar variables according to the underlying user perception of the variables. Hierarchical cluster analysis performed using the BIB survey results produces a dendrogram that illustrates the underlying relationships among the variables by grouping similarly quantified entities at various stages of relationship formation. Ten clusters were identified based upon these underlying relationships: page progression/targeting strategy, basic visual structure, navigation, clarification/simplification, relevance/speed, trust/flexibility, marketing, appeal/diversion, multimedia, and accessibility. Figure 1 illustrates the dendrogram containing the fifty-seven variables that were identified, ranked, and clustered.

[insert figure 1 about here]

Figure 1. Ten Clusters Consisting of 57 Ranked Variables and the Corresponding Dendrogram

To the knowledge of the authors, none of the survey participants had disabilities, which may account for the relatively low contribution (low rank) of the variable 'Accessible for Users with Disabilities' to webpage design and overall preference. The disconnection between disabilities and participants most likely caused the 'Accessibility' variable to not possess a strong underlying relationship with any other variables, therefore forming its own cluster. This cluster will not be considered in further description of high-level cluster relationships.

Figure 2 illustrates the dendrogram with the nine named clusters at the left side (excluding Accessibility). Moving right across the figure, the relationships among clusters are obtained. When clusters join a neighbouring cluster, they form a larger cluster, until finally all variables and clusters become one. At each intersection, these clusters are named to illustrate not just the clustering of variables, but the meaningfulness of the relationships among all of the variable clusters.

[insert figure 2 about here]

Figure 2. High-Level Definition of Dendrogram Clusters

The results of the content analysis and cluster analysis show that the perceived usability and overall preference of a webpage depend upon many design factors, some of which are performance related, some appearance related, while others feature related. The results also show

that performance-related design variables are regarded as the most important ones in the perceived usability and overall preference of a webpage. To complement this survey study and to obtain a more detailed understanding of how users make overall preference judgements in the face of performance-appearance tradeoffs, and how users' performance and preference change as a function of specific design variables, three experiments were conducted.

3. Experiment One

The survey results showed that technical performance and appearance both contribute to perceived webpage usability and technical performance was judged as being more important than appearance factors. Experiment 1 further examines the potential tradeoff between technical performance and appearance factors in user judgement through a two factor experiment. One of the experimental factors, time-to-load, was selected from the technical performance cluster, while the other, use of colour images, was from the visual appeal cluster.

3.1. *Method*

3.1.1. *Participants*: A total of seven undergraduate students (four males and three females) with the age range 18-22 years participated in Experiment One. Each had normal (20/20) or corrected-to-normal vision and normal colour vision. All participants had at least four years of prior Internet experience and accessed the Internet on average at least two hours a day. The participants were compensated $10 for approximately one hour of their time.

3.1.2. *Apparatus*: Twenty-four webpages were created based upon the layout of news articles from the usnews.com site (2004). All articles were approximately the same length. The webpages were uploaded to the Internet and linked from a dummy site to each individual page. The webpages were presented on a 17 inch visual display terminal using a 400 MHz desktop computer. The Internet connection varied from a slow modem with the highest variability in loading speed that took over 40 seconds to load each page, a fast modem with moderate variability in loading speed that took between 10 and 40 seconds to load each page, and a local area network (10 Mbps unswitched) connection with the lowest variability in loading speed that took less than 10 seconds to load each page. Participants used a computer mouse as the input device.

3.1.3. *Conditions*: The experiment was a repeated measures design with two within-subject factors. The within-subject factors were the level and relative variability of loading speed and the level of aesthetic content. Three levels of loading speed were accomplished by varying the speed of the internet connection. Slow loads took over forty seconds, medium loads varied between ten and forty seconds, and fast loads were under ten seconds. Two levels of aesthetic content were obtained by designing webpages that were text only with no colour or text with colour and images. Each of the six conditions (three levels of loading speed and two levels of aesthetic content) was presented four times to each participant, for a total of 24 webpages viewed by each participant. The order of presentation of the 24 pages was randomized.

3.1.4. *Procedure*: Prior to starting the first trial, the experimenter presented the experimental protocol, answered any questions the participant had, and then the participant read and signed an informed consent form.

For each trial, participants sat down at the computer and were presented with a link. When ready, the participant was instructed to click on the link and view the webpage that was presented. Participants were instructed to imagine themselves reading news online while surfing the Internet at their leisure. Immediately after viewing each page, the participant rated their overall preference in the adjacent experimental room. The rating was performed immediately after viewing each webpage. While the participant rated the page, the experimenter simultaneously prepared the next trial by modifying the Internet connection and displaying the link for the next webpage to be viewed.

3.1.5. *Performance measures*: An overall user preference rating considering both the loading speed and the aesthetic appeal on a visual analog scale from 1 (low overall preference) to 9 (high overall preference) was recorded for each webpage.

3.2. *Results and Discussion*

A 2 x 3 repeated measures analysis of variance (ANOVA) was performed for all participants in each condition (168 data points from seven participants, two levels of aesthetic content, and three levels of loading speed each presented four times). The statistical analysis was a within-subject design. The within-subject measures included webpage aesthetic content (text only with no colour and text with colour and images) and webpage loading speed (slow, medium, fast).

Figure 3 displays the mean overall user preference for loading speed by each aesthetic content condition. The main effect of loading speed (slow mean=4.43, medium mean=5.64, fast mean=6.73) was significant ($p<0.0001$). Figure 4 displays the mean overall user preference for aesthetic content by each loading speed condition. The main effect of aesthetic content (text only with no colour mean=4.45 and text with colour and images mean=6.75) was also significant ($p<0.0001$). There was no significant interaction between webpage aesthetic content and webpage loading speed ($p=0.791$).

[insert figure 3 about here]

Figure 3: Overall User Preference of Loading Speed over Two Levels of Aesthetic Content

[insert figure 4 about here]

Figure 4: Overall User Preference of Aesthetic Content over Three Levels of Loading Speed

Results show that overall user preference increases as aesthetic content increases. Overall user preference also increases as loading speed and variability decreases. However, overall user preference did not differ between webpages with low aesthetic content and fast loading speeds and webpages with high aesthetic content and slow loading speeds. While it is not surprising that users dislike webpages that are slow and lack visual appeal, this experimental finding does offer quantitative and empirical evidence that users are willing to sacrifice loading speed for a more

aesthetically appealing webpage. It is therefore important to know which appearance-related design variables are effective in creating a higher aesthetic appeal and how these variables may affect user's performance. Two experiments were conducted as our first steps toward addressing this issue.

4. Experiment Two

In Experiment 2, two design variables related to webpage appearance—image size and font size, were used to examine how they affect the user's aesthetic preference, ease-of-use judgement, and performance. These two variables were selected because they are among the most commonly used webpage design variables for achieving appearance effects, which is the focus of this and the following experiment.

4.1. *Method*

4.1.1. *Participants*: A total of 20 people (eight males and 12 females) with the age range 22-29 years participated in Experiment Two and Experiment Three. Each had normal (20/20) or corrected-to-normal vision and normal colour vision. All participants had at least four years of prior Internet experience and accessed the Internet on average at least two hours a day. The participants were compensated $10 for approximately one hour of their time.

4.1.2. *Apparatus*: Twenty webpages were created based upon The New York Times on the Web (2004) for each participant, 400 in all. A total of twenty different news articles were selected for

the experimental stimuli and then modified to reduce the length to between 180 and 190 words. There were two images that corresponded to each article, a small image that ranged from 124-300 pixels high by 175-200 pixels wide, and a large image that ranged from 224-382 pixels high by 552 pixels wide. The two image sizes corresponded to the small picture/large picture format of The New York Times on the Web. There were ten font sizes possible for each article: 7.5, 8.5, 9, 9.5, 10, 10.5, 11, 12, 13, or 14 point. These font sizes were selected as representatives of 'PC Magazine's 2003 Classic 100 News and Entertainment Webpages' (2003).

The content of the article, the image size, and the font size varied for each webpage. Twenty articles, two image sizes, and ten font sizes combined to create four hundred unique pages. No article content/image size/font size condition was replicated within-subject or between-subjects. Each participant viewed each article once, each image size ten times, and each font size twice.

Each webpage was presented to the participant on a traditional 17 inch CRT visual display terminal with 60 Hertz refresh rate and 1280x1024 pixel resolution. The webpages were saved on a local 1.8 GHz laptop computer that was used to view the pages. Participants used a mouse with a scroll wheel as an input device. The experimenter sat facing the participant and had another display of what the participant was viewing. Between trials, the experimenter used the laptop's touch pad to prepare the next experimental trial for the participant.

4.1.3. *Conditions*: The experiment was a repeated measures design with two within-subject factors. The within-subject factors were the image size and the font size. Two levels of image size were accomplished by using either the small or the large image that accompanied the article. Ten levels of font size were obtained by modifying the font size in the HTML code. Each of the

20 conditions (two levels of image size and ten levels of font size) was presented once to each participant, for a total of 20 webpages viewed by each participant. The order of presentation of the 20 webpages was randomized for each participant.

4.1.4. *Procedure*: Prior to starting the first trial, the experimenter presented the experimental protocol, answered any questions the participant had, and then the participant read and signed an informed consent form. Participants were then instructed to sit at the computer workstation facing the experimenter. The participant was told to please adjust the position of the mouse, monitor, and chair to be most comfortable for them. Focal distance from the participant's eyes to the monitor remained approximately the same for all participants.

The participants were instructed to imagine themselves reading news online while surfing the Internet at their leisure. For each of the 20 experimental trials, a dummy page was displayed and the participants were instructed to click on a link that would open a webpage. The participant then read the article text displayed at their own pace and then clicked a link at the bottom of the page when finished.

At the conclusion of each trial, participants answered a four-choice multiple choice comprehension question about the article and then rated their aesthetic preference for the webpage and their ease of interaction with the webpage, respectively, each on a scale from 0 (low) to 10 (high). Participants were told not to consider the content of the article when doing the rating. The total time the participant needed to finish reading each page was also recorded.

4.1.5. *Dependent measures*: Three dependent measures were collected for each webpage trial: An aesthetic preference rating on a scale from 0 (low aesthetic preference) to 10 (high

aesthetic preference), an ease-of-interaction rating on a scale from 0 (hard) to 10 (easy), and the time needed to finish reading a webpage.

4.2. *Results and discussion*

A 2 x 10 repeated measures analysis of variance (ANOVA) was performed for all participants in each condition (400 data points from 20 participants, 2 image sizes, and 10 font sizes). The statistical analysis was a within-subject design. The independent variables were webpage image size (small or large) and webpage font size (7.5, 8.5, 9, 9.5, 10, 10.5, 11, 12, 13, or 14 point).

Figure 5 displays the mean interaction time, the mean aesthetic preference rating, and the mean ease of interaction rating by each image size condition. The main effect of image size on interaction time was not significant ($p=0.1771$), nor was the main effect of image size on ease of interaction ($p=0.0887$). The main effect of image size on aesthetic preference was significant ($p=0.0253$).

[insert figure 5 about here]

Figure 5: Mean Interaction Time, Aesthetic Preference, and Ease of Interaction over Image Size

Figure 6 displays the mean interaction time, the mean aesthetic preference rating, and the mean ease of interaction rating by each font size condition. The main effect of font size on interaction time was not significant ($p=0.4913$), but the main effect of font size on user aesthetic

preference was significant ($p<0.0001$). The main effect of font size on ease of interaction was also significant ($p<0.0001$).

[insert figure 6 about here]

Figure 6: Mean Interaction Time, Aesthetic Preference, and Ease of Interaction over Font Size

The main finding of this experiment is that user aesthetic preference and ease of interaction ratings increase as font size increases, and user aesthetic preference increases as graphic size decreases, although user performance was not affected by the two design variables. These quantitative and empirical findings provide useful insight to designers that performance alone may not a good indicator of webpage effectiveness, which may be reflected through user's aesthetic preference or ease-of-use judgements.

5. Experiment Three

To continue the investigation started with Experiment 2 on the role of appearance-related design variables in webpage design, two variables, link style and column width, were examined in Experiment 3 to see how they affect user's aesthetic preference, ease-of-use judgement, and performance. These two variables were selected because they are among the most commonly used webpage design variables for achieving appearance effects.

5.1. *Method*

5.1.1. *Participants*: Experiment Three participants are the same as those participants in Experiment Two.

5.1.2. *Apparatus*: Forty webpages were created, based upon four articles from The New York Times on the Web (2004). The four news articles had the following number of words and paragraphs: 1) 645 words, 14 paragraphs; 2) 643 words, 14 paragraphs; 3) 756 words, 20 paragraphs; and 4) 749 words, 20 paragraphs. Each of the four articles represented the 10 experimental conditions once, making a total of 40 webpages.

There were two levels of column width and five levels of link style. The narrow column width was 60% of the wide column width. The five different link style combinations were also selected for each article: 1) black underlined; 2) blue underlined; 3) blue not underlined; 4) navy blue not underlined; and 5) blue underlined when scrolled over, not underlined otherwise. These link styles were selected because they are commonly used. For example, they appear on PC Magazine's 2003 Classic 100 News and Entertainment webpages (2003) or on The New York Times on the Web (2004).

Each of the six links in every article were words 7-9 letters in length, were never repeated in the document, were never the first word of the paragraph, and were never proper nouns or abbreviations. No word chosen to be a link was ever reproduced within-subject. Each of the 40 webpages had a unique combination of article content, column width, link style, and link content.

Each webpage was presented to the participant on a traditional 17 inch CRT visual display terminal with 60 Hertz refresh rate and 1280x1024 pixel resolution. The webpages were

saved on a local 1.8 GHz laptop computer that was used to view the pages. Participants used a computer mouse with a scroll wheel as an input device. The experimenter sat facing the participant and had another display of what the participant was viewing. Between trials, the experimenter used the laptop's touch pad to prepare the next experimental trial for the participant.

5.1.3. *Conditions*: The experiment was a repeated measures design with two within-subject factors. The within-subject factors were the column width and link style. The two levels of column width and five levels of font style were accomplished by modifying the webpage code. Each of the 10 conditions (two levels of column width and five levels of font style) was presented four times to each participant, for a total of 40 webpages viewed by each participant. The order of presentation of the 20 webpages was randomized.

5.1.4. *Procedure*: Prior to starting the first trial, the experimenter presented the experimental protocol, answered any questions the participant may have had, the participant then read and signed an informed consent form. Participants were then instructed to sit at the computer workstation facing the experimenter. The participant was told to adjust the position of the mouse, monitor, and chair to be most comfortable for them. Focal distance from the participant's eyes to the monitor remained approximately the same for all participants.

Participants viewed a webpage containing a link to the experimental webpage at the beginning of each trial. Before clicking the link to the experimental webpage, the participant was told a word that was the trial's target link. Upon viewing the experimental webpage, they were instructed not to read the article, but simply to search for the target link. Upon locating the target

link, they were told to click it immediately. There were always six links, so the participant was targeting one of the six links. Although each article was viewed ten times each, none of the six links were ever the same.

At the conclusion of each trial, participants answered a four-choice multiple choice comprehension question about the article and then rated their aesthetic preference for the webpage and their ease of interaction, respectively, each on a scale from 0 (low) to 10 (high). Participants were told not to consider the content of the article when rating the webpages. The total time the participant spent viewing each webpage (from the time they clicked the dummy link to the time they clicked the target link) was also recorded.

5.1.5. *Dependent measures*: Three dependent measures were collected for each webpage trial: An aesthetic preference rating on a scale from 0 (low aesthetic preference) to 10 (high aesthetic preference), an ease-of-interaction rating on a scale from 0 (hard) to 10 (easy), and the time needed to search webpage and find the target link.

5.2. *Results*

A 2 x 5 repeated measures analysis of variance (ANOVA) was performed for all participants in each condition (400 data points from 20 participants, 2 levels of column width, and 5 levels of link style each presented 4 times). The statistical analysis was a within-subject design. The within-subject measures included webpage column width (narrow or wide) and webpage link style (black underlined; blue underlined; blue not underlined; navy blue not underlined; and blue underlined when scrolled over, not underlined otherwise).

Figure 7 displays the mean interaction time, the mean aesthetic preference rating, and the mean ease of interaction rating for each image size condition. The main effect of column width on interaction time was not significant ($p=0.7168$), nor was the main effect of column width on ease of interaction significant ($p=0.8144$). The main effect of column width on user aesthetic preference was significant ($p<0.0001$).

[insert figure 7 about here]

Figure 7: Mean Aesthetic Preference, Ease of Interaction, and Interaction Time over Column Width

Figure 8 displays the mean interaction time, the mean aesthetic preference rating, and the mean ease of interaction rating by each link style condition. The main effect of link style on interaction time was not significant ($p=0.1010$), although the main effect of link style on user aesthetic preference was significant ($p<0.0001$). The main effect of link style on ease of interaction was also significant ($p<0.0001$).

[insert figure 8 about here]

Figure 8: Mean Aesthetic Preference, Ease of Interaction, and Interaction Time over Link Style

Results show that users aesthetically preferred the wider column width. Their aesthetic preference also varies with link style, blue with underline being the most aesthetically preferred,

then blue, then blue with scroll over underline, then black with underline, and finally navy blue without underline being the least aesthetically preferred. Ease of interaction also varies with link style; users finding blue with underline the easiest, followed by blue without underline, then blue with scroll over underline, then black plus underline, finally navy blue without underline being the hardest to interact with. The link style results are of interest because it is apparent that the colour blue, which is traditionally used for link, thereby perceptually coded, is truly more aesthetically preferred and easier to interact with.

Although link style and column width contributed to significant differences with respect to ease of interaction and user aesthetic preference, there was no significant difference in interaction time. This result is of particular note to designers because interaction time is often used as a measure of webpage 'goodness,' i.e., if a user finds what he/she is looking for in the quickest amount of time, the webpage is better/more preferred. This study has shown that is not necessarily the case.

6. General Discussion

This series of studies yielded 57 webpage design variables that were ranked according to their perceived importance in webpage usability and overall preference and grouped into ten clusters in terms of their underlying relationship. The results show that technical performance and aesthetic factors are both important issues in webpage design. Although this result does not appear to be surprising or counterintuitive, it does offer concrete empirical evidence to this issue.

The three experiments supplied the basis for a more detailed understanding of the quantitative relations among several specific design variables. Experiment 1 showed the users

are willing to sacrifice technical performance to some extent for a more aesthetic interface. Experiments 2 and 3 showed that user performance alone is not a reliable predictor of webpage effectiveness, since user's aesthetic preference ratings and ease-of-use ratings often diverge from their performance. These findings provide useful insights to webpage designers that while performance may be relatively easy to measure, they should consider other factors as well, such as webpage aesthetic preference.

Future study may expand the scope of the research to investigate further issues such as age, type of webpage browsing task (general purpose browsing or a directed fact-finding search), content of webpages (user interest or choice of the researcher, Marchionini & Shneiderman 1988), and different methods of achieving high levels of aesthetic effects.

References

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**, 670-682.

BAUERLY, M. and LIU, Y. 2008, Effects of symmetry and grouping on interface and design aesthetics, *International Journal of Human-Computer Interaction*, 24, 275-287.

BERNARD, M. L., CHAPARRO, B. S., MILLS, M. M. and HALCOMB, C. G. 2003, Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text, *International Journal of Human-Computer Studies,* **59**, 823-835.

BURSTEIN, C. D. 2002, *Viewable with any browser: Accessible site design guide* [online]. Available from: http://www.anybrowser.org/campaign/abdesign.html [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

COX, J. and DALE, B. G. 2002, Key quality factors in web site design and use: An examination, *International Journal of Quality and Reliability Management,* **19**, 862-888.

DE GRAAFF, H. 2004, *HCI index* [online]. Available from: http://degraaff.org/hci/ [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

DENNERLEIN, J. T. and JOHNSON, P. W. 2006, Different computer tasks affect the exposure of the upper extremity to biomechanical risk factors, *Ergonomics,* **49**, 45-61.

DUNN-RANKIN, P. 1983, *Scaling Methods* (Hillsdale, N.J.: Erlbaum).

ERICSSON, M. 1999, *HCI resources: Guidelines, styleguides, standards* [online]. Available from: http://www.ida.liu.se/labs/aslab/groups/um/hci/guidelines.html [Accessed 18 Mar 2008, originally accessed 16 Feb 2004 from: http://www.ida.liu.se/~miker/hci/guidelines.html].

GIBBS, S. and SZENTIVANYI, G. 1997, *Index to multimedia information sources* [online]. Available from: http://sjcpl.lib.in.us/rhshomepage/multimedia.html [Accessed 18 Mar 2008, originally accessed 16 Feb 2004 from: http://viswiz.gmd.de/MultimediaInfo/].

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

HOM, J. 1998, *The usability methods toolbox* [online]. Available from: http://jthom.best.vwh.net/usability/usable.htm [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

INSTONE, K. 2003, *Usable web* [online]. Available from: http://instone.org/ [Accessed 18 Mar 2008, originally accessed 16 Feb 2004 from: http://usableweb.org/].

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.

KUROSU, M., and KASHIMURA, K. 1995, Apparent usability versus inherent usability. In *CHI'95 Conference Companion*, (New York: ACM/Addison Wesley), 292-293.

LAVIE, T. and TRACTINSKY, N. 2004, Assessing dimensions of perceived visual aesthetics of web sites, *International Journal of Human-Computer Studies,* **60**, 269-298.

LIU, Y. 2003a, Engineering aesthetics and aesthetic ergonomics: Theoretical foundations and a dual-process research methodology, *Ergonomics*, **46**, 1273-1292.

LIU, Y. 2003b, The aesthetic and the ethic dimensions of human factors and design. *Ergonomics*, **46**, 1293-1305.

MARCHIONINI, G. and SHNEIDERMAN, B. 1988, Finding facts vs. browsing knowledge in hypertext systems, *IEEE Computer,* **21**, 70-80.

MARION, C. 2003, *Software design smorgasbord* [online]. Available from: http://mysite.verizon.net/resnx4g7/ [Accessed 18 Mar 2008, originally accessed 16 Feb 2004 from http://www.chesco.com/~cmarion/]

NEW YORK TIMES. 2004, *The New York Times* [online]. Available from: http://nytimes.com/ [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

NIELSEN, J. 2004, *Jakob Nielsen on usability and Web design* [online]. Available from: http://www.useit.com/ [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

NIELSEN//NETRATINGS. 2008, *Global Internet index: Average usage* [online]. Available from: http://www.nielsen-netratings.com/news.jsp?section=dat_gi [Accessed 18 Mar 2008].

OLIVER, K. 2002, *Web interface design* [online]. Available from: http://www.edtech.vt.edu/edtech/id/interface/index.html [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

PC MAGAZINE. 2003, *PC Magazine's top 100 web sites* [online]. Available from: http://www.pcmag.com/article2/0,4149,908003,00.asp [Accessed 18 Mar 2008].

PEARSON, R. and VAN SCHAIK, P. 2003, The effect of spatial layout of and link colour in web pages on performance in a visual search task and in interactive search task, *International Journal of Human-Computer Studies*, **59**, 327-353.

PERLMAN, G. 2004, *HCI sites* [online]. Available from: http://www.hcibib.org/hci-sites/ [Accessed 18 Mar 2008, originally accessed 16 Feb 2004].

THURING, M. and MAHLKE, S. 2007, Usability, aesthetics and emotions in human-technology interaction, *International Journal of Psychology*, **42**, 253-264.

TRACTINSKY, N. 1997, Aesthetic and apparent usability: empirically assessing cultural and methodological issues. In *Proceedings of CHI'97*, (New York: ACM/Addison Wesley), 115-122.

TURNER, S. 2002, The HEP test for grading web site usability. *Computers in Libraries*, **22**, 37-39.

U.S. NEWS & WORLD REPORT. 2004, *usnews.com* [online]. Available from: http://www.usnews.com [Accessed 18 Mar 2008, originally accessed 5 Mar 2004].

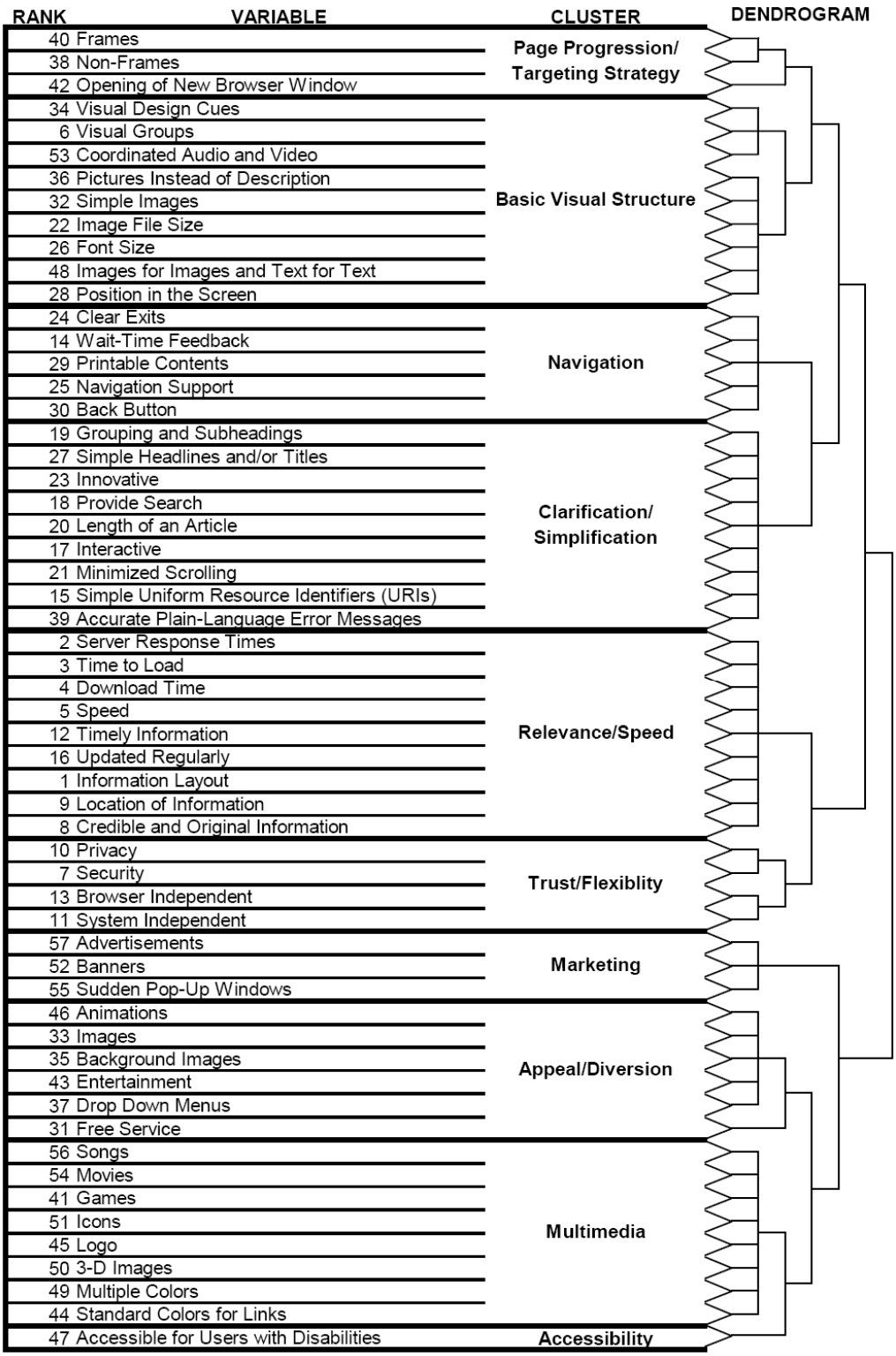

Figure 1. Ten Clusters Consisting of 57 Ranked Variables and the Corresponding Dendrogram

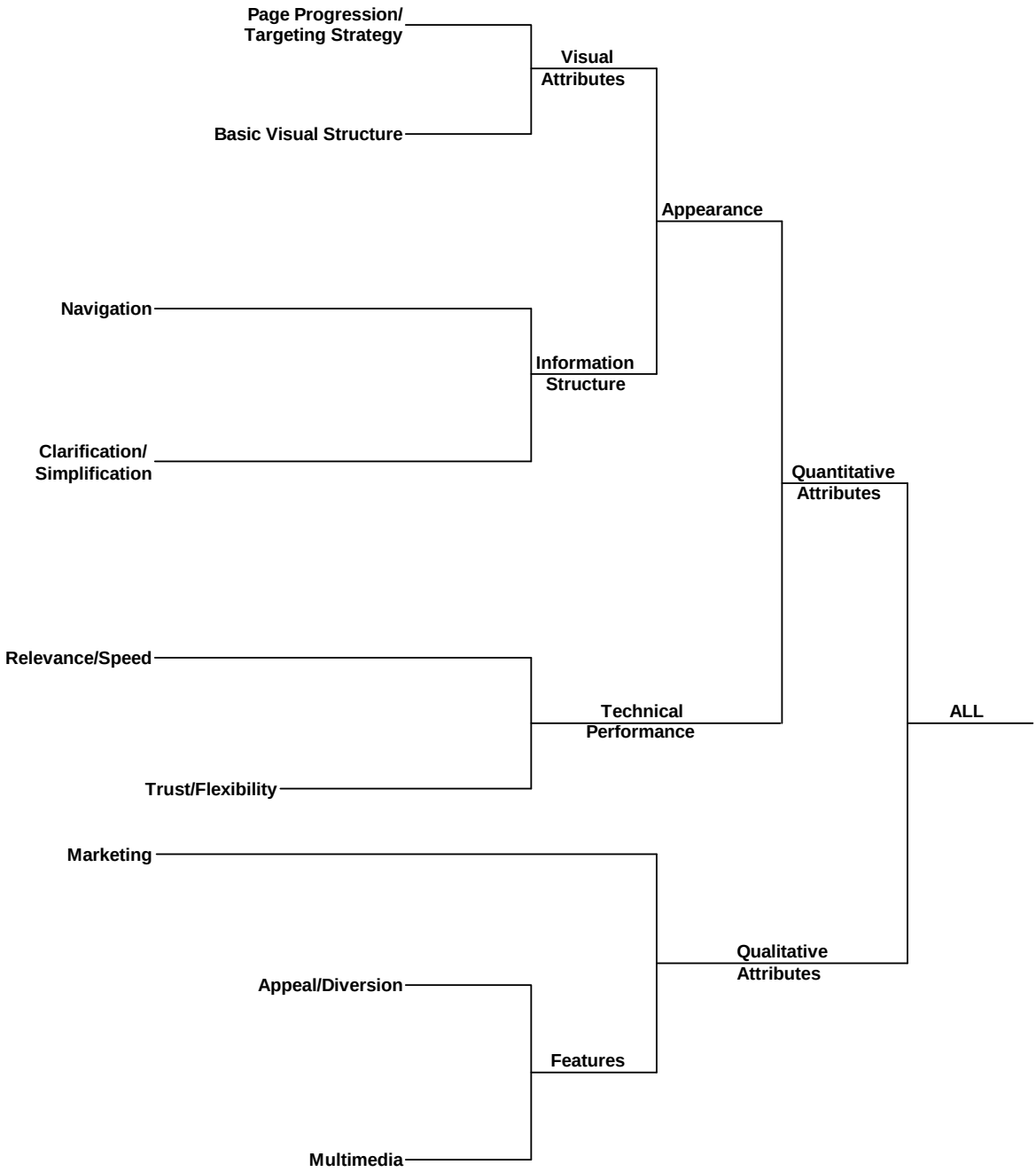

Figure 2. High-Level Definition of Dendrogram Clusters

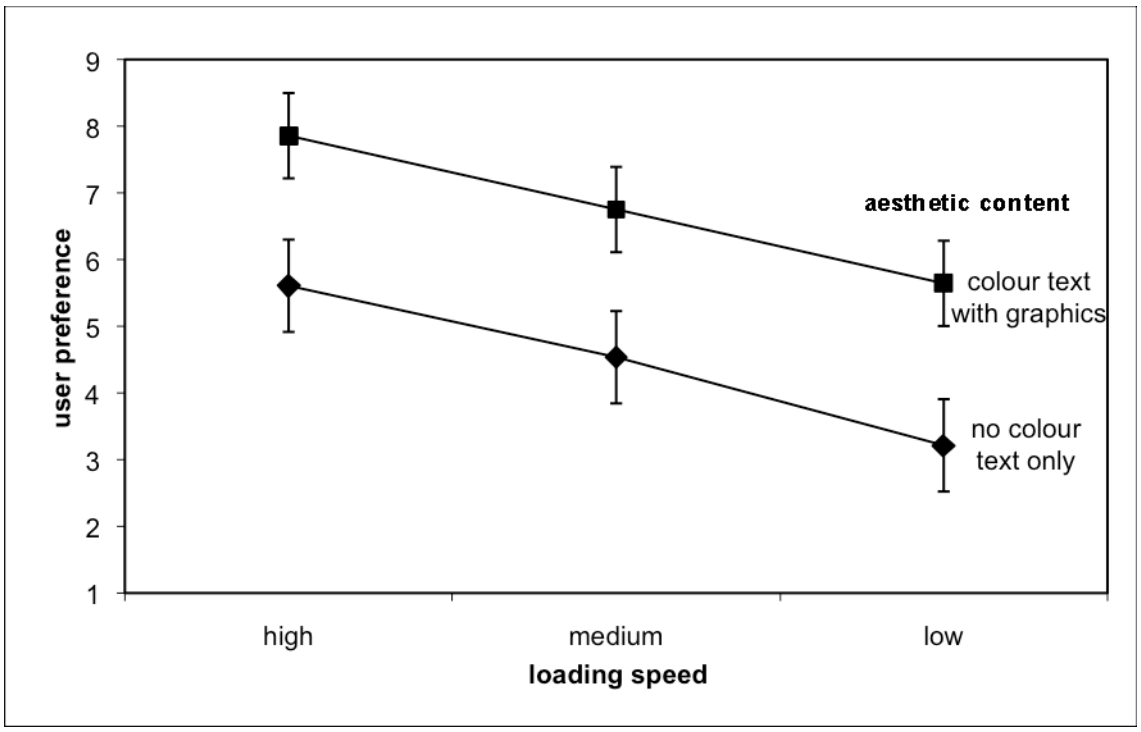

Figure 3. Overall User Preference of Loading Speed over Two Levels of Aesthetic Content

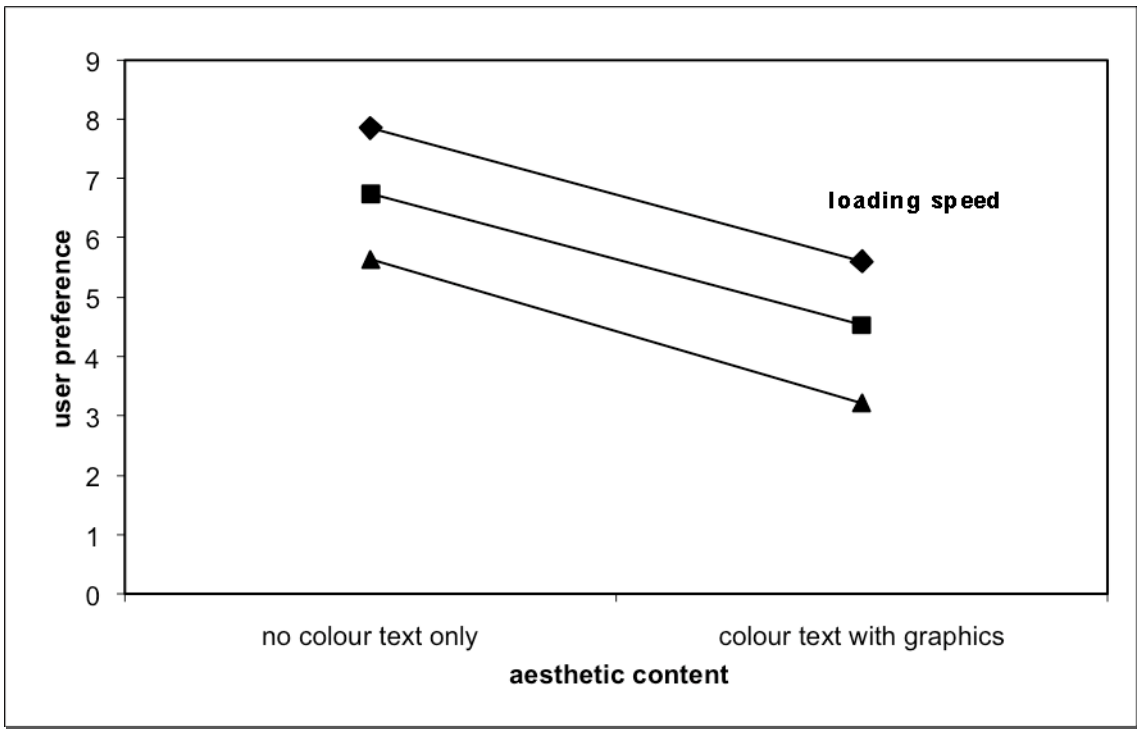

Figure 4. Overall User Preference of Aesthetic Content over Three Levels of Loading Speed

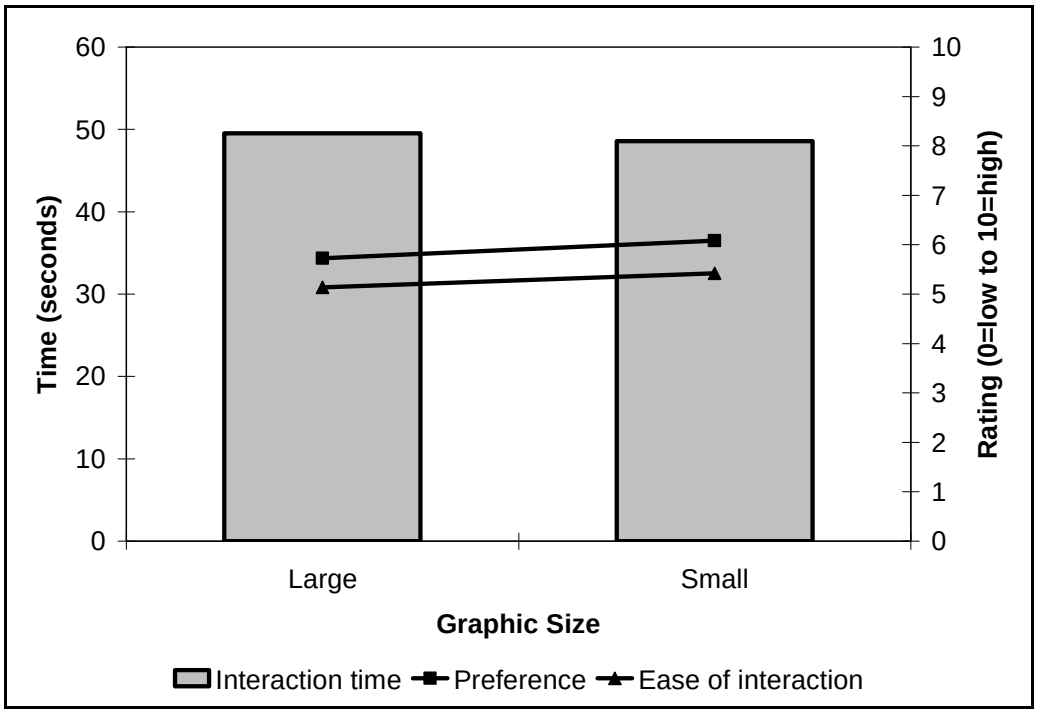

Figure 5. Mean Interaction Time, Aesthetic Preference, and Ease of Interaction over Image Size

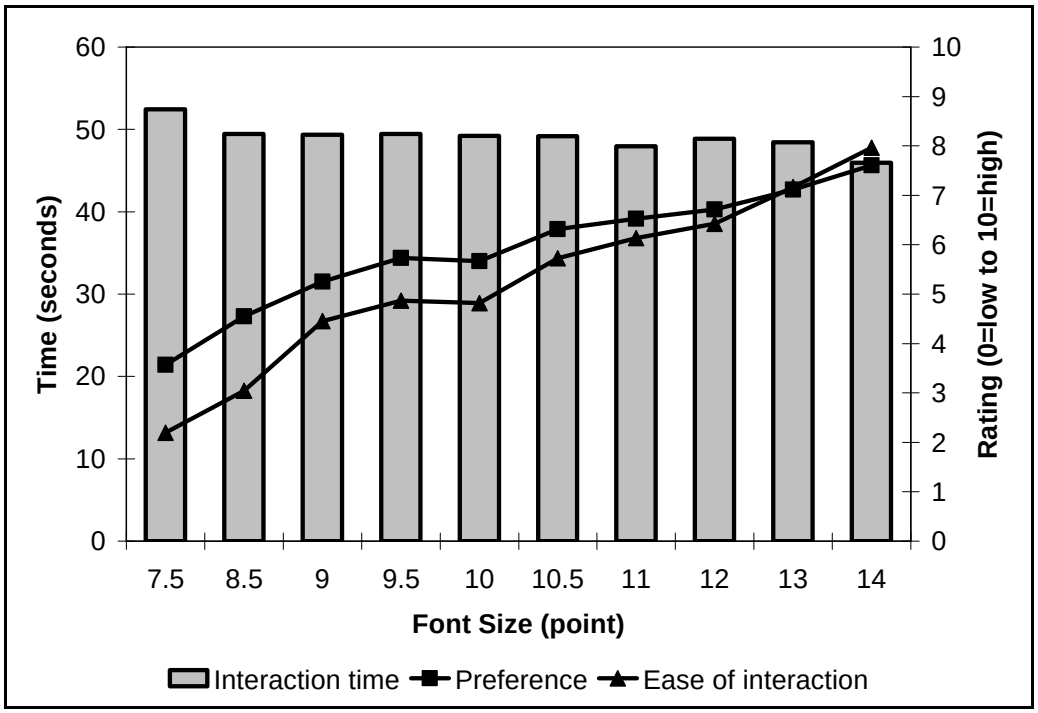

Figure 6. Mean Interaction Time, Aesthetic Preference, and Ease of Interaction over Font Size

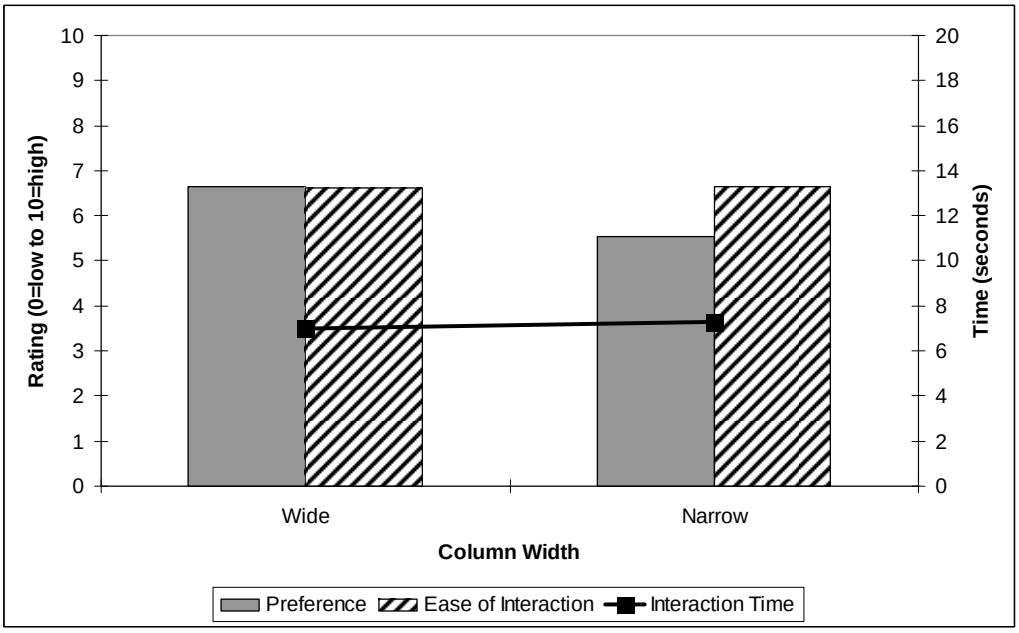

Figure 7. Mean Aesthetic Preference, Ease of Interaction, and Interaction Time over Column Width

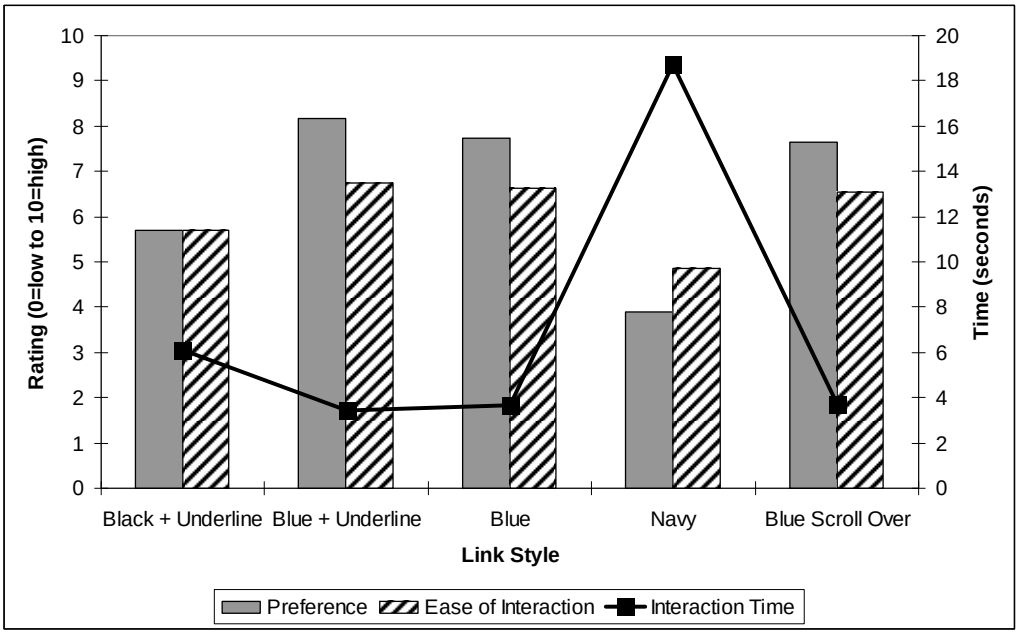

Figure 8. Mean Aesthetic Preference, Ease of Interaction, and Interaction Time over Link Style

Figure 1. Ten Clusters Consisting of 57 Ranked Variables and the Corresponding Dendrogram

Figure 2. High-Level Definition of Dendrogram Clusters

Figure 3: User Preference of Loading Speed over Two Levels of Aesthetic Content

Figure 4: User Preference of Aesthetic Content over Three Levels of Loading Speed

Figure 5: Mean Interaction Time, Aesthetic Preference, and Ease of Interaction over Image Size

Figure 6: Mean Interaction Time, Aesthetic Preference, and Ease of Interaction over Font Size

Figure 7: Mean Aesthetic Preference, Ease of Interaction, and Interaction Time over Column Width

Figure 8: Mean Aesthetic Preference, Ease of Interaction, and Interaction Time over Link Style