VISUAL COMPLEXITY RANKINGS AND ACCESSIBILITY METRICS

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

September 2009

By
Eleni Michailidou
School of Computer Science

Contents

List of Tables

List of Figures

Abstract

The World Wide Web (Web) has become the major means of distribution and use of information by individuals around the world. Web page designers focus on good visual presentation to implicitly help users navigate, understand, and interact with the content. The rapid and constant advancement of technology introduced new ways to present information that leads to visually complex Web pages. Problems arise, though, for people with disabilities, especially those who are visually impaired, because implicit visual cues presented on a Web page cannot be accessed and used.

We assert that, identifying the areas that are complex for sighted users will have direct benefits for blind and visually impaired users. We theorise that by understanding sighted users' visual perception of Web page complexity we can understand the cognitive effort required for interaction with that page. This is an important contribution to the Web accessibility area because by using visual complexity, an identifiable measure, as an implicit marker of cognitive load, Web pages can be designed that are easier to interact with.

Results from user evaluations provided statistical models that, based on the density and diversity of Web page structural elements (such as text, tables, and images), can significantly predict sighted users' perception of Web page visual complexity. The framework is then implemented into the ACTF Eclipse framework by extending the aDesigner accessibility tool to the ViCRAM tool. The tool automatically analyses a Web page with respect to its visual complexity. For each Web page a complexity score, that determines the page's level of visual complexity, and an overlay heatmap, that mimics a user's visual complexity perception by noting the areas that are most visually complex, are generated.

A user and technical evaluation support our assertions and show that the tool can significantly predict the level of visual complexity of a Web page. Therefore, users can have an initial perception of the visual layout of the page and designers

can use this framework to balance Web page visual complexity with usability and accessibility.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

I would like to express my deep appreciation to my supervisors, Dr. Simon Harper and Mr. Sean Bechhofer, whose continuous guidance and encouragement helped me complete this project. I have been very fortunate for both of them. They were always available, supportive, open to discussions and taught me how to be a researcher and not just a student. Thanks are own to the School of Computer Science for funding this project and Dr. Robert Stevens who proposed to continue my studies to a PhD.

I am very thankful to all the HCW Lab members, especially to Darren Lunn, Dr. Yeliz Yesilada and Dr. Caroline Jay for all their support through discussions and assistance on coding and statistics. Darren, from a PhD-mate, has become a friend whom unaccounted-for proofreading, technical help, user evaluations assistance and discussions cannot be valued!

Many people volunteered during the user evaluations, for which I would like to thank. I appreciate the support from the External Affairs and Academic Support Offices during the 'visit-days' user evaluations. My HCW Lab colleagues are thanked for participating in all pilot experiments providing valuable comments.

I am also thankful to the IBM Tokyo Research Laboratory members, especially to Dr. Kentarou Fukuda, for their assistance during the implementation phase of this project. The IMG members are also thanked for their friendly atmosphere during dinners and coffee breaks, and especially Yeliz and Dr. Alexandra Nenadic for their support and friendship.

I am grateful to my family and friends (from Cyprus, Greece, UK, Arizona) for all their encouragement during my time completing this thesis and especially to my partner, Vasilis, for his patience, love and support. Special thanks and this thesis dedication go to my parents, Michael and Nitsa, for all their endurement, belief in me and love through these educational years of my life owned to them and provided by them.

Chapter 1

Introduction

The World Wide Web (Web) has been characterized as a "great equalizer", cutting across cultural boundaries, as well as breaking down both personal and geographical barriers [Sullivan & Matson 2000]. It has also become a tool for business, communication, learning, leisure, and a whole host of anticipated and unanticipated activities across a broad spectrum of the population [Germonprez & Zigurs 2003]. Access to these resources "is considered an entitlement by the majority of the industrialized world" [Asakawa 2005].

Most Web pages focus on good visual presentation to implicitly help users navigate, understand, and interact with the content. Research [Hoffos & Sharpless 1991; Faraday & Sutcliffe 1998] shows that the use of visual aesthetics, such as sharp layout, snappy captions and interesting images, can transform a wall of dry text into a presentation which users will approach enthusiastically. Studies expound the value of text and images and suggest approaches that try to effectively design how they should be combined together [Marks 1994; Fischer 1996; Faraday & Sutcliffe 1998]. For example, the use of 'contact points', which are places in the text where the contents need to be related with the image, or the use of co-references between text and images [Faraday & Sutcliffe 1998] are approaches for effectively designing a combination of text and images.

Visual perception is a process of extracting knowledge about objects and events in the environment and depends both on the stimulus and the individual's characteristics [Lavie & Tractinsky 2004]. That is, human cognition is required to recognize the relationship between the stimulus objects, which is achieved by bringing existing knowledge to the environment. Web page design features, such as pattern, animations, and layout, can influence user's visual processing. With information on the Web constantly increasing, more structural elements and style sheets are used to organise this information leading to pages becoming visually complex that impose users with more cognitive load.

Cognitive and semantic aspects of a stimulus play an important role in visual and scene perception [Henderson *et al.* 1999; Rayner 1998]. Eye movements are driven by properties of the visual world and processes in a person's mind [Rayner 1998]. Eye tracking and usability evaluation studies try to investigate and understand user behaviour [Rayner 1998; Jacob 1991; 1995; Freedman 1984; Sanders *et al.* 1979; Lohse & Johnson 1996; Russo & Leclerc 1994] with an increasing interest to Web page behaviour [Granka *et al.* 2004; Jay *et al.* 2006; McCarthy *et al.* 2003; Outing & Ruel 2006]. A general conclusion is that user interaction depends on the visual factors (nearby visual features) and scene semantics (general knowledge about the scene layout). Cognitive overload is a result of the boost of information presented on the Web. Understanding how this information and cognitive overload affects user perception and Web interaction can lead to solutions that improve users' Web experience. We believe that an initial step towards this goal is to understand Web page visual perception and relate a user's implicit understanding of Web page visual complexity with its layout.

Web accessibility refers to the practice of making Web pages accessible to all users, especially to those with disabilities [Paciello 2000]. When Websites are accessible, people with disabilities are able to perceive, understand, navigate, interact and contribute to the Web. Access to, and movement around, complex hypermedia environments, of which the Web is the most popular example, has long been considered an important and major issue in the Web design and usability field [Chen 1997; Furuta *et al.* 1997].

For example, when sighted users reach a Web page, they can scan the page and obtain a comprehension of it in an average of 5 seconds. By that time they have decided if the page is relevant to their task and have moved towards the part of the page that interests them. What happens, though, when visually impaired people try to interact with such resources? Problems arise because the implicit visual cues presented on a Web page cannot be accessed and used by impaired people, especially visually impaired people [Plessers *et al.* 2005]. For example, due to the visual organization a sighted user can easily detect a menu structure, where a screen reader can only detect a collection of links. Therefore, the meaning of the menu concept is lost [Plessers *et al.* 2005]. This happens because they use assistive technologies, such as screen readers, that create an audio rendering of the page. Visually impaired users have to listen to the entire page being read aloud from the top-left corner of the screen to the bottom-right in order to decide if it is useful for their task. These assistive technologies can relate a structural overview when the user accesses a page for the first time. This information includes, for example, the number of headings in the page based on the heading tags in the underlying HTML[1]. However, if the Web page is not appropriately designed and does not comply with accessibility guidelines, this information can be misleading [Yesilada *et al.* 2003b]. In addition, these technologies cannot provide instant access to any part of the page unless the user is familiar with them. Takagi *et al.* [2004] demonstrated that screen readers can take from 10 seconds to 3 minutes to reach the main content of a Web page. This is because some sites provide accessibility features, such as skip-links and heading tags, and depending on user's familiarity and knowledge of the assistive technology can reduce the browsing time. Other sites, though, do not consider accessibility at all increasing browsing difficulty [Takagi *et al.* 2004].

The need to organise and control information overload and attract user's attention with visual elements can cause visually complex pages that impede sighted users in completing their tasks, due to cognitive overload, or visually impaired users, due to accessibility problems. Guidelines[2] and laws[3] have been introduced to force designers to create accessible Websites. However, it was observed that some designers only paid attention to the letter of the guidelines and regulations, without understanding the real importance of following them [Asakawa 2005]. For example, Ivory & Megraw [2005] determined that the number of tables that were used to control page layouts increased within a three year period (2000 - 2003). Considering that W3C guidelines suggests avoiding layout-tables as they hinder visually impaired users when accessing content due to the limitations of the assistive technologies, many designers are still ignorant of these guidelines or fail to understand how impaired users access the Web [DRC 2004].

Other than following the regulations and guidelines for more usable and accessible Web pages, designers and developers have to understand how blind and visually impaired people use the Web. This will help them determine the elements of the page that are more usable and less accessible in order to find ways

[1]HTML - HyperText Markup Language
[2]W3C Web Accessibility Initiative, http://www.w3.org/WAI
[3]Section 508, http://www.section508.gov

of decreasing accessibility problems. Poor coding is an additional problem and as recent research on Web accessibility shows, most Web sites fail to satisfy even the basic accessibility requirements [DRC 2004]. Even though there is an increase in the awareness of Web accessibility [Asakawa 2005], and regulations and guidelines exist, Web site navigation by blind users is still difficult [Harper & Yesilada 2007].

We assert that, identifying the areas that are complex for sighted users will have direct benefits for blind and visually impaired users. Takagi *et al.* [2007] posit that the scanning navigation of blind users within the page corresponds to the eye-movements of sighted users moving between Areas of Interest. This conceptualisation of navigation is different to how it is used for sighted users as navigation commonly refers to movement via links between pages. Because, the term 'eye–movements' cannot be used for blind users, Takagi *et al.* [2007] adopt the term hypertext navigation to mean link following, and scanning navigation to mean eye–movements. They confirm their thoughts that the scanning navigation of blind users within the page is comparable to the eye–movements of sighted users in their second study (Takagi *et al.* [2007] §4 page 30). They elaborate that scanning navigation is found to comprise of exhaustive element by element scanning with just an audio glance at the start of each element; or gambling strategy scanning in which multiple elements are jumped to see if faster progress is made to the desired information, followed by exhaustive jumps to home-in on that information.

For both scanning techniques, navigation was based on blind users making use of landmarks [Yesilada *et al.* 2003a; 2007] contained within the page. These landmarks consisted of intended landmarks that were inserted into the page by the author, such as heading tags, paragraph tags, and list tags that added structure to the content of the page, and unintended landmarks that contained strong information scent, such as meaningful or relevant text. Therefore, a page that has an increased number of landmarks can affect a blind user's navigation by increasing the amount of time spent for interaction. Indeed, in Yesilada *et al.* [2008], an eye tracking study was conducted that showed a strong relationship between Web page areas that receive visual attention by sighted users and the objects found in the classification system developed by Yesilada *et al.* [2003a] and supported by Takagi *et al.* [2007].

In this case both strategies would be affected by the visual complexity of Web pages - because the visual complexity is directly related to the structural complexity of the underlying code. Therefore, understanding the visual complexity, ergo the structural complexity, will have a direct benefit to blind users.

As demonstrated by Yesilada *et al.* [2008], sighted users tend to mostly fixate on the same structural elements used by blind users to build their navigational model. Thus, an increased complexity for sighted users implies an increased complexity for blind users.We thus examine the relationship between visual complexity and the diversity and density of structural elements contained within a page and as initially demonstrated in Harper *et al.* [2009] a positive relationship exists. Therefore the more structural elements a page has the more complex it is for sighted users to interact with due to cognitive overload. As Takagi *et al.* [2007] showed, these same structural elements are used by blind users to build their navigation model.

1.1 Problem Statement

The visual complexity of Web pages has been shown to affect the difficulty of using Web pages but is also regarded as a subjective decision by the user [Harper *et al.* 2009]. In the thesis we attempt to describe visual complexity and understand how it can be used to benefit Web interaction and design. We theorise that by understanding sighted users' visual perception of Web page complexity we can understand the cognitive effort required for interaction with that page.

Visually impaired users find many Web pages to be complex from an interaction standpoint. However, measuring interaction requirements is normally performed individually based on the user, the page, and the task at hand, such as searching or browsing. While it takes sighted users less than 5 seconds to gain a comprehension of the Web page it can take visually impaired users much longer (see Chapter 3). Indeed, studies [Hoffos & Sharpless 1991; Faraday & Sutcliffe 1998] have shown that if a page is too visually complicated visually impaired users will often not even try to interact with the content or give up after a short time [Lunn 2009]. Therefore, if it is possible to give a visually impaired user some notice of the expected complexity of the interaction required before link traversal then the time wasted on unproductive audio interaction, scanning, and glancing could be reduced.

By using visual complexity, an identifiable measure, as an implicit marker of cognitive load, we can contribute to the Web accessibility area. This is because Web pages can be designed that are easier to interact with, especially for visually impaired users.

1.2 Research Questions

The aim of this research is to investigate the relationship between visual complexity and general cognitive complexity. We aim to understand the way in which sighted users interact with Web resources and classify those resources based on visual complexity, and therefore cognitive load. In this way we expect to examine if visual complexity is synonymous with the more general complexity of Web pages. By defining visual complexity in this concept, we plan to give designers the ability to check their pages to help them balance visual complexity with usability and accessibility. In order to reach our aim a series of questions arise which are discussed below.

Q1 Does a relationship exist between visual complexity, Web interaction and accessibility? We believe that the increased number of visual landmarks on a page can impede visually impaired user navigation by increasing the time and effort to reach the information they seek. The relationship between Web accessibility and visual complexity will be investigated through the literature and by observation. We assert that the number of visual landmarks on a Web page affects the level of visual complexity. As explained before, visually impaired users, and specifically blind users, were found to perform hypertext navigation in a similar fashion to sighted users, with scanning navigation that corresponds to sighted users' eye movements [Takagi *et al.* 2007]. Therefore, we believe that by interpreting how sighted users perceive the visual presentation of a Web page, how they rate a Web page with respect to its visual complexity, and how they interact with it will provide a generalizable sample set. Investigating sighted users' Web behaviour using eye movement tracking methods gives a better understanding of users' perception and cognition of page presentation. Users' glancing habits provide important information on where users look first when they reach a Web page, where they pay more attention and for how long they concentrate on specific parts of the page. By using eye movement analysis we will examine the interplay between users' Web behaviour and page design, structure and accessibility.

Q2 Does visual complexity depend on Web page design? Design elements and techniques constantly change in order to create more attractive sites. Studies try to identify the design characteristics of the most effective sites, how these characteristics change over time and how users' cognitive load is affected. A Web page is designed by a combination of a large set of variables such as images, text, tables and links that vary in types such as colour, size, and position. We aim to examine through user evaluations how Web page layout and explicitly the structural elements used to design a page affect users' complexity and aesthetic perception. Conclusions will be drawn about user perception and cognition in relation with the visual presentation and structure of Web pages and will provide the basis on defining visual complexity.

Q3 Can Web page layout be used to predict the visual complexity of a Web page? Results from questions 1 and 2 will be built into a framework that, based on the density and diversity of Web page structural elements such as text, tables, and images, will be used to predict the level of visual complexity. The framework will then be implemented into a tool that can automatically analyse a Web page with respect to its visual complexity. In this way, for each Web page, users and designers will be provided with a complexity score that determines the page's level of visual complexity and an overlay heatmap that notes the areas that are most visually complex. In order to answer this question and support our initial assertions, a user and technical evaluation will take place to examine the visual complexity prediction significance.

1.3 Thesis Structure

The rest of this report is structured as follows:

Chapter 2: Background and Related Work - A literature review of the state of the art in Web accessibility showed that even if regulations and guidelines exist they are not always followed, resulting into the need for better understanding of page design. Investigating visual perception revealed that the link between page layout and perception relates to print documents only. Since the Web is an interactive medium the investigation of user behaviour is an important factor into understanding visual perception. Eye tracking studies are currently used but mainly evaluate a Website with respect to its usability and accessibility level. We believe that identifying the connection between Web page layout and user cognition will give further insights into users' perception, interaction and page layout.

Chapter 3: Eye Tracking Experiment - An eye tracking study is presented where sighted users' browsing behaviour on nine Web pages was investigated to determine how the page's visual clutter related to sighted users' browsing patterns. The results showed that salient elements attracted users' attention first, users spent more time on the main content of the page and tended to fixate on the first three or four items of menu lists. Common gaze patterns began at the salient elements of the page, moved to the main content, followed by the header, right column and left column of the page and finished at the footer area. This study showed that eye movement can help us understand how the visual presentation of Web pages influences sighted users' behaviour and perception. It was also determined that the more visually complex a Web page is, based on an initial definition [Michailidou 2005], the more time it takes for users to gaze through it and the more disoriented their scanpath is. We believe that this happens due to the high number of structural elements used to represent the page which acted as the basis for future studies to examine how the structure can predict visual complexity.

Chapter 4: Towards Defining Visual Complexity - Three pilot experiments, two user and an online evaluation, were conducted to understand how page layout affects user's perception. During the studies users' Web page rankings based on visual appearance and the effect that block structure has on user perception were investigated. These studies acted as background work into conducting an online pairwise experiment by defining a set of variables to modify each Web page into its chunk rendering. Corner, blocks, boxes and top-left-corners were defined and identified on a set of Web page's chunk renderings. Users had to compare this set of Web pages based on their visual complexity. Results showed that the visual complexity score was significantly and positively related to the organisation of the page determined by the number of boxes, blocks, corners and top-left-corners.

Chapter 5: Visual Complexity and Aesthetics - Users' perception of visual complexity and aesthetic appearance of Web pages was evaluated via an extended study based on the results derived from Chapter 4. A strong correlation between users' visual complexity perception of a Web page and structural elements was determined. Regression analysis derived models that estimate the visual complexity level of a Web page using the structural elements of the page. These models were used to design a tool that can predict a Web page's visual complexity.

Chapter 6: Visual Complexity Algorithm - Results from previous user evaluations were built into a framework based on the density and diversity of Web page structural elements that provided a visual complexity score prediction. The framework was implemented into the ACTF Eclipse framework, as an extension of the aDesigner software, and automatically analyses a Web page with respect to its visual complexity. For each Web page, users and designers are provided with a score that determines the page's level of visual complexity and an overlay heatmap that notes the areas that are most visually complex. In order to evaluate the effect of the framework and identify issues for tool improvement, a technical evaluation was conducted. Results showed that the tool can significantly predict the level of visual complexity of a Web page when compared to the previously collected user rankings, with space for improvement.

Chapter 7: User Evaluation of the Complexity Algorithm - A final user evaluation was performed to examine the success level of the tool. During an online experiment, users were asked to rank Web pages based on their visual complexity. These rankings were compared with the rankings that the algorithm returned for the same pages. Results showed a strong and significant correlation between users' visual complexity perception of a Web page and the algorithm prediction rankings. This demonstrates that the visual complexity level of Web pages can be automatically predicted up to a certain ratio based on the structural elements of the page. The tool can therefore be used by users to gain an initial perception of the visual layout of the page and designers to improve their designs towards balancing Web page visual complexity with usability, aesthetics and accessibility.

Chapter 8: Conclusions and Future Work - Investigating users' visual perception and Web interaction enabled us to understand the linkage between visual complexity, cognition and Web page layout. This research demonstrated that an initial step into enhancing accessibility for visually impaired users is by understanding sighted users' perception. The last section of this chapter highlights the directions for future research that could enhance our knowledge for designing accessible and less visually complex Web pages. This includes the investigation of the relationship between audio complexity, Web accessibility, visual complexity and aesthetics along with suggestions for further improvement of the tool.

1.4 Publications

Results derived throughout the research led to the following peer-reviewed publications:

1. Eleni Michailidou. **ViCRAM: visual complexity rankings and accessibility metrics** ACM SIGACCESS Accessibility and Computing (86):24-27, 2006. http://doi.acm.org/10.1145/1196148.1196154

   A description of the ViCRAM project was used as an introduction to the research field. The ViCRAM motivation, research goals, status and future plan of the project were described.

2. Simon Harper, Eleni Michailidou and Robert Stevens. 2009. **Toward a Definition of Visual Complexity as an Implicit Measure of Cognitive Load** ACM Transactions of Applied Perception (TAP) 6, 2 (Feb. 2009), 1-18. http://doi.acm.org/10.1145/1498700.1498704

   The results of a preliminary evaluation towards the definition of visual complexity were demonstrated. The evaluation provided evidence that there is a possible link between visual complexity and cognitive load which established the purpose of this research.

3. Caroline Jay, Darren Lunn and Eleni Michailidou. **End User Evaluations** In Web Accessibility: A Foundation for Research, S. Harper and Y. Yesilada, editors, Human-Computer Interaction Series. Springer, London, 1st edition, September 2008. http://www.springerlink.com/content/978-1-84800-049-0

   This chapter outlined the processes that can be used throughout the design life-cycle to ensure Web accessibility, describing their strengths and weaknesses, and discussing the practical and ethical considerations that they entail. The chapter also considers an important emerging trend in user evaluations: combining data from studies of 'standard' Web use with data describing existing accessibility issues, to drive accessibility solutions forward. The personal contributions was derived from the ViCRAM project experience on user evaluations and the effect and contribution that this project has.

4. Simon Harper, Eleni Michailidou, Huangmao Quan. **Fingerprinting the Visual Complexity of Web Pages** ACM Transactions on the Web (TWEB) - In Submission, 2008.

   This paper demonstrated that identifying areas of high complexity can be achieved by identifying areas high in block structure elements. By using computational algorithms we can capture this metric and place Web pages in a sequence which correlates with our previous human sequenced experiments.

5. Eleni Michailidou, Simon Harper and Sean Bechhofer. **Visual Complexity and Aesthetic Perception of Web Pages** SIGDOC '08: In Proceedings of the 26th Annual ACM international Conference on Design of Communication (Lisbon, Portugal, September 22 - 24, 2008). 215-224. http://doi.acm.org/10.1145/1456536.1456581

   Results of an online user evaluation that investigated user perception of the visual complexity and aesthetic appearance of Web pages were described. A strong and high correlation between users' perception of visual complexity, structural elements (links, images, words and sections) and aesthetic appearance (organisation, clearness, cleanliness, interestingness and beautifulness) of a Web page was determined which provided robust support for our research questions.

6. Eleni Michailidou, Simon Harper and Sean Bechhofer. **Investigating Sighted Users' Browsing Behaviour to Assist Web Accessibility** ASSETS '08: In Proceedings of the 10th international ACM SIGACCESS Conference on Computers and Accessibility (Halifax, Nova Scotia, Canada, October 13 - 15, 2008). 121-128. http://doi.acm.org/10.1145/1414471.1414495

   An eye tracking study was presented where sighted users' browsing behaviour on Web pages was investigated to determine how the page's visual clutter is related to sighted users' browsing patterns. A relationship between visual complexity and layout of Web pages and users' perception was revealed that supported our research.

7. Eleni Michailidou, Simon Harper, Grace Mbipom. **Aesthetically Pleasing Web Pages Do Not Impede Accessibility** New Review of Hypermedia and Multimedia Journal - In Submission, 2009.

An empirical investigation into the relationship between users' perception of visual aesthetics and the accessibility level of Web pages is described. Results suggest that aesthetically pleasing pages, especially clean, clear and organised pages, did not impede accessibility. In this way we highlighted the erroneous notion that Web accessibility produces unappealing page designs.

8. Eleni Michailidou, ViCRAM Tool, Accessibility Tools Framework, Eclipse Technology Project, In Commiting Process, 2009.

   The ViCRAM prototype tool is based on the visual complexity algorithm, derived from complexity prediction models generated through user evaluations and provides two features. First, designers are provided with a complexity score that determines the page's level of visual complexity and the second feature is an overlay heatmap that highlights the areas that are most visually complex on the page.

A number of **Technical Reports** written throughout the project are listed in Appendix K.

Chapter 2

Background and Related Work

Web accessibility refers to the practice of making pages on the Web accessible to all users, especially to those with disabilities [Paciello 2000]. Researchers in this area try to find solutions that improve Web accessibility and decrease the gap that exists between impaired and non-impaired users. This chapter is an introduction to the Web accessibility problem. We describe the different components, including content and technologies used to assist users with disabilities. We examine what problems these technologies have, guidelines and regulations formed to decrease Web accessibility problems, along with a variety of evaluation and validation tools that help designers develop a more accessible Web.

In addition, we introduce and explain the idea of providing feedback with respect to the visual content of a Web page to visually impaired people by understanding sighted users' visual perception. As the authors in Takagi *et al.* [2007] explain, visually impaired users, and specifically blind users, perform hypertext navigation in a similar fashion to sighted users and also scanning navigation which corresponds to sighted users' eye-movements (see page 20). Since eyes provide sighted people with an abundance of information about the outside world [Buhmann *et al.* 1999], we will investigate eye movements to examine how sighted users perceive Web page presentation. This chapter is also a literature review on how Web page structural elements and overall design affect users' interaction through researches that relate presentation with usability frameworks and through eye movement behaviour. We believe that providing information about the visual content can assist in designing accessible and less complex pages. This means that the source code of the page is more accessible by browsers and technologies that impaired users use.

2.1 Assistive Technologies

Assistive technologies are products used by people with disabilities to help accomplish tasks that they cannot otherwise complete. When these technologies are used with computers, they are referred to as adaptive software or hardware [Brewer 2004a]. These technologies are continuously designed and improved to facilitate visually impaired people access their computer systems. These include voice recognition systems, screen readers, screen magnifiers, Morse code input devices and alternative keyboards [Nguyen & Petty 1996]. Some of these technologies are used along with graphical desktop browsers, text browsers, multimedia players, or plug-ins. This section focuses on the technologies that support visually impaired users' Web interaction describing their main features along with the problems they face. In this way, a better understanding is formulated on visually impaired user interaction and the needs for improvement are identified and discussed.

2.1.1 Screen Readers

A screen reader is a software application that attempts to identify and interpret what is being displayed on a screen. Screen readers are used by individuals who are visually impaired or have dyslexia. These systems interpret what is displayed on a screen and directs it either to speech synthesis for audio output, or to refreshable Braille for tactile output [Brewer 2004a]. Screen readers rely on the Web page's source code as the basis for audio rendering. Internet browsers create the visual presentations and screen readers do the audio renderings of the presented page. It is important to note that these applications do not understand what they are reading in terms of the document's structure.

Screen readers read from the top left corner of the screen to the bottom right, line by line, one word at a time which makes it even harder for the user to understand and follow the Web page's content. For example, a common Web page layout uses two columns with the main content in the right column and a menu on the left. Visually impaired users, whose screen readers read from the source code, listen first to the entire left column in order to access the right column. Therefore, the ability to move between main page elements, such as paragraphs and links, is limited, time consuming but crucial.

Another example showing the failure of a screen reader to produce the correct output due to the page's source code is the need to use the 'alt' attribute for images. The contents of the 'alt' attribute within an 'img' tag are read aloud by the screen reader software to describe an image. Tables 2.1 and 2.2 are examples of inaccessible and accessible source code, respectively. The example is shown using the HyperText Mark-up Language (HTML). Each table shows the source code for an image of a dog along with the corresponding audio outputs from a well known screen reader, the Home Page Reader IBM [2006]. Table 2.1 shows the HTML code of an image without the 'alt' tag and Table 2.2 is the code of the same image with the use of the 'alt' tag. The screen reader produces a more descriptive output during the second example. The problem of inaccessible elements on Web pages similar to the example is caused by designers and even if regulations exist are not always followed and even if they have helped to raise the awareness of Web accessibility, they are not sufficient to provide Web pages that are accessible to all [Harper *et al.* 2004].

Table 2.1: Inaccessible Code and the Screen Reader Output

Inaccessible code: <img src="dog.jpg" width="200" height="150">
Screen Reader Ouput: Image with no alt text: dog.jpg

Table 2.2: Accessible Code and the Screen Reader Output

Accessible code: <img src="dog.jpg" width="200" height="150" alt="A Golden Retriever sitting in a garden, next to a young girl.">
Screen Reader Ouput: A Golden Retriever sitting in a garden, next to a young girl.

As is shown above, most screen readers access the source code, which provides a more efficient audio output and support. However, as we explain in more detail in Section 2.8, the visual perception of a sighted user is manipulated with the use of various visual aesthetics that produce an attractive Web page. While more technologies are developed to produce and improve the presentation and layout of a Web page, the visual presentation becomes more complex. As Asakawa [2005] described, "interfaces such as JavaScript and Flash are integrated into Web pages" leading to more visually complex pages than screen readers can deal with. When a page's design focus is on the aesthetics that will induce a user to become involved in the website [Hoffmann & Krauss 2004], it makes audio interaction even harder [Yesilada *et al.* 2003b].

Some of the most well known screen readers, such as Jaws [Henter-Joyce 2006] and Window-Eyes [GW-Micro 2006], aim to provide an easier way of accessing Web pages that follow Web accessibility and design guidelines. For example, Jaws provides the Document Presentation Mode (DOM) which allows the user to read tables by row, just as they are presented on the screen. With this feature, the user can change how Jaws structures HTML pages and other virtual documents (such as PDF documents) in a virtual buffer. It supports screen layouts that render the pages in the virtual buffer in a way that is more similar to what a sighted user sees on the screen. That is, all items in a single table row are displayed on the same line, and each column is separated by a vertical bar. That way, the user understands better the overall table design as well as the relationship between table cells in each row. Also, Jaws relies on Microsoft Active Accessibility (MSAA) information when reading menus, which allows greater accuracy when reading menu items.

2.1.2 Browsers

Specialized browsers are Web browsers that can additionally provide audio output. They differ from screen readers in that they can be used as stand alone audio browsers. Some of the most well known self-talking browsers are BrookesTalk [1998], EmacsSpeak [2006], and Home Page Reader [IBM 2006].

BrookesTalk is a speech output browser which is independent of visual browsers and of text to speech software applications as it uses Microsoft speech technology. BrookesTalk focuses on summarising pages and facilitating Web searching and browsing. It is a function key driven Web browser which means that it provides only keyboard access using the function keys. Information retrieval and language processing techniques are used by the browser to provide different views of Web pages. It includes the functionality of a standard Web browser for visually impaired users in that it can break up the text part of a Web page into headings and links and read out paragraphs.

Concentrating mainly on summarizing pages, BrookesTalk also provides several functions including heading list, link list and extracted keyword lists. Heading lists can be considered as a good indicator of page content. However, headings are often represented as images or eyecatching objects rather than standard heading elements. The summary of the page includes title, number of words, links, headings, and images [BrookesTalk 1998] which is important to the user. It also provides different synopses of a Web page to help users decide whether it will be useful to them or not [Zajicek *et al.* 1998]. While the summary provides an overview of the page content, it does not include information about the layout or other information that is important to enable navigation through the page. In addition, users' familiarity with the tool is required for more efficient use.

BrookesTalk offers, also, a special search facility, a configurable text window for visually impaired users and a standard visual browser so that "visually impaired users can work alongside other people who can fully utilise a standard graphical interface" [Zajicek *et al.* 1999]. However, enabling users to search within the Web, does not mean that they can also understand the retrieved search results.

IBM Home Page Reader (HPR) converts Web page contents into text and synthesized speech using the HTML source. It uses a numeric keypad that helps to navigate Web pages and individual keys and combinations of keys to represent control commands [Asakawa & Itoh 1998]. It, especially, provides several reading modes, such as table navigation, frame navigation and items reading mode [IBM 2006], which are different choices of how to read a Web page. Although these reading modes may reduce the number of key combinations and simplify the command language, it may be confusing because users tend to maintain the mode they are in.

HPR also addresses issues concerning orientation and provides two important functionalities: position and summary. The browser gives a simple description about the current location that the user is by giving a percentage of the page with respect to the position. For example, if you are in the centre of the page it says "At 50% of the page". However, for such information to be useful, the user needs to have a good understanding of the overall physical layout of the page. Therefore, there is a danger that the provided information can confuse visually impaired readers rather than improve their spatial orientation [Yesilada 2005]. HPR also provides a summary of the page which includes the title and the number of links, items, forms, frames and tables, but it does not mention how they are related to each other and how the layout of the page is, which is the key for being able to navigate the page.

Most existing browsers, like BrookesTalk and IBM HPR, use HTML source code to render the page in audio by generating the Document Object Model (DOM) representation of the page and then rendering the page based on that model [Yesilada 2005]. This enables them to provide different functionalities to interact with the structure of the Web page, such as a variety of navigation modes (IBM Home Page Reader), or provide additional information of the layout of the page, such as a summary (BrookesTalk). Even if these browsers try to be compliant with conventional Web browsers, they still provide less functionalities or are not as accurate as conventional browsers. In addition, accessing the HTML source code, brings these browsers closer in achieving difficult functionalities but can cause them problems when the source code is not compliant with the accessibility guidelines.

2.1.3 Problems with the Assistive Technologies

As previously explained in this chapter, these technologies often fail to provide useful Web content to visually impaired users. This is due to the software itself and the content presentation. On one hand, most of the software that provide audio output have usability problems. For example, the software users might not be able to alter the page renderings to meet their needs. Features such as keyboard shortcuts, mouse adaptations, content enlargements, and jumping to linked pages [Asakawa & Itoh 1998] are provided only to a limited number of software tools. It is unavoidable that these technologies cannot solve all the problems that visually impaired users face. For example, some Web pages that contain many images, frames, and tables are not easy for a screen reader or browser system to read. If a page contains many images without any alternative text, users have to hear a large number of URL addresses.

On the other hand, even if the user-side technology has become more accessible and usable (with technology's improvement), the content is less accessible due to the page layout and use of visual cues [Asakawa 2005]. This is the main reason that Web content guidelines and regulations are introduced. If designers follow the regulations and guidelines, they will produce more accessible Web content [W3C 2006]. Unfortunately, these guidelines are not always followed.

2.2 Web Accessibility Regulations

Many organizations have rules for producing accessible Web content. Some of them proposed by organizations including the American Foundation for the Blind (AFB), the World Wide Web Consortium (W3C), the Royal National Institute for the Blind (RNIB), IBM, the Norman Group (Nielsen), and the ARA Section508. It is important to note that these guidelines are comparable with the W3C Web Content Accessibility Guidelines (WCAG) [W3C 1999; Harper *et al.* 2001]. This is because WCAG covers the key points from the rest of the organizations as well as the user agents and authoring tools.

Currently, there are three federal laws (in the U.S.) that are associated with Web accessibility: the Americans with Disabilities Act of 1990 (ADA) and Sections 504 and 508 of the Rehabilitation Act of 1973. Section 508 is the most commonly referenced legal guide governing U.S. Web accessibility. This is because it is grounded in the U.S. Access Board's Electronic and Information Technology Accessibility Standards, which in turn is based on the World Wide Web Consortium's WCAG.

This section describes some of the most important regulations and guidelines that were developed by W3C.

2.2.1 W3C's Guidelines

The World Wide Web Consortium's (W3C) Web Accessibility Initiative (WAI) has developed several accessibility guidelines for Website information, software and assistive technologies. These have been separated into three different guidelines (WAI):

Web Content Accessibility Guidelines (WCAG) explain how to make Website's content accessible to people with disabilities. These guidelines are mainly intended for Web content developers, such as page authors, site designers, Web authoring tool developers and Web accessibility evaluation tool developers.

Some examples of the WCAG guidelines are to provide: equivalent alternatives to auditory and visual content, clear and consistent navigation mechanisms, such as orientation information, navigation bars and a site map, increase the likelihood that a person will find what they are looking for at a site, and provide context and orientation information to help users understand complex pages or elements. They also provide guidelines to design for device independence by using features that enable activation of page elements via a variety of input devices.

Authoring Tool Accessibility Guidelines (ATAG) explain how to make authoring tools accessible to people with disabilities. Authoring tools allow people to produce Web pages and Web content. A primary focus of ATAG is defining how tools help Web developers produce Web content that conforms to Web Content Accessibility Guidelines. For example, ATAG require that Web authoring tools generate valid markup, be configured to prompt for accessibility content such as alternative text for images, captions for audio, and descriptions for video. In addition, they require to provide ways of evaluating accessibility, and provide user-configurable editing views without affecting document markup.

User Agent Accessibility Guidelines (UAAG) explain how to make user agents accessible to people with disabilities, particularly by increasing access to Web content. User agents include Web browsers, media players, and assistive technologies, which are software that some people with disabilities use for interacting with computers. UAAG require that browsers and media players provide documentation of accessibility features in an accessible manner, provide access to content through a variety of navigation mechanisms (sequential navigation, direct navigation, searches, and structured navigation), and implement interoperable interfaces to communicate with other software (assistive technologies, the operating environment, and plug-ins) [Brewer 2004b].

Even though guidelines and regulations played an important role in spreading Web accessibility awareness, it was observed that some designers only paid attention to the letter of the guidelines and regulations, without understanding the real importance of following them [Asakawa 2005]. Many designers are still ignorant of these guidelines or fail to understand how impaired users access the Web. Studies like [DRC 2004] show that designers have an inadequate understanding of the needs of disabled users and how to create accessible Websites.

Other studies show that the application of the guidelines is subject to interpretation: two designers applying the same set of guidelines to the same set of pages generate different results [Ivory & Hearst 2001; Yesilada 2005]. Some of the fundamental problems with the existing guidelines [Ivory *et al.* 2003] are that they are often difficult to understand and to apply. In addition, the needs of different types of users may conflict and there is little guidance on how to address these conflicts. Moreover, designers may not be able to address all guidelines at authoring time. Therefore, since guidelines and regulations are not always followed in the right way, if at all, new solutions, approaches and improvements are in need that can assist on the proper designing of tools and Websites.

2.3 Validation Methods

In order to ensure beneficial implementation of the WCAG, it is necessary to have effective testing methods. Evaluation of the accessibility of a Website is more complex than simply validating the markup of a Website. This is because implementation of accessibility requirements is not only a matter of using valid markup, but also of issues such as the appropriateness of equivalent alternatives, and organization of information on a Website [Brewer 2004b]. In addition, by understanding how disable users access the Web, developers can become more expert on identifying accessibility problems. Effective evaluation requires a "combination of automated and manual testing by someone with expert knowledge of WCAG and of relevant testing tools and approaches" [Brewer 2004b]. Hence, a combination of automated evaluation tools, design guidelines and user testing is a better approach for validating Website accessibility and designing sites and tools as well.

This section describes how automated evaluation tools and user testing help validate Web accessibility. It is important to make clear that some of the tools that are described here check for both usability and accessibility of a Web page. It is hard to distinguish between usability and accessibility because they often go together. Accessibility is the ability of being reached where usability is the ability of being used. Making a Website accessible allows someone who is blind, for example, to use it efficiently with a screen reader. On the other hand, making a Website usable means that a person can navigate through the site, learn from it, enjoy using it and achieve his task in an ease manner.

2.3.1 Automated Tools

Currently, there are validation tools that help determine if a Website is accessible. Software programs or online services perform a static analysis of pages or sites regarding their accessibility by evaluating HTML and Cascading Style Sheet (CSS) [WAI 2005]. The W3C separates the tools in two categories: tools that perform tests for a variety of accessibility issues, including tools that test for one or a limited aspect of accessibility, and tools that run on an ongoing basis such as proxies, and Web services. Some of the tools that are now available include AccessEnable [RetroAccess 2002], Lift [UsableNet-Inc. n.d.], WebXACT [WatchFire 2006], aDesigner [Takagi *et al.* 2004], and Doctor HTML [Tongue & Inc. 1997]. Most of these tools check for conformance to Section 508 or W3C guidelines and then generate a report that provides some guidance for repairing the pages.

2.3.1.1 Evaluation of Website Accessibility

Tools like AccessEnable, Lift, aDesigner and WebXACT are used to test for Website accessibility. WebXACT is a Web accessibility desktop testing tool provided originally by WatchFire [2006] and currently part of IBM Rational Software. It was designed for small Websites to help expose barriers to accessibility and encourage compliance with existing accessibility guidelines, including Section 508 of the US Rehabilitation Act and the W3C's WCAG, on a page-by-page basis by generating a suggestions' report. Lift, in addition to the suggestive report, provides wizards to walk developers through code modifications [Ivory *et al.* 2003]. AccessEnable is a Web-based tool that can make some changes to pages automatically and the developer can then download the modified pages.

The above tools also check if Web pages enable other forms of input, such as keyboard or Braille devices. For example, WebXACT looks for keyboard shortcuts for forms and links but does not analyze the shortcuts to see if they are meaningful, consistent with common interface shortcuts, or conflicting with each other [Ivory *et al.* 2003]. Also, Lift checks for the use of cascading menus, which the developers determined to be problematic for users with disabilities [Coyne & Nielsen 2001]. Most checks entail looking for the existence of page and frame titles, alternative text for images and other objects and headings for tables and labels for form elements, so that screen readers can process this information. None of the tools, however, can evaluate whether alternative text, titles, labels, headings or table summaries are meaningful to the users [Ivory *et al.* 2003].

There are also tools such as aDesigner [Takagi *et al.* 2004] that simulate how a page is accessed by a screen reader user. The aDesigner is a disability simulator that helps Web designers ensure that their pages are accessible and usable by visually impaired people. Web developers use aDesigner to test the accessibility and usability of Web pages for low-vision and blind people. Therefore, aDesigner focuses on evaluating usability of pages as opposed to validating them against guidelines. The usability issues addressed by aDesigner include elements such as the degree of color contrast on the page, the ability of users to change the font size, the appropriateness of alternate text for images, and the availability of links in the page to promote navigability. The tool also checks the page's compliance with accessibility guidelines. The result of this analysis is a report that lists the problems that would prevent accessibility and usability problems. With this information, Web content developers get immediate feedback and can make the necessary modifications to address these obstacles before the content is published. Even though aDesigner supports a more sophisticated approach compared to others, designers still need to learn and use a separate tool to repair their pages. However, the use of such tools could be one of the easiest and more reliable solutions into examining a site and understanding how visually impaired users interact with such pages.

2.3.1.2 HTML and CSS Validation

HTML and CSS evaluation tools, focus on the Web page source code. They validate the source code against W3C specifications. Such tools include W3C HTML Validation Service [W3C 2000], W3C CSS Validator [W3C 1998], and Doctor HTML [Tongue & Inc. 1997].

The W3C HTML Validation service checks HTML documents for compliance with W3C HTML Recommendations and other HTML standards. It also provides several reporting characteristics that can be specified including showing the parsed DOM tree and the outline of the current page.

Doctor HTML is a commercially available tool which can be used to analyse either a single Web page or an entire Website. It provides a variety of tests including a document structure test which looks for unclosed HTML tags, and image syntax which checks image commands such as height, width, and 'alt' tags.

The W3C CSS Validator and the CSS Accessibility Analyser check the validity of CSS that a Web page uses against the W3C's specifications. They also test colour contrast and to ensure that relevant sized are specified in relative units of measurement.

These tools check for the proper use of HTML and CSS, by identifying incorrect coding and then pointing out possible solutions. This is helpful for designers and developers who may understand coding in HTML and CSS but who are not familiar with relevant techniques for accessibility.

2.3.2 User Testing

Another successful approach when evaluating a Web page is user testing. User testing is a family of methods for evaluating a user interface by collecting data from people actually using the system or the interface. This improves accessibility because human judgement is merged into the evaluation. Indeed, user testing suggests that it is not possible to determine a site's accessibility using the software alone. For example, an automated validation tool can check for the existence of ALT tags but not for the quality of the text within the ALT tag; a user is the best tool to evaluate that.

This type of evaluation with special user groups can incur greater time and monetary costs due to special arrangements for testing [Mankoff *et al.* 2005] but can bring significant results, by finding specific and detailed problems. However, user testing is demanding and hard to achieve due to financial resources, time consumed and requirements for testing, such as user interface expertise.

A possible alternative is using Web developers to perform the user testing. As Asakawa [2005] suggests, by learning why the checkpoints are defined in the evaluation process, such as understanding the use of 'alt' tag, Web designers can develop a more accessible page or at least follow as many guidelines as possible. Also, a developer who learns how a screen reader or other accessibility technology works would be able to understand how a visually impaired user access the Web. In this way they would be able to identify more easily any possible problems. For example, by learning a screen reader's navigation methods they will understand how their designed pages are accessed by voice browser users [Asakawa 2005].

As Nielsen [1994] described, user evaluation can occur through a variety of usability inspection methods. These can be methods where the interface is normally inspected by a single evaluator at a time, such as heuristic evaluation and cognitive walkthrough, or group methods, such as consistency inspection. These tests can offer excellent opportunities for observing how well an interface supports users' needs. However, there are issues that limit the use of those methods. For example, people with adequate user interface experience are hard to find, techniques are difficult to apply before an interface exists, and any recommendations that come at a later stage of the development are too late to implement [Jeffries *et al.* 1991].

Most of these evaluation techniques tend towards the interface developers and not just the user interface experts. Jeffries *et al.* [1991] tried to evaluate the above techniques, showing how well they work, what kinds of interface problems they are best-suited to detect, whether developers who are not user interface specialists can actually use them, and how they compare in cost/benefit terms. They found that heuristic evaluations produced the best results during their experiment, but the use of this technique is limited due to the need of skilled user interface professionals that are hard to find. Cognitive walkthroughs, as they suggest, can be used by software engineers to identify some important usability problems but heuristic evaluation and usability testing are more accurate techniques.

User testing is important during the evaluation process. A user, an expert or a developer can recognize problems that an automated tool can not. Even if the identification of the accessibility and usability problems through a user is not an end to the development process, it is a big factor in it, because user evaluation help eliminate problems and improve interfaces [Karat *et al.* 1992].

2.4 Transcoding

Web content is used on desktop computers as much as on small-screen devices and as voice output. Therefore, Web content has to be rendered to be suitable for each device. A mixture of important and less important information and visual effects make this task difficult [Asakawa & Takagi 2000]. Transcoding is a technology used to adapt computer application displays and Web content so that they can be viewed on any of the increasing number of diverse devices on the market. Working like an interpreter, the content is translated to suitable formats for various platforms, regardless of protocol, application, screen size, and language used. For example, a large full color image may be reduced with regard to size and color depth, removing unimportant portions of the content.

A variety of transcoding techniques have been proposed to solve the problem of mismatch between Web content and various small screen devices [Whang *et al.* 2001]. These techniques are classified into three categories: client-side, server-side and proxy servers. In client-side techniques, hand-held devices receive the whole content of a HTML page from a Hypertext Transfer Protocol (HTTP) server and convert the content format locally in the hand-held device. In server-side techniques, Web content is transcoded in the HTTP server and the hand-held devices receive the reduced, reformatted Web content. Instead of accessing the Web page directly, a proxy server intercepts the communication between the client device and the Web server and adapts the content before it is rendered on the client device. There are a variety of systems that use transcoding including WebAlchemist [Whang *et al.* 2001], Betsie [Education n.d.], and Web Access Gateway [Brown & Robinson 2001].

WebAlchemist [Whang *et al.* 2001] is a prototype Web transcoding system, which automatically converts a given HTML page into a sequence of equivalent HTML pages that can be properly displayed on a hand-held device. It is based on a set of HTML transcoding heuristics managed by the Transcoding Manager (TM) module that extract partial semantics from syntactic information such as the table width, font size and cascading style sheet. WebAlchemist generates readable, structure-preserving transcoded pages, which can be properly displayed on hand-held devices.

Betsie [Education n.d.] is a filter program developed by the BBC. It transcodes the BBC Website into accessible pages, mainly for visually impaired people. Their approach is divided into two functions, one for converting a Web page to a text only page with color and font size settings, and one for moving indexes to the bottom of a page.

Web Access Gateway [Brown & Robinson 2001] transcodes already existing Web pages automatically by removing images and creating a text only page. It serializes all text information by removing tags, change font size and color settings, and enlarge images.

As described above, transcoding techniques can take an HTML or XML[1] document, analyze it and provide a separate version of it. In order to understand how to transcode the document, external annotations are used. Semantic transcoding uses semantics or external annotations to help the machine to understand the document's contents so that transcoding can have higher quality [Harper & Bechhofer 2005; Nagao *et al.* 2001]

One approach that used semantic transcoding is the Dante project [Yesilada *et al.* 2004], whose main focus was to enable visually impaired people to navigate through Web pages. Their goal was to (1) analyse Web pages to identify objects that support mobility and travel; (2) discover their roles; (3) annotate them using ontologies (the WAfA ontology) in order to make their roles explicit and (4) transform pages based on the annotations to enhance the provided mobility support [Yesilada *et al.* 2004]. The Dante approach uses semantic annotation

[1]XML: Extensible Markup Language

of Web pages to provide screen readers with semantic knowledge. This approach helped to better facilitate the audio presentation of the Web page, because the annotations used make the "implicit knowledge of the role the page objects fulfil explicit" [Yesilada *et al.* 2003a; 2004; Plessers *et al.* 2005]. The notion of travel is introduced by Harper *et al.* [2001] through the TOWEL project that tried to improve the accessibility of Web pages for visually disabled and other travellers by drawing an analogy between virtual travel and travel in the physical world.

The SADIe project [Bechhofer *et al.* 2006; Lunn *et al.* 2008] aims to improve access to Web content for visually impaired users by transcoding the page into a format more suited to the sequential audio stream generated by a screen reader. The SADIe platform[2] provides a solution that combines the benets of heuristic and semantic transcoding to oer an accurate yet highly scalable transcoding solution. SADIe accomplishes this through three operations: *Defluff*: Removes elements that provide little or no information to the page; *Reorder*: Rearranges the page so that important areas of content appear near the top of the page; *Menu*: Displays the menu of the Website at the bottom of the page where it can be easily found but does not interfere with the main content of the page.

We believe that understanding how both sighted and visually impaired users interact with the Web provides implicit information that should be used by such approaches for more accessible Websites.

2.5 Web Page Presentation

Most of the Web pages focus on good visual presentation to implicitly help users navigate, understand, and interact with the content. Studies [Hoffos & Sharpless 1991; Faraday & Sutcliffe 1998] showed that sharp layout, snappy captions and interesting pictures can transform a wall of dry text into a presentation which users will approach enthusiastically. Researchers want to make designers aware of the power of the visual image but without downplaying the importance of the text. Other studies expound the value of the combination of text and image [Marks 1994; Fischer 1996; Faraday & Sutcliffe 1998] and approaches like the use of 'contact points', or co-references between text and images [Faraday & Sutcliffe 1998] try to effectively design a combination of text and images. Contact points

[2] An experimental prototype is available at: http://hcw.cs.manchester.ac.uk/experiments/sadie/

are places in the text where the contents need to be related with the image, such as getting details of an object's appearance, or how to perform an action. This is helpful for sighted viewers. What happens, though, when visually impaired people try to interact with such resources? Problems arise because the implicit visual cues presented on a Web page cannot be accessed and used by impaired people, especially visually impaired [Plessers *et al.* 2005]. This is why within the literature, visual complexity of an HTML document is described through Website accessibility and usability frameworks.

This section introduces the concept of visual presentation of Web page design. We try to explain how authors relate design and aesthetics and how and why users are affected from the appearance of the page with respect to the aesthetics of the page.

2.5.1 Web Page Visual Aesthetics and Design

The term aesthetics originates from the Greek which is "the science of how things are known via the senses" pointing to perception [Zettl 1999; Hoffmann & Krauss 2004]. With the use of elements of visual aesthetics the perceptions of the viewer are manipulated. On a Web page design, these elements can be colours, text style and size, pictures and animations. They induce the user to "unknowingly, unconsciously, and unsuspectingly choose to become involved in the message and the Website of concern" [Krauss 2004; Hoffmann & Krauss 2004]. This is achieved by involving the user in the communication process using elements of visual aesthetics in concert to support the intended message.

In the visual aesthetics survey of Hoffmann & Krauss [2004], the authors point to studies that try to determine Web user's perceptions of the aesthetic qualities of Websites. These studies showed that visual clarity and visual richness are two of the most important aesthetic dimensions. Visual clarity refers to clean, clear, and symmetric designs and visual richness to creativity and originality of the Websites's aspects. The above, as they explain, strongly influence the 'clients' of a Website.

A Web page design type is determined by the arrangement of elements or details. It is the "creative art of executing aesthetic or functional design" [Merriam-Webster 2006]. It is clear from earlier sections of this literature review that Website designs can either facilitate or impede the users from the Web resources that are available.

Ivory & Megraw [2005] presented a study of Website design patterns from 2000 to 2003 by examining questions like "what are the design characteristics of the most effective sites?" and "have these characteristics changed over time?". They tried to differentiate each period with respect to the design elements and techniques they used, such as links, graphics, text formatting, colours, and stylesheets. They found that Website designs became increasingly graphical in nature, reliant on browser scripts and less consistent. The significant design changes that they found was caused by text and link formatting, graphics, tables, style sheets, scripts, and HTML coding. For example, designers use larger font sizes and hyperlinks without lines underneath them making the users not to notice them. The use of graphics doubled within three years, using organizational and ornamental graphics such as bullets, form buttons and icons.

Additional studies exist that try to identify Web page design metrics that predict whether a site is highly rated with respect to complexity [Ivory *et al.* 2000; 2001; Germonprez & Zigurs 2003]. These studies relate Website design guidelines with complexity explaining that the way a Website is presented depends on the way the page itself is designed and what elements (metrics) are used.

Ivory *et al.* [2000; 2001] performed a quantitative analysis on the Web page attributes (e.g. number of fonts, images, and words) using a large collection of Web sites. The Web sites were judged based on content, structure and navigation, visual design, functionality, interactivity, and overall experience and were selected from categories such as financial, educational, community, health, service and living. They found that "Web pages and sites differ from each other on many dimensions, such as layout quality, screen coverage and information quality" [Ivory *et al.* 2000]. Also, in the same study, metrics concerning page composition, such as word and link count, page formatting, such as emphasized text and text positioning and overall page characteristics, such as page size, help to predict (with 63% accuracy) whether a page can be assigned with high or low rating from human judges with respect to usability. Using the first study and the same characteristics of the page-level elements, in [Ivory *et al.* 2001], they developed profiles of Web pages that distinguished between good and bad pages with respect to design and usability. The above studies [Ivory *et al.* 2000; 2001; Ivory & Megraw 2005] were part of a project that tries to develop techniques to empirically investigate all aspects of Website design, something that helps on identifying the components of a visually complex page.

Germonprez & Zigurs [2003] proposed three major dimensions of factors and their associated factors that impact the complexity of a Website. These are cognition, content and form and as they show in the study, these three factors and their association affect how individuals “perceive a Website, the content that is located at the site and the manner in which the Website is constructed”. That is, human cognition affects how a user retrieves and uses information in a Website. Content of the Website and the amount of information that is available affects complexity since it can cause information overload on a Website. The form of the Website with respect to user interface, navigation, and structure is the third factor that affects Website complexity.

The above studies are the closest in identifying the visual complexity of a Web page. Our empirical study [Michailidou 2005] using the above characteristics and results focused on Web page design and the structure of the page that identify visually simple and complex designs. A framework was developed with the help of the Web page’s elements such as paragraphs, tables, lists, and menus, helped to identify the main structural factors that cause visual complexity. It was suggested that visual complexity of Web pages depends on the presentation of the page’s elements and by the density and diversity of the elements that are presented, where diversity is the variety of factors and density the amount of each element.

An important difference is that during the above studies, participants or judges were using Web sites instead of Web pages used by Michailidou [2005]. A Website is a collection of one or more Web pages. In Michailidou [2005] the research was focused only on analyzing the structure of the Web pages.

Hence, the studies [Ivory *et al.* 2000; 2001; Germonprez & Zigurs 2003] relate the metrics with usability and try to come up with frameworks that help Website design improve accessibility and usability and to “develop tools to help designers assess and improve the quality of their Web sites” [Ivory *et al.* 2001]. Studies like [Michailidou 2005] tried to implement those metrics for the development of a tool that would help visually impaired people better understand the design of a Web page and know what to expect when reaching a page. This extension is one of this thesis project objectives.

In Ivory & Megraw [2005], the authors determined that the number of tables that were used to control page layouts increased within three years. Having in mind the W3C regulations, increased table use hinders visually impaired users from accessing the Web page’s content due to the limitations of the screen reader.

Methods of evaluating the usability and accessibility of Web sites are described in Section 2.4. However, it is important to relate the Web page's structure and visual presentation with usability and accessibility issues. Visually impaired people cannot distinguish whether a Website uses colours, bigger or smaller font sizes, animations or specific layouts. What they can tell is whether they can access and read the content of the page easier. However, as we explain later in this chapter, visually impaired user interaction can be interpreted as similar to sighted user browsing behaviour.

Other than following the regulations and guidelines for more usable and accessible Web pages, designers and developers have to understand how visually impaired people interact with the Web. This will help them determine the elements of the page that are more usable and less accessible. Another approach is to understand how sighted users perceive the visual presentation of a Web page and how they read and access the Web. Interpreting how sighted users rate a Web page with respect to its visual complexity can give feedback for a better design or can provide supportive information to visually impaired users. In order to achieve this, technologies such as eye tracking, provide the opportunity to observe and understand sighted user Web behaviour.

2.6 Eye Movement

Eye movement tracking methods can reveal how people read and perceive a visually complex Web page. Using eye movements as a user-to-computer communication medium [Jacob 1995] can help in identifying and attempting to solve accessibility problems.

Eye trackers initially supported research in human visual data acquisition, but as technology continued to evolve, led to applications where understanding of human perception, attention, search, tracking and decision making are of great importance. These applications include medical image analysis [Reuter & Schenck 1985], advertising [Lohse & Johnson 1996; Russo & Leclerc 1994], analysis of the performance of aeroplane pilots [Sanders *et al.* 1979], visual communication and control devices [Jacob 1991; 1995; Freedman 1984].

This section focus on eye movement by describing the relevant characteristics of the human eye, existing eye tracking technology and studies that examine and analyse eye movement.

2.6.1 Physiology of Eye Movements

An eye consists of the pupil, iris, cornea and sclera [Kolb *et al.* 2005]. The pupil allows light to enter the eye and the iris controls the size of the pupil so that more or less light is allowed to enter the eye. The cornea covers both the pupil and the iris allowing the production of a sharp image at the retinal level and the sclera, the "white of the eye", forms part of the supporting wall of the eyeball and is continuous with the cornea.

As we look straight ahead, the visual field can be divided into three regions: fovea, parafovea, and peripheral [Rayner 1998]. The fovea, located near the center of the retina, is densely covered with receptors which helps for a higher acuity vision, by covering 2° of the vision. Outside the fovea, within the parafovea, the acuity is not as good as it extends to 5° on either side of fixation and it is even poorer in the peripheral area (the region beyond the parafovea). One can not see an object clearly using just peripheral vision. Hence, to adequately see an object, one must move the eyeball to make the specific object appear directly on the fovea [Jacob 1995; Rayner 1998].

While reading, looking at a scene or searching for an object, the eye does not generally move smoothly over the visual field. It makes continuous movements called *saccades* and between the saccades, our eyes remain relatively still during *fixations* for about 200-300 ms [Jacob 1995; Rayner 1998].

Saccades are quick jumps of the eye from area to area with velocities as high as 500° per second. They are used to orient the eyeball to cause the desired portion of the visual scene to fall upon the fovea [Jacob 1995; Rayner 1998; Josephson & Holmes 2002]. The saccade is ballistic, which means that once it begins, it takes a destination/path that is not possible to change. Hence, the destination is selected before the movement begins, usually through the peripheral vision since most of the destination lies outside the fovea. Each saccade takes 30-120 milliseconds and usually covers 15° to 20° but can cover from 1° up to 40°.

Fixations follow saccades and are the periods that the eye is relatively immobile. Hence, fixations indicate where the eye pays more attention distinguishing the object that is viewed at the moment[Rayner 1998; Jacob 1995]. Although it determines where the eye pays more attention, the eye does not remain completely still but has small motions within a 1° radius, called nystagmus. The fixations usually last between 200 and 600 milliseconds, followed by another saccade.

2.6.2 Eye Tracking Technologies

The capability to track the eye gaze has been available for several decades. Early techniques for doing this were based on scleral search coils (which are still used today for certain applications). This technique involves applying a coil of wire to the eye and in the case of human subjects, a contact lens or scleral annulus, in which the coil is embedded, is acutely placed in a locally anaesthetised eye for the length of an experiment, typically no longer than 45 minutes Paul J Murphy *et al.* [2001]. In recent years, head-mounted and remote camera-based systems have been developed to allow more natural and less cumbersome methods of gaze tracking. Until now, video-based solutions have either required the use of helmet-mounted equipment or have struggled to deal with head-movement. The most recent revolutions in eye tracking technology (such as Tobii eye trackers [Tobii 2006]) provide very high quality tracking without interfering with the user environment of the test subject and are known as video based tracking. All types of eye tracker provide recording facilities for analysing the data. Infrared light reflects off the eye and the infrared image is recorded by video camera with a filter. Image processing software calculates locations of pupil(s) and reflection off the cornea.

Eye tracking provides a non-intrusive, on-line measure of allocation of visual attention. They are relatively expensive but useful for studying a wide variety of questions in various areas of research. It is important to note that the best choice of eye tracking system depends on the task and research area that one is planning to study Duchowski [2002].

2.6.3 Eye Tracking Studies

Eye movements are fundamental to the operation of the visual system. Research around eye movement tries to provide insight into the cognitive processes with the brain, neuroscience, psychiatry, psychology, ergonomics, advertising and design. Studies on design and Web page visual perception are limited but lately increasing. A detailed literature review of eye movement and application areas is written by Rayner [1998]. The author examines studies of eye movements in reading and other information processing tasks such as music reading, typing, visual search, scene perception, mathematics, and problem solving. The author proves through extensive literature that eye movement data are very informative in revealing moment-to-moment processing activities such as those mentioned above. Since our interest is to help visually impaired people understand faster and easier the Web page design we will limit our eye movement literature to studies about visual perception, and scene perception.

2.6.3.1 Scene Perception

Studies in scene perception show that observers understand and comprehend the visual information of a scene within 100 milliseconds (ms) but they need a few more milliseconds to consolidate it into their memory [Potter 1976]. For example, by watching television and flipping rapidly through the channels one can grasp each channel's meaning with a single glimpse of each picture (channel). This is called the "gist of a scene" [Friedman 1979; Oliva 2005].

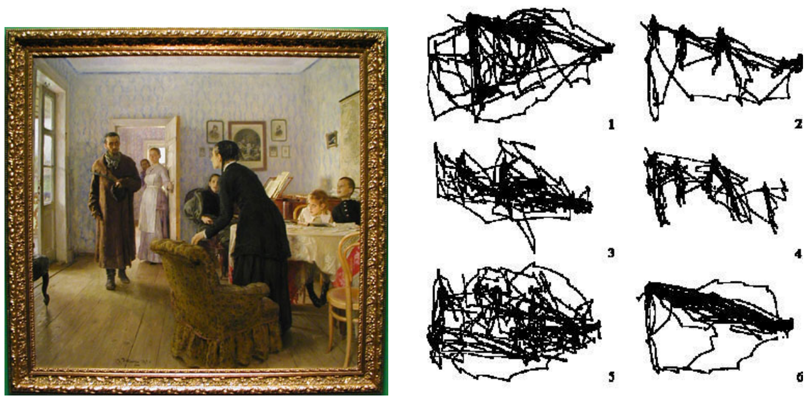
As [Oliva 2005] describes, there are two visual gists: the perceptual and conceptual gist. The perceptual gist is when the brain determines the image properties that provide a structural representation of scene, like the color, texture and volume. The conceptual gist includes the semantic information of the scene, which is improved after the perceptual information.

Knowledge about an object's features can also influence eye movement. Researchers suggest that eye movement is goal directed [Hamker 2003; Cater & Chalmers 2003]. According to the model of Hamker [2003], the author shows that visual perception controls eye movement. Viewers first see an object, comprehend what they see and then move their eyes. The nerves that supply the muscles that move each eyeball are responsible for planning an eye movement because these nerves cause the visual perception. Grouping information is another factor that determines the scanpath of the eye. People tend to construct models from our daily visual experience. These models are developed from things that we have seen before and help recognize an object because we associate them with those memories/models. Since humans have the ability to organize perceptual input, the development of the models is supported with grouping information by distinguishing different parts of an image that are uniform in color or uniform in texture [Buhmann *et al.* 1999].

Scene perception influence can come from knowledge, memories, beliefs or goals that viewers may bring to the image [Richardson & Spivey 2004]. This type of visual perception is also known as top-down [Yarbus 1967; Richardson &

Spivey 2004; Cater & Chalmers 2003]. During this process the brain directs the eye to focus on one or more objects in a scene that are relevant to an observer's given goal [Cater & Chalmers 2003], such as looking for a street sign. Yarbus was one of the first to study how the eye moves when looking at complex images and to note that the sequence of saccades that an observer follows depends on the task or questions that had been asked about the scene. During his experiment, he presented to various subjects the Ilya Repin's painting "An Unexpected Visitor" (Figure 2.1 (a)) and asked them different questions, where each questions was then associated with one common scanpath. The questions asked during the experiment are (a) "What do you see in the picture" (scanpath 1), (b) "What are the ages of the people in the painting?" (scanpath 2), (c) "What had the family been doing before the visitor arrived?" (scanpath 3), (d) "What clothes are the people wearing?" (scanpath 4), (e) "Where was each person and object in the picture" (scanpath 5), (f) "How long had the visitor been away?" (scanpath 6). Figure 2.1 shows the picture shown and scanpaths that were revealed for each question.

The scanpaths, as shown below in Figure 2.1 (b), differed dramatically according to each question. As Fahle [2004] describes, perceptual learning is any relatively permanent change of perception as a result of experience. Visual learning usually leads to fast improvements of performance in perceptual tasks. "The improvement is often very specific for the exact task trained, the precise stimulus orientation, the stimulus position in the visual field, and the eye used during training" [Fahle 2004].



(a) An Unexpected Visitor *(b) Observer's Scanpaths*

Figure 2.1: Repin's Picture and Scanpath

Bottom-up visual processing is another type of visual attention. Bottom-up models suggest that people spend the greatest amount of time focusing on salient parts of an image on the grounds that they are most interesting. That is, a viewer does not have a task goal in mind.

2.6.3.2 Visual Complexity

Scenes are composed of numerous objects, textures and colours which are arranged in a variety of spatial layouts. A variety of studies tried to study visual complexity and understand how a cognitive system represents the level of complexity of a scene [Oliva *et al.* 2004; Rayner 1998; McConkie & Currie 1996; Heaps & Handel 1999; Heylighen 1997].

Heylighen [1997] describes the perception of complexity with respect to the variety in the stimulus, by saying that the perceived visual complexity can increase as a function of the quantity and range of objects or as a function of the variety of materials and surface styles while the number of objects and surfaces remain the same.

Heaps & Handel [1999] defined complexity as "the degree of difficulty in providing a verbal description of an image" (also in Oliva *et al.* [2004]), and showed that textures with repetitive and uniform oriented patterns were less complex than disorganized patterns. A visual pattern is also described as complex if its parts are difficult to identify and separate from each other [Oliva *et al.* 2004].

Similarly with the scene perception, the complexity perception of an image depends on the amount of grouping, the quantity of the parts an observer perceives in the scene, familiarity with the scene and existing knowledge of objects inside the scene. In Oliva *et al.* [2004] the authors conclude that visual complexity is mainly represented by the perceptual dimensions of quantity of objects, clutter, openness, symmetry, organization, and variety of colours.

2.7 Visual Complexity and Eye-tracking

Eye movements are driven both by properties of the visual world and processes in a person's mind. Since eye movements can be tracked by modern technology with great speed and precision, they can now be used as a powerful input device, and have many practical applications in human-computer interactions [Richardson & Spivey 2004]. Combining the results from the presented literature survey of eye movements, scene perception and visual complexity we can say that eye movement tracking can be an invaluable tool not only for psychologists but for the area of Web page design.

As we described in the previous section, studies exist that try to reveal the Web page design composition element that can predict whether a site is highly rated with respect to complexity [Ivory *et al.* 2000; 2001; Germonprez & Zigurs 2003; Michailidou 2005]. Additional studies, though, exist using eye-tracking technologies to try to analyse the user behaviour in Web search or to reveal possible problems in Web design [Granka *et al.* 2004; Pan *et al.* 2004; McCarthy *et al.* 2003; Outing & Ruel 2006; Goldberg *et al.* 2002; Josephson & Holmes 2002; Jay *et al.* 2006].

Trying to answer questions such as 'how do users interact with a list of ranked results of WWW search engines?', Granka *et al.* [2004] performed an eye-tracking study to analyze the amount of time that Web users spent viewing the results of a Web search, as well as how thoroughly searchers evaluate their results set. They showed that the user focuses on the first two links of the results and as soon as the user has to scroll, the ranking of the links becomes less of an influence for attention. In addition to selecting links, scrolling was a factor in reading the abstracts that were provided on the search list.

McCarthy *et al.* [2003] tried to test the best position for the navigation menu on Web page by performing an eye-tracking study. They tested whether placing the menu in an unexpected position had negative impact on search performance. Analysing their eye-tracking data showed that Web users quickly adapted to an unexpected screen layout. However, it was obvious, during their study, that most of the users expected the menu to be on the left hand side of the screen.

Jay *et al.* [2006] wanted to examine "how the presentation of Web pages on standard displays makes them easier for sighted people to use" by performing an eye tracking study. They investigated how sighted readers used the presentation of the BBC News Web page to search for a link by comparing the standard presentation with the BBC's text-only version. They concluded that presentation of information assisted task completion and users spent more time to complete a task in a text-only version of a Web page.

Eye movement behaviour involves different levels of cognitive processes, including oculomotor and semantic processes [Henderson *et al.* 1999]. Pan *et al.* [2004] tried to determine the ocular behavior on a single Web page: whether it is determined by individual differences of the subjects, different types of Web sites, or the task that they had to perform. The results of the study showed that Web page viewing behaviour is driven by: 1) the gender of subjects, 2) the order of Web pages being viewed and 3) the interaction between site types and order of the pages being viewed. The authors concluded that the results of the scanpath analysis revealed a possible relationship between scanpath variability among individuals and the structural/visual complexity of the Web page.

Eye-movement measures were taken in Josephson & Holmes [2002] while subjects repeatedly viewed three different kinds of Internet pages: a portal page[3], an advertising page and a news story page. The resulting scanpaths were compared using a methodology (string-edit) that measures resemblance between sequences. They found that on the Web pages with complex visual digital images, some viewers' eye movements followed a habitually preferred path across the visual display. In this study, some scanpaths were strongly similar among different viewers which could indicate that factors such as features of the Web site or memory are important.

Goldberg *et al.* [2002], tested navigation among portlets, when at least two columns exist. Their eye-tracking results showed that navigation was biased towards horizontal search (across columns) as opposed to vertical search (within columns). Within a portlet, the header bar was not reliably visited prior to the portlet's body, evidence that header bars are not reliably used for navigation cues. They suggest that critical portlets should be placed on the left and top of the Web portal area, and that related portlets do not need to appear in the same column.

In Eyetrack III [Outing & Ruel 2006] the authors observed people for one hour as their eyes followed mock news Websites and real multimedia content. They extensively tested eye movements across several news homepage designs and noticed a common patten: The eyes most often fixated first in the upper left of the page, then hovered in that area before going left to right. Only after pursuing the top portion of the page for some time did their eyes explore further down the page. They also noticed that with news homepages, readers' instincts were to first look at the logo and top headlines in the upper left. Another important conclusion during their study was how people read headlines. They noticed that

[3]A web portal presents information from diverse sources in a unified way and offer services such as e-mail, news, stock prices, information, and entertainment. MSN, Yahoo! and iGoogle are such examples.

when people looked under headlines on news homepages, they often only looked at the left one-third of the blurb. In other words, most people just looked at the first couple of words, and only read on if they are engaged by those words.

2.8 Discussion

As we mentioned earlier, in addition to following regulations and guidelines for more usable and accessible Web pages, designers and developers have to understand how visually impaired people interact with the Web. This will help them determine the page elements that are more usable and less accessible [Yesilada *et al.* 2007]. Takagi et al Takagi *et al.* [2007], during their studies, showed that blind participants adhered to their familiar scanning navigation methods when they faced accessibility problems . This navigation involved gambling or exhaustive scanning by using landmarks in order to find the boundaries of the content area of a Web page and therefore build mental models of the page (see discussion on page 20). Yesilada *et al.* [2008] presented an eye tracking study that showed a strong relationship between Web page areas that receive visual attention and travel objects previously identified to hinder visually impaired users accomplish same tasks as sighted users [Yesilada *et al.* 2007].

We suggest that the increased visual landmarks on a page can impede visually impaired user navigation by increasing the time and effort to reach the information they seek. As explained in this chapter, the increased number of visual landmarks lead to a visually complex page. We believe that it is important to interpret how sighted users perceive the visual presentation of a Web page, how they rate a Web page with respect to its visual complexity and how they interact with it. Feedback can be extracted and used as guidelines for better designs, using approaches such as transcoding (see Section 2.4), or can provide supportive information to visually impaired users about the visual complexity of the page which will enhance their experience by providing an initial perception of the page through their assistive technologies (see Section 2.1).

As the authors in Takagi *et al.* [2007] explain, visually impaired users, and specifically blind users, perform hypertext navigation in a similar fashion to sighted users and also scanning navigation which corresponds to sighted users' eye-movements. Eye tracking technologies are now increasingly used in studies such as [Granka *et al.* 2004; McCarthy *et al.* 2003; Outing & Ruel 2006; Jay *et al.* 2006; Yesilada *et al.* 2008] that analyse the user behaviour in Web search or to reveal possible usability and accessibility problems.

We focus on the design and structure of the page itself to identify the characteristics that make a page layout visually complex or not, through a user's cognition. In addition, we aim to develop a framework that will help visually impaired people on better understanding the design of a Web page by giving them feedback about the visual complexity of the Web page. This framework can also be used to guide a transcoding process for the design of a more accessible Web page.

2.9 Summary

The Web has become a ubiquitous information source and communication channel. The Internet user population is constantly increasing and people try to access more and more information. However, as we described in this chapter, visually impaired users face accessibility problems. Assistive technologies exist but provide limited access; laws and guidelines are formed but not always followed; and new technologies have been develop that try to find solutions. In this chapter we described the Web accessibility research area's concerns and attempts to improve accessibility problems which enabled us to investigate and analyse further the methodology for achieving our aims through the research questions (see page 22).

We can see through the literature that eye movements can provide researches with "a rich, dynamic data source concerning the temporal dynamics and psychological processes that led up to the responses of the eye movement" [Richardson & Spivey 2004]. These properties can also be of great value to designers and engineers, as they allow for detailed measurements of how a user is interacting with a device. Even if it is mostly used for usability evaluations, one can see its application in determining specific scanpaths relative to each Web page structure. If each design is associated with a scanpath and fixation points, it can provide feedback that will be used for the better development of accessibility tools. If it is possible to give a visually impaired user some notice of the expected complexity of the interaction required before link traversal then the time wasted on unproductive audio interaction, scanning, and glancing could be reduced.

Chapter 3

Eye Tracking Experiment

As discussed in Chapter 2, eye movements are driven by properties of the visual world and processes in a person's mind [Rayner 1998]. Eye movements can now be tracked using eye tracking technologies which are used as a powerful device for many practical applications in human-computer interaction. Investigating sighted users' Web behaviour using eye movement tracking methods gives a better understanding of users' perception and cognition of page presentation. Users' glancing habits provide important information on where users look first when they reach a Web page, where they pay more attention and for how long they concentrate on specific parts of the page. Most eye tracking studies, such as [Granka *et al.* 2004; Jay *et al.* 2006; McCarthy *et al.* 2003; Outing & Ruel 2006], try to evaluate a Website with respect to its usability and accessibility level.

This chapter describes an eye tracking study where our formative study [Michailidou 2005] was extended by using eye movement analysis to understand users' behaviour with respect to page design, structure and perception. Sighted users' browsing behaviour on Web pages was investigated to determine how the page's visual clutter is related to sighted users' browsing patterns and specifically, how sighted Web users read and navigate within a Web page and how they perceive the page's visual complexity. This gave supportive feedback on previous research [Michailidou 2005] and expanded our understanding of where Web users pay more attention and how they perceive the page content and enhances our research in defining visual complexity.

3.1 Methodology

Eye tracking technologies can record and show specific page areas that users gaze on and the number of fixations an area receives. Gaze plots indicate where a user's attention falls and fixation counts the amount of attention [Rayner 1998]. During this study, participants' eye movements were tracked while looking at nine Web pages. Fixation count, gaze time and gaze plots for each Web page were analysed to investigate the relationship between browsing behaviour and page visual presentation. Data, stimuli screenshots, images and videos collected during the experiment can be found in the associated Technical Report (see page 266).

3.1.1 Study Design

Each participant was asked to browse the home page of nine Websites by just looking at them and state whether they liked each page or not. We believe that by asking them to perform serendipitous browsing (purely random, unstructured, and undirected activity [Cove & Walsh 1988]) their behaviour and attention to any parts of the page would be affected mainly by each user's cognition and the scanpaths would not be biased by any specific task (i.e. pointing to a link, reading aloud the news headline). The Web pages used are listed in Table 3.1. This study was a within-subjects design (all subjects followed the same procedure and looked at the same pages) with five variables: *Header* (H), *Main Content* (C), *Right Column* (R), *Left Column* (L), and *Footer* (F). These five areas (see Figure 3.1) can be found in most Web site layouts and analysing eye movement data based on these areas can help investigate the participants' browsing behaviour throughout the page.

3.1.2 Procedure and Participants

Each participant was sat 60 cm from the monitor and the observer explained the instructions and the measures that were going to be taken. The process started with the subject looking at an index page with a list of links that they had to follow in order to view all nine pages. The participants were asked to browse the Web page without following any link on that page and state whether they like the page. Thirty three undergraduate students, graduate students and staff within the School of Computer Science at the University of Manchester volunteered to

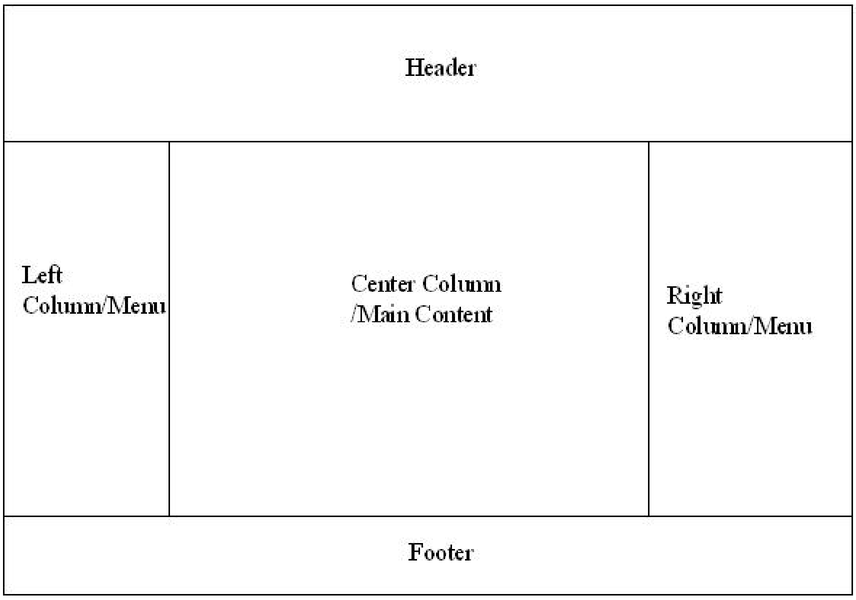


Figure 3.1: The five Web page areas used for the analysis of the eye movement data

participate in this study. The procedure was explained to each participant at the time of the study. All participants were daily Internet users with normal or corrected vision.

3.1.3 Materials

The experiment was run on a SONY VAIO VGN-FS315S laptop and stimuli were displayed on a SONY VAIO SDM-HS75P monitor. A TOBII x50 Eye Tracker, positioned at the base of the monitor, tracked the participant's gaze. The TOBII ClearView Analysis software was used to record and analyse eye movement and event data. The pages participants viewed were selected to be representative of sectors such as public information, business, academic, entertainment, leisure, Web services including search engines, and personal home pages [Amitay *et al.* 2003]. Pages were a subset of the twenty pages used during an empirical study [Michailidou 2005] previously conducted, where a preselection process took place in order for the set to be representative of all types of complexity. Table 3.1, also, shows the complexity level that each page was identified and described with during the previous study. As shown in the table, the pages are described either as visually simple or complex.

Table 3.1: Web Pages used for the Experiment and their respective Visual Complexity Level and Score [18]

Web Page Name	**URL**	**Visual Complexity Level**	**Visual Complexity Score (*seconds*)**
John Rylands Library Catalogue	http://rylibweb.man.ac.uk/catalogue/	Simple	3.75
MINT Group	http://mint.cs.man.ac.uk/index.html	Simple	3.75
Vodafone UK	http://www.vodafone.co.uk	Simple	4.92
Google Search Results	http://www.google.com/(Visual+Complexity)	Simple	6.17
Peve Group	http://www.cs.man.ac.uk/peve	Simple	8.17
IMG Group	http://img.cs.man.ac.uk/	Complex	12.67
Gene Ontology	http://www.geneontology.org/	Complex	14.67
Computer Science Manchester	http://www.cs.manchester.ac.uk/	Complex	14.75
BBC UK	http://www.bbc.co.uk/	Complex	16.25

3.2 Results

The primary aim of the experiment was to understand how visual attention is affected by a page's layout and its correlation with the visual complexity level. Questions such as 'where does the user look first' and 'which areas does the user pay more attention to' are going to be answered by analysing the eye-tracking data.

The ClearView software provides gaze plots, fixations and scanpaths over the Web page that will help identify users' common behaviour. Gaze plots display fixations, gaze positions, and scan-paths superimposed over the Web page and show how an individual's visual attention was allocated during a task. Hotspot visualizations summarize the gaze positions from multiple participants and create a coloured map over a Web page which indicates the areas that received the most fixations. The parts of the Web page that receive the most attention constitute areas of interest (AOI). The position, type (image, title, paragraph etc), order of fixation, frequency of fixation, mean fixation time and total dwell time of AOIs on a Web page will be identified and compared across tasks. In addition, fixation counts, times to first fixation and gaze times on each of the five areas shown in Figure 3.1 were used for our data analysis in order to examine where people look, which areas they focus more and which order their gaze pattern follows.

A detailed hotspot and gaze plot analysis of each Web page that was used during the experiment, including the page's elements that were viewed first, the most, least or not at all can be found in Appendix B. The rest of the section describes the most important qualitative and quantitative results derived from the experiment data analysis. Terminology of various statistical terms used throughout the analysis can be found in Appendix A.

3.2.1 Gaze Time and Fixation Count

Table 3.2 displays the mean gaze time and mean fixation count for each Web page and the percentages with respect to the total average gaze time and fixation count respectively. Gaze time represents the total time, in milliseconds, of all fixations in each Web page. Fixation count represents the number of fixations in each page. The fixation count is more useful when it is related to the areas of interest of each page but it can support the gaze time results.

The table lists the Web pages in ascending order of the average gaze time.

Table 3.2: Gaze Time and Fixation Count for each Web Page

Web Page Name	Average Gaze Time (seconds)	Average Gaze Time (%)	Average Fixation Count (units)	Average Fixation Count (%)
John Rylands Library Catalogue	6.0	6.4	888	7.1
Peve Group	8.9	9.6	1217	9.8
MINT Group	9.1	9.8	1251	10.0
Google Results	9.3	9.9	1200	9.6
Computer Science Manchester	10.1	10.8	1378	11.0
BBC UK	10.3	11.1	1363	10.9
Vodafone UK	11.3	12.1	1488	11.9
IMG Group	11.4	12.2	1458	11.7
Gene Ontology	12.5	13.4	1685	13.5

The least gaze time occurred on the Library Catalogue page with an average of 6 seconds. The longest gaze time occurred on the Gene Ontology page with 12.5 seconds. It is important to note that, as Table 3.1 shows, the Library Catalogue page is considered with low level visual complexity where Gene Ontology with high level. The Web pages Peve, Mint and Google Results were described as visually simple and had less than 10% of the total gaze time each, which supports our initial conclusions that gaze time increases as the complexity level of the page increases as well. The fixation counts support the observations made using gaze time data with the exceptions of Computer Science and Mint Web pages that had more fixation counts but less gaze time than the BBC and Google Result pages respectively.

3.2.2 Five Web page Areas

The ClearView software assisted in defining the five variables/areas for each Web page using the Areas of Interest (AOI) tool. Table 3.3 lists the five areas that each page was divided into and their respective average fixation and gaze data. For each area of the page the proportion of times a given participant looked at it for the first time was calculated. The time to first fixation indicates the order of fixation and hence provides evidence for the scanpath that the participants followed while browsing on each Web page.

Fixating on a particular part of the page means that you pay more attention or you are looking for something, where as glancing does not necessarily mean that you process what you look at. From the same table we notice that fewer fixation counts as well as less gaze time occurred on the footer areas. This is why the designers also use small font sizes for the footer area of the Web page.

A significant difference was determined between these five areas[1], with F(2.974, 68.409) = 103.53, p<.001. *Post Hoc* pairwise comparison analysis, using Sidak adjustment, revealed that Header and Main Content areas were significantly different from each other and from the other areas (p<.004). Table 3.3 shows that the Main Content was the area visited most for the first time followed by the Header area. This suggests that users expect to find the most important information in the main content area or the appearance of salient elements (such as eyecatching logos, pictures, and flashing images) attracts their attention first.

Table 3.3: Areas of Attention - Mean values

Areas	**First Fixation (*0-1*)**	**Gaze Time (*seconds*)**	**Fixation Count**	**Gaze Order**
Header (H)	0.61	2.5	10.4	1.85
Main Content (C)	0.91	5.6	13.6	1.55
Right Column (R)	0.12	1.3	9.5	2.83
Left Column (L)	0.09	2.3	8	2.87
Footer (F)	0	0.1	0.4	3.77

A difference was determined[2] between the average gaze time spent on each of the five areas for each participant with F(2.218,70.975) = 88.889, p<.0001. *Post Hoc* analysis, using Sidak adjustment, revealed a significant difference between almost all five areas (p<.001). A statistical significance between the Header and Right Column areas was not determined, probably due to the fact that not all pages had any type of information on the right or left side of the page. Based on the gaze time, Table 3.3 lists a possible attention order in which the Main Content is looked at for the longest time, which reveals the attention that this area receives.

Analysis on the number of times that each participant fixated on each of the five areas revealed the same results as the average gaze time analysis. ANOVA test showed a significant difference between the fixations spent on each region with F(3.431, 109.798) = 88.557, p<.001, using Huynh-Feldt adjustment. *Post Hoc* analysis determined that almost all areas were significantly noticed ($p \leq .02$) and as Table 3.3 shows as well that the Main Content area received the most

[1]By running repeated measures ANOVA on *arcsine* transformation of the proportions, using Huynh-Feldt adj.

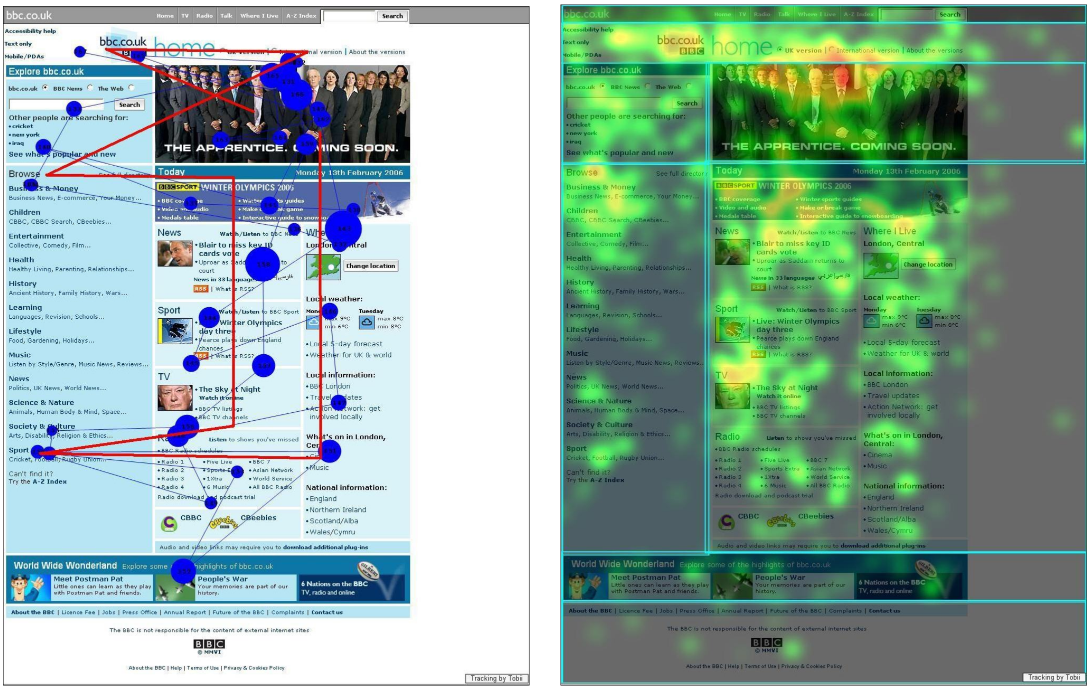
[2]By running repeated measures ANOVA on the average gaze time, using Huynh-Feldt adj.

fixations. However, the fixation counts between the Header and Left Column areas and between the Right Column and Left Column areas were not significantly different, which as explained before, it was expected.

By recording the length of time that each participant fixated at each area for the first time a possible gaze and fixation order was determined. ANOVA revealed a significant and robust effect of gaze time for each area with $F(4, 92) = 82.057$, $p < .001$. *Post Hoc* pairwise comparison, using Sidak adjustment, revealed a clear and expected pattern of significant differences among almost all individual areas, $p \leq .001$. We were not able to determine significant gaze order between the Main Content and Header areas or between the Right and Left Column as their values were close. Therefore, this data analysis shows that the areas Header and Main Content are looked at before the Right and Left Column but the order could not be clarified. Even if we can not make solid conclusions, the number listed on the last column of Table 3.3 can be interpreted as a possible gaze order (the smaller the number means the earlier this area was visited): Main Content, Header, Right Column, Left Column and Footer. Figure 3.2(a) shows one of the participants' gaze plot derived from looking at the BBC UK page. This gaze plot was selected to represent the common gaze sequence, as a pattern was determined between all the participants' gaze plots from each page. The blue spots (circles) connected with a line show the participant's gaze sequence and the red line (a line connecting page areas) shows the most common scanpath among all participants resulted by a combination of manual inspections of all the participants' gaze plots and the above data analysis.

3.2.3 Salient Elements

Hotspots (generated by the software) can show exactly the areas that participants fixated most as colour points are used to designate where participants have been looking. The warmer the colour is on the hotspot (the closer to red) the more fixations occurred on that area. Figures 3.2(b) and 3.3 show the hotspots of BBC UK and Vodafone UK pages that both had a large central image. The hotspots show that the most fixations occurred on these salient elements (eye-catching and instantly noticeable areas such as advertisements, big images and animations) which was common in all of the experiments' pages. This suggests that users' fixation occurs on the attractive elements of a page. This is also suggested by the manual hotspot and gazeplot analysis described in Appendix B.



(a) A Participant's Gaze Plot showing a Common Scanpath

(b) Participants' Hotspot showing the areas with most fixations

Figure 3.2: BBC UK (Visually Complex): (a) Gaze Plot and (b) Hotspot

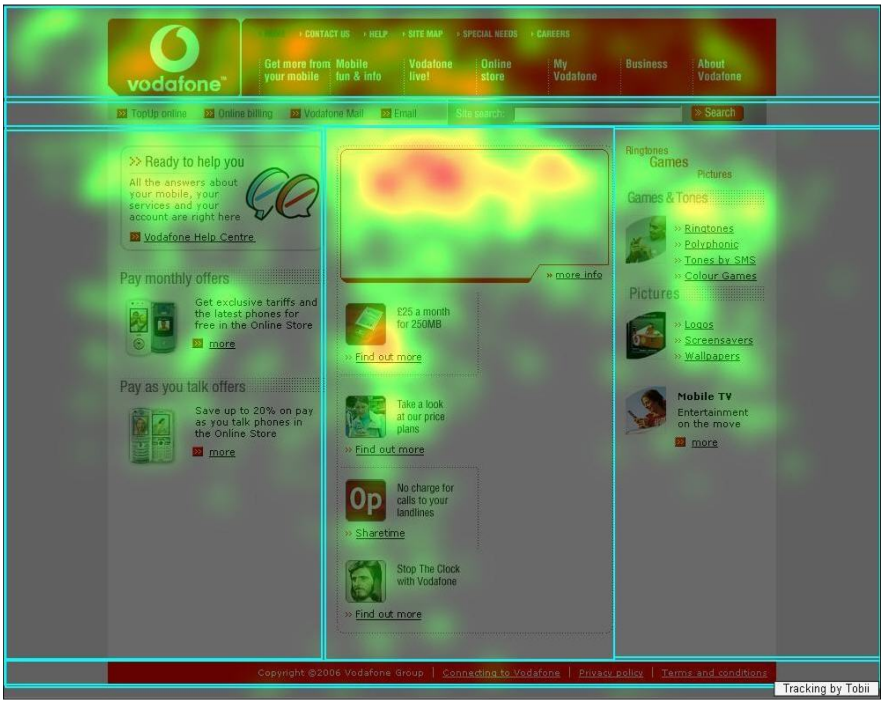


Figure 3.3: Hotspot Visualisation for Vodafone

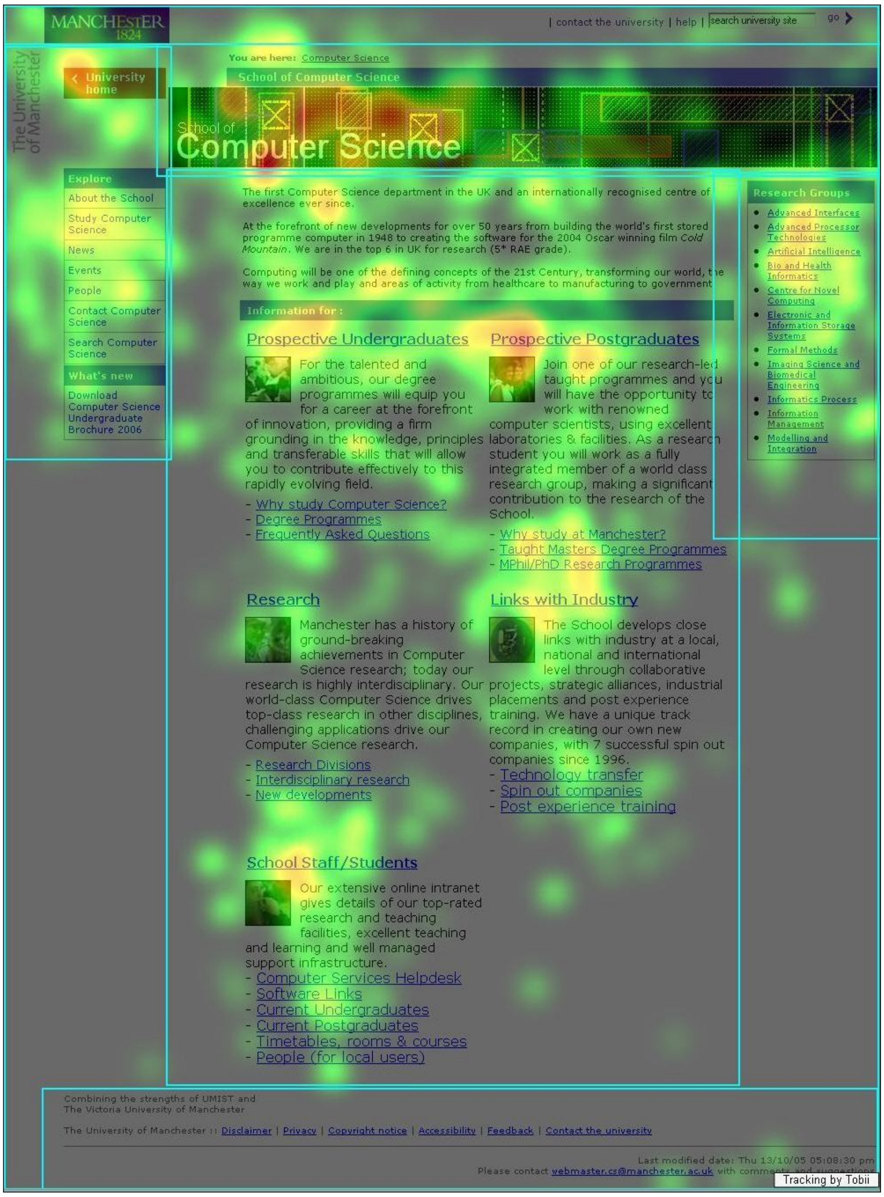


Figure 3.4: Hotspot Visualisation for Computer Science Department

3.2.4 Menus

With the help of gaze plots and hotspots a common browsing behaviour was noticed between the participants. Menus or long lists were not completely read (see Figures 3.2 and 3.4). Instead, when participants looked on a menu, they actually fixated to the first three or four items of the menu list.

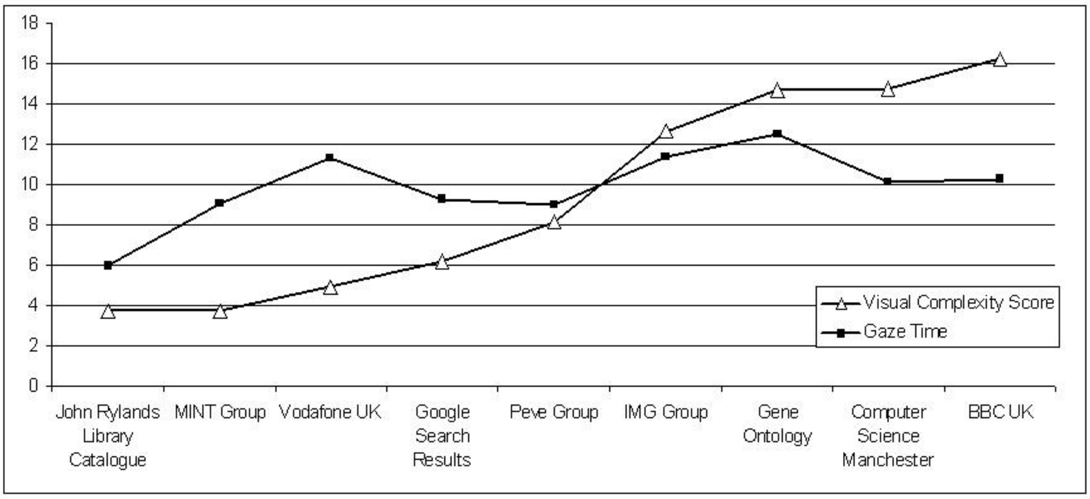


Figure 3.5: Web page Complexity Scores and Gaze Times

3.2.5 Visual Complexity

In order to understand how visual presentation affects participants' behaviour we categorized the pages used in this study in two categories: *Simple* and *Complex*. The number of graphics, images, colours, text, links, and information on each page was used to assign each page in the appropriate group using the methodology defined in [Michailidou 2005]. Table 3.1 lists the visual complexity level and average score that was determined from the methodology described in [Michailidou 2005] and Table 3.2 the attention occured on each Web page. One can notice that each of the simple pages has less than 10% of the overall gaze time and fixation count. The page with the least percentage of both average gaze time and fixation count was the Library Catalogue Web page that is ranked as simple. The page with the highest percentage of average gaze time and fixation count, 13.4% and 13.5% respectively, was the Gene Ontology Web page. The Gene Ontology page is ranked as visually complex page and having the highest average gaze time and fixation count provides support to our initial framework that describes the pages as visually complex or that the participants spent lot of time to familiarize themselves with the page or that they found something interesting that made them fixate longer. The IMG and Vodafone pages, ranked as visually complex, have also high percentages of both gaze time and fixation count. This could, again, mean that participants found the pages interesting or that page's elements (salient or not) attracted their attention.

By investigating the gaze time spent on each of the two types of pages by running a t-test it was shown that, on average, participants spent more time browsing visually complex pages (Mean=10.4, Standard Error=0.9) than visually simple pages (Mean=9.2, Standard Error=0.8, $t(32)$=-3.298, p=.002, r=.5). A correlation analysis between the visual complexity scores and the gaze time revealed a positive and significant relationship, with $r = .6$, p (one-tailed) $< .5$. Figure 3.5 shows the gaze time and complexity score for each page, in ascending order of the complexity scores. The relationship determined above between the two variables ($r = .6$) is also shown by these two lines as both have the trend to increase.

Figure 3.6 shows a participant's gaze plot on one of the visually simplest pages (Mint Group) and Figure 3.2(a) shows a gaze plot on one of the most visually complex pages (BBC UK). By manually examining all of the participant' gaze plots one-by-one and comparing to each other, we noticed that the more visually complex the page was the more scattered and disordered (without specific pattern) their scanpath was.

3.2.6 Limitations

During the experiment and the data analysis, we identified some limitations with the study's structure and were considered for the follow-up user evaluations. These are:

- Web pages with advertisements, static or flash, and lot of graphics or sponsored links changed during the experiment at least once. For example, the Manchester Online Web page (initially being one of the tested Web pages) changed at least three times. Hence, three participants looked at three different versions of the same Web page. Also, this happened with the Google Results page and the BBC home page. This did not cause any particular problems on the data analysis, because (a) the ClearView software enabled

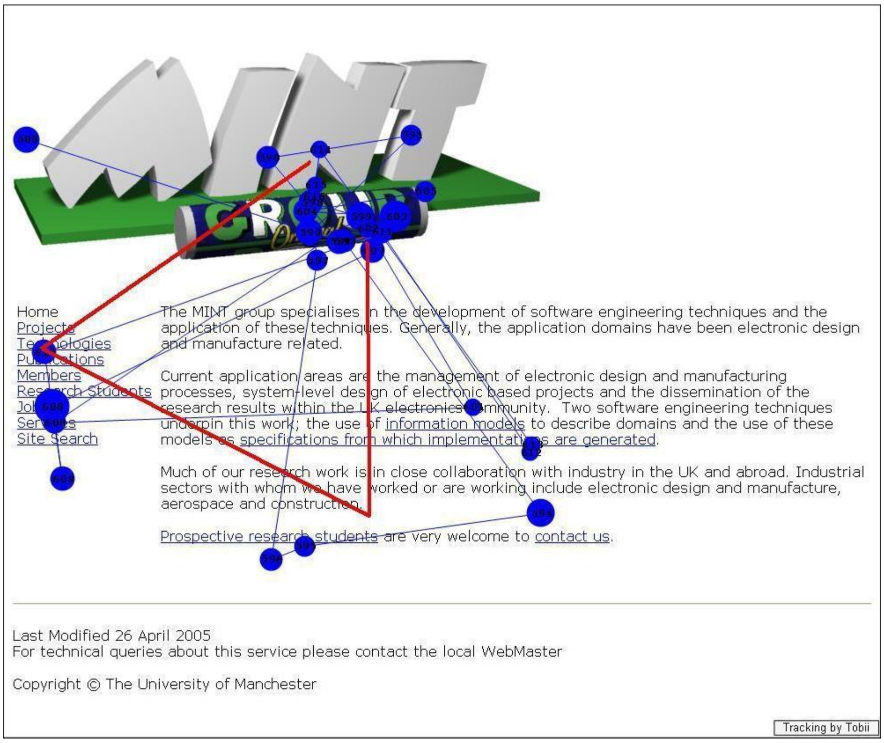


Figure 3.6: Gazeplot - MINT Group (Visually Simple)

us to analyse the data with respect to the defined AOI and (b) the structure of the Web page did not change.

- Vodafone's page kept redirecting the participants to the same page meaning that navigating back one page was not sufficient to get to our main site with the list of testing pages by clicking the Back button as directed by the evaluator. Instead we should ask them to click on the Home button which would be preset to re-direct to the index page.

- The Manchester Online page had pop-ups that we were not able to block through our browser, causing a distraction to the participants while glancing through the page. Due to the pop-ups the software failed to save all of the subject's data. Hence, we had to drop the Manchester Online Web page from our data collection.

- One of the main problems during the study was the subject's familiarity with most of the pages. The examined page were selected from research groups within the department, news pages and e-commerce pages. The participants stated that they had browsed most of these pages at least once before. Pages that they did not like at all or pages that they already knew received the least gaze time and hence less attention. In addition, some participants did not scroll through the whole pages due to their familiarity with the page, even if they liked it or not which caused a generalizability problem. The experiment's main focus was to test the time that each subject needed to form an opinion about the visual representation of the page. Since time and familiarity with the page were biased, any statistical significance derived from the data analysis should be only used as feedback for further studies.

3.3 Improving Web Accessibility

Other than understanding user's visual perception and its relation with Web page complexity, this eye tracking study gave further insight into Web accessibility for both sighted and visually impaired users. One might consider: "How can this eye tracking study help to improve visually impaired users' Web access?". Even if regulations and guidelines exist that, when followed, create accessible Web pages, visually impaired users are still hindered by the visual appearance of the Web [Yesilada *et al.* 2008; Takagi *et al.* 2007; Yesilada *et al.* 2007]. We believe that understanding how sighted users browse the Web reveals interesting paradigms which enable both sighted and visually impaired users to browse a page fast, effortlessly and efficiently. These paradigms, if appropriately used along with the existing regulations and guidelines, may improve Web accessibility. For example, with appropriate CSS and HTML important information can be more accessible for screen readers through an improved audio rendering while leaving the visual rendering of the page intact. We suggest three points that should be considered by Web page designers to benefit visually impaired users' Web experience. These suggestions focus on areas of attention, interaction patterns, menu design, and the presentation of structural and aesthetic elements.

It is important to note that these are suggestions derived from this eye tracking study and since any significance could not be generalizable due to design issues, one should not discard the interesting observations that were extracted.

3.3.1 Areas of First and Most Attention

Provide the title of the page followed by the main content without any salient elements impeding the reading order.

Table 3.3 shows that the Main Content and Header areas are looked at first for the most time and Main Content is the one looked at for the longest. Taking into consideration that Main Content and Header areas attract users' attention most, designers should deliver these sections first in an accessible way. For example, the source code of the title of the page should be provided first followed by the first subject of the main content area. The visual rendering of the page should still be the same but screen readers should be able to access the main content before any menu links and without the users having to learn any software shortcuts.

The observation that the Main Content areas attract user attention first and for the longest is shown to be affected by users' expectation to find information in the centre of the page or the appearance of salient elements (such as big images, logos and animations). The latter is usually a result of a user's peripheral vision which affects the eyes' destination and sequence [Rayner 1998]. From the gaze plots (Figures 3.2(a) and 3.6) and hotspot visualizations (Figures 3.2(b) and 3.4) provided here, we notice that salient elements were looked at first and had the most fixations. Even if this attention shows that the images might be important to provide first in the source code, they interfere with the context and visually impaired users will continue facing accessibility problems. This is due to the fact that assistive technologies, such as screen readers, go through all these images and advertisements before reading the main content of the page. Therefore, visually impaired users on some Websites can spend almost two minutes [Takagi *et al.* 2004] listening to advertisements and image descriptions before reaching the main content's heading. We suggest that the source code of big images and animations should be placed after the Main Content and Header areas of the page. Using appropriate CSS these images can be visually presented in the same location and style as before. Hence, the page will remain visually the same but visually impaired users will benefit by listening to the most important context of the Main Content first and the advertisements last. However, further research is necessary to find efficient ways of distinguishing an advertisement from an image relevant to the content.

3.3.2 Interaction Pattern

Provide the content of the page in the following order: Header, Main Content, Left Menu, Right Menu/Column, Footer.

Based on the gaze times, first fixations, and the fixation counts, providing a Web page's content in the following order can increase accessibility: (1) Header information should be limited to the title of the page, (2) Main Content must be followed with no advertisements or structural images, (3) Left Column and (4) Right Column should be then in order followed by the (5) Footer area.

Providing an audio rendering of the page in this way will decrease interaction time. Most users decide if they want to stay at a page when they find what they are looking for, which is usually in the main content area. If not, menus are then searched for links that can help them complete their task. Due to the fact that not all pages used in this study had right and left columns/menus we were not able to distinguish the order in which participants tend to visit them. However, previous studies, such as [McCarthy *et al.* 2003], show that users expect to find the menu on the left hand side of the page but they can easily adapt to any new layout as long as it is consistent throughout the site.

3.3.3 Menu Design

Decrease the length of the menu lists by providing 3±2 links.

Figures 3.2(b) and 3.4 show that lists of links, such as menus, were not thoroughly gazed at. Participants tended to mostly fixate on the first three to four links. We believe that this happened for a number of reasons such as that participants were either uninterested in the content of the page, they might have reached what they were looking for, or they already had an idea of what the rest of the menu lists contained, but mainly because they did not have a specific task in mind. Studies, such as [Outing & Ruel 2006] describe that participants do not read or look at all links in a list and tend to read the first three words in each link. Also, the authors in [Cutrell & Guan 2007] demonstrated that participants looked at the first 3 or 4 search results at a search result Web page, no matter which result they eventually clicked on.

We suggest that the menu lists should be limited to 3±2 links by providing serial access to the rest of the links [Horton 2006]. During another eye tracking study [Brown *et al.* 2009], it was demonstrated that suggestions that are further down the list are less likely to be viewed, and receive fewer and shorter fixations than those at the top. Specifically, the percentage of people viewing a suggestion was greatest if it appeared in the top three list positions, and dropped markedly after this point. Evidence also showed that the first suggestion is regarded by participants as the most important, then the second, then the third, with earlier suggestions receiving more, and longer, fixations [Brown *et al.* 2009]. We believe that smaller lists of links with the availability to accessing information in a serial manner could be useful not only for visually impaired users but for small screen devices (such as PDAs and mobiles). Providing to the user a limited number of options with the most common ones, may decrease interaction time and make the user more satisfied.

3.4 Discussion

During this study, it was shown that the more visually complex a Web page is, the more time it takes for users to gaze through it (see Tables 3.1 and 3.2) and the more disoriented their scanpath is (see Figures 3.2(a) and 3.6). We believe that this happens due to the high number of structural elements used to represent the page such as text, links, figures, subjects, animations, and input forms to name but a few. Users' gaze order can follow a stable and standard flow when they look at pages with simple layout. On the other hand, pages with a large number of different subjects, images, links and sections can disturb users' gaze flow and lead to a scattered scanpath. We assert that by removing or minimizing unnecessary visual clutter, such as advertisements, and following the reading pattern suggested above leads to a page which is easier and faster to access by visually impaired users without having to be expert users of their assistive technologies.

This study's results provide further insight on how visual presentation affects sighted users' browsing behaviour. We suggest to use the above points as guidelines to transcode a Web page into one with less visual clutter but also in a way that can be rendered for visually impaired users as sighted users tend to browse. These suggestions could be used and incorporated into a transcoding approach that is currently proven to provide Web pages that are easier to access by visually impaired users (see [Bechhofer *et al.* 2006]).

3.5 Conclusions

Eye tracking methodologies have been used to investigate cognitive processes for over 100 years [Rayner 1998] but have recently proved to be a powerful tool in understanding Web behaviour. Here we have described an eye tracking study where we investigated how sighted users perceive the visual presentation of Web pages, including on which page areas they glance first, for how long and in which order. Analysis of this study's results revealed interesting browsing patterns. For example, Header and Main Content areas are the ones that users tend to fixate for the first time and for the longest. This is due to users' common gaze pattern, expectations to find information in the centre of the page or the appearance of salient elements (such as big images, logos and animations). Other interesting observations are that menus were not completely read, no specific reading order between right and left menus or columns was determined and the footer area was the last and least gazed at.

In addition, this study provided implicit information on users' common browsing scanpaths. After gazing at dominant graphics the participants scanned through the main content of the page. They did not fully read the text in the main content area, and they paid more attention to the links and the first sentence of each block or paragraph. Then, participants tended to look at and read the first three links on the menu, either on the left or right hand side of the screen.

We were also able to notice a relationship between visual presentation and users' browsing behaviour. Participants spent more time looking at a more visually cluttered page than a simpler one. In addition, the more visual cues the page had, the more scattered and disordered the participants' scanpath was.

We showed that eye movement can help us understand how the visual presentation of Web pages influences sighted users' behaviour and perception with respect to visual complexity. Even if we could not make robust conclusions based on quantitative statistical analysis, we suggest three key points that should be used by Web page designers and software developers to improve visually impaired users' Web experience. These key points denote the areas that users tend to focus more and most, users' common browsing behaviour, menu design and location and how visual complexity affects Web browsing. These suggestions must be considered as guidelines for designing Web pages but also to transcode pages into simpler and more accessible ones.

This study acted as our initial step into understanding and defining a page's structure and navigation pattern and identifying islands of visual complexity for each page. These observations contributed towards the investigation of this project's first research question, Q1, by examining the relationship between visual complexity and Web interaction and making interesting qualitative conclusions of a possible connection. The next step from this experiment is to examine how structural elements affect users' visual perception and how we can modify them to create visually simpler pages towards answering the rest of the research questions.

Chapter 4

Towards Defining Visual Complexity

As discussed in Chapter 2, complexity can be defined as "the degree of difficulty in providing a verbal description of an image" [Heaps & Handel 1999; Oliva *et al.* 2004] and perception of an image depends on the amount of grouping a user unconsciously performs, familiarity with the scene and existing knowledge of objects inside the scene. Website visual perception is affected by cognition, content and form [Germonprez & Zigurs 2003]. Developing a framework with the help of the Web page's elements such as paragraphs, tables, lists, and menus, an empirical study [Michailidou 2005] identified a number of structural elements as factors of visual complexity. It was suggested that visual complexity of Web pages depends on the presentation of the page's elements and by the density and diversity of the elements that are presented. One of this study's objectives, as described in Chapter 1, is to examine through user evaluations how Web page presentation and explicitly the structural elements used to design a page affect users' complexity perception. This chapter describes a series of pilot user evaluations that based on the initial conclusions from the empirical study investigated this relationship. The results from these studies form the basis towards defining and understanding visual complexity of Web pages.

4.1 Web page Comparison - Pilot Evaluation I

The first attempt into understanding user Web page perception was a pairwise comparison user study. During this evaluation, ten Web pages are compared based on their visual complexity by sighted users so that a page ranking would be determined based on the visual complexity of the Web pages. The following sections describe in further detail the study design and the results revealed.

4.1.1 Methodology

During this study we used the pairwise comparison methodology during which objects are presented in pairs to one or more judges. The basic experimental unit is the comparison of two objects, A and B, by a single participant who must choose one of them. Usually, the judge is allowed to declare a tie, or asked to record a preference on some finer scale, a ranking or a score [Herbert 1988]. During this study, objects are the images and judges are the participants who had to decide which image is the most visually complex.

A balanced paired-comparison experiment occurs when all possible comparisons between the objects are made by all judges. The method of judgement used in this experiment is the 7-point preference scale (-3, -2, -1, 0, 1, 2, 3), reading "strong preference for A", "preference for A", "slight preference for A", "no preference", "slight preference for B", etc. This scale gives the opportunity to retain a close comparison between pairs of objects and create a ranking sequence.

With the ten Web page images, a complete factorial paired-comparison design would be if each participant compared all possible combinations between the ten pages, that is 45 comparisons. However, as it is described in Section 4.1.1.1 the experiment took place during visits by prospective students to the School and a limited time was available for the execution of this study. Therefore, due to this limitation, each group of participants performed 20 comparisons only.

4.1.1.1 Participants and Procedure

The study took place during visits by prospective students to the School of Computer Science. Fifty-two (52) prospective undergraduate students volunteered to participate, 48 male and 4 female. Five of the participants were aged between 26 and 45 and forty-seven aged between 16 and 25. All had normal or corrected vision and 42 participants had English as their native language. The participants reported that they use the Internet daily or a few times a week for business/work (43%), email/chat and for special interests (93%), and for online purchases (71%).

Each group of participants was sat behind a wide table placed within a meter in front of a television screen. A picture taken during one of the sessions is shown in Figure 4.1. A presentation was made by the observer who introduced the participants to the experiment and described what they had to do. Then, the participants were asked to read the information sheet provided to them and if agreed with the experiment procedure to sign the consent form. After completing a demographic data questionnaire, the observer showed them an example of what they had to do and the participants then performed the study which was completed within ten minutes. During the experiment the participants were asked to compare two images at a time and record on a 7-point preference scale which they believe is most visually complex. The participants used an answer sheet provided to them to record their answers. The scripts read for introducing the experiment and the materials used for providing consent and the answer sheet can be found in Appendix D.

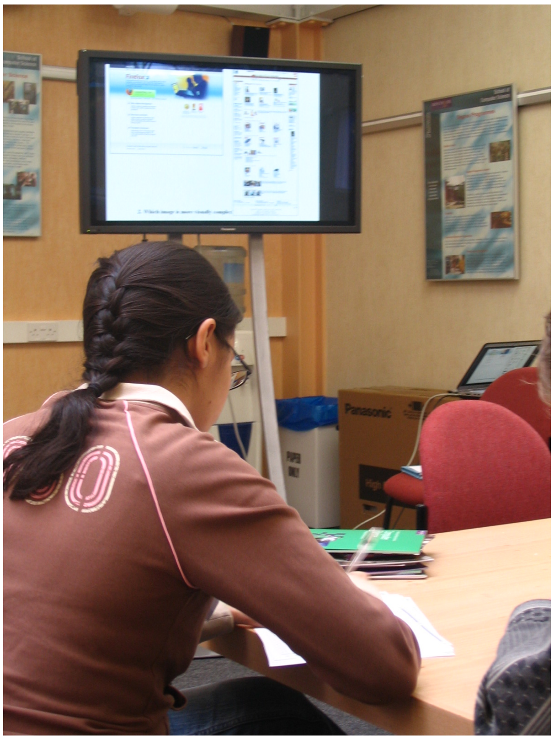


Figure 4.1: Set-up during Pilot Evaluations

4.1.1.2 Materials

Pages were similar to the pages evaluated during the eye tracking study described in Chapter 3. Screenshots of all ten images are shown in Appendix C. Table 4.1 lists the Web pages used for the experiment and their complexity level determined using the framework previously defined [Michailidou 2005; Harper *et al.* 2009]. Each page was represented by a screen shot taken at approximately the same time and date. The images were displayed on a Panasonic TH-42PHD8 42-inch plasma colour television. Each participant was provided with the information sheet, consent form, demographic data questionnaire, an answer sheet and a black pen. After the experiment, all of the equipment was returned to the observer.

Table 4.1: Pilot I - Pages used in Experiment and their Visual Complexity level

Page	**Name**	**Visual Complexity Level**
P1	Amazon	Complex
P2	Annotea	Simple
P3	BBC	Complex
P4	Firefox	Simple
P5	Flickr	Complex
P6	GoogleSearch	Simple
P7	MSN	Complex
P8	Orkut	Simple
P9	WAI	Simple
P10	Yahoo	Complex

4.1.2 Results

Due to the fact that during our evaluation participants had to make the same comparison more than once, and also gave the participants to compare various random subsets of 20 pairs, made our design complex and intricate pattern of interdependencies. Even though it was difficult to extract much information, we ran some tests that gave us important feedback. The statistical results described below are not robust but were considered as an initial and directive input for our future studies. Terminology of various statistical terms used throughout the analysis can be found in Appendix A.

Table 4.2: Pilot I - Mean Values for Comparisons Significant at ≤ .05

Pages	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0	0	0.769	-1.081	-1.452	-1.38	0	-1.455	-1.542	0
P2	0	0	1.688	1.188	0	0	0	0	0	1.409
P3	-0.769	-1.688	0	0	-1.455	0	0	-1.053	0	0
P4	1.081	-1.188	0	0	0.778	0	0	0	0	2.107
P5	1.452	0	1.455	-0.778	0	0	1.737	-0.762	0	1.421
P6	1.38	0	0	0	0	0	1.429	0	0	1.682
P7	0	0	0	0	-1.737	-1.429	0	0	0	-0.546
P8	1.455	0	1.053	0	0.762	0	0	0	0	2.053
P9	1.542	0	0	0	0	0	0	0	0	1.867
P10	0	-1.409	0	-2.107	-1.421	-1.682	0.546	-2.053	-1.867	0
Mean	0.6141	-0.4285	0.4965	-0.2778	-0.4525	-0.4491	0.3712	-0.5323	-0.3409	0.9993

To read the table: row x & column y = the mean value for pair XY

4.1.2.1 Ranking Sequence

For the data analysis, we used -3, -2, -1, 0, 1, 2, 3 for the responses. During the study, image comparisons were made for both A *vs* B and B *vs* A, where B and A are two different images. We collapsed these responses by changing signs of BA responses and adding them to AB. In the case that the same participant responded to both AB and BA the average of the two responses was taken. Tables D.1 and D.2 lists the comparison data given by each participant. One can easily notice that not all participants compared all possible combinations. This experiment design problem caused any further data analysis to be questionable. However, we can still make qualitative data analysis and run a one-sample t-test [Field 2005] from which the results can be only used as initial feedback for further studies.

The mean value of each column of Tables D.1 and D.2 show how on average participants compared their complexity. A zero mean value means that the participants overall believe that the two pages do not differ. Instead of using these mean values for creating our ranking sequence, we first test, for each column, a null hypothesis that its true mean is zero. The more a sample (observed) mean differs from zero and the more consistent the individual responses are, the more significant its difference from zero is. To compare the means and test the null hypothesis (that all images have the same level of visual complexity) we ran a one-sample t-test on all comparisons.

Table 4.2 shows the means for all the pairs that are significant at ≤ .05. That is, for every paired comparison where the t-test revealed a significance we list the mean value. Otherwise we set it to zero. Tables D.3 and D.4 list the one-sample descriptive statistics (mean values of all comparisons) and the t-test results respectively.

When we sum the columns of means (see Table 4.2), we have a complexity mean value for each Web page. The higher the number, the more complex a given site was overall judged and a zero value means that it was judged as neutral. Table 4.3 lists the Web pages used during the evaluation with their complexity mean value in ascending order. We can also see the Web pages' preassigned level of visual complexity and how they were ranked by the participants. The only outlier in this sequence was the Flicker (P5) Web page where we described it as visually complex but the participants believe that it was one of the visually simplest pages. This is because based on our initial model we identified the specific page with a high number of images and a diversity of colours that fitted in the complex range of pages, where the participants might only noticed the limited number of information. Figure 4.2 shows the Web pages in order of visual complexity and the level of complexity for each page (S-simple, C-complex), which demonstrates that except the Flickr page, our initial assumption is supported by the participants' rankings.

Table 4.3: Pilot I - Web Page Complexity Means and Ranking Scores

Web Page ID	Web Page	Visual Complexity Level	Complexity Mean	Ranking Sequence
P8	Orkut	Simple	-0.5323	1
P5	Flickr	Complex	-0.4525	2
P6	GoogleSearch	Simple	-0.4491	3
P2	Annotea	Simple	-0.4285	4
P9	WAI	Simple	-0.3409	5
P4	Firefox	Simple	-0.2778	6
P7	MSN	Complex	0.3712	7
P3	BBC	Complex	0.4965	8
P1	Amazon	Complex	0.6141	9
P10	Yahoo	Complex	0.9993	10

4.1.2.2 Discussion

Due to experimental design biases, any statistical analysis results would not be robust. However, these results provide an idea and basis for further investigation. We were able to reveal a possible relation between the preassigned complexity level and the participants' perception on visual complexity using an initial model

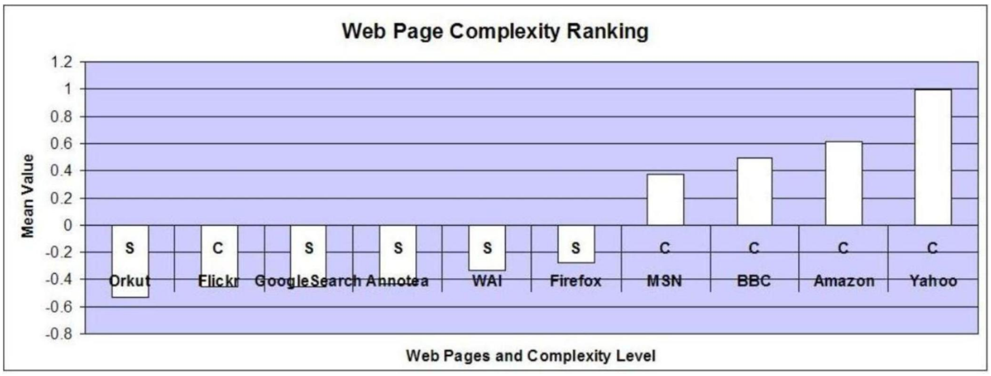


Figure 4.2: Pilot I - Ranking Order based on Complexity Mean Values

[Michailidou 2005]. In order, though, to specify the factors that affect complexity perception, further studies need to be performed to investigate the effect that the layout has on user perception.

4.2 Web page Comparison: Edge Factor - Pilot Evaluation IIa

As it was discussed in the literature, Web page designers focus on good visual presentation to implicitly help users navigate, understand, and interact with the content. With the rapid and constant advancement of technology, new ways are constantly being introduced to present information that leads to visually complex Web pages. Studies [Hoffos & Sharpless 1991; Faraday & Sutcliffe 1998] show that distinct layout, smart captions and interesting pictures can transform a wall of dry text into a presentation which users will approach enthusiastically. Complexity can be defined as "the degree of difficulty in providing a verbal description of an image" [Heaps & Handel 1999; Oliva *et al.* 2004]. Textures with repetitive and uniform oriented patterns are less complex than disorganized ones. A visual pattern is also described as complex if its parts are difficult to identify and separate from each other [Oliva *et al.* 2004]. Complexity perception of an image depends on the amount of grouping, the quantity of the parts an observer perceives in the scene, familiarity with the scene and existing knowledge of objects inside the scene. Visual complexity is mainly represented by the perceptual dimensions of quantity of objects, clutter, openness, symmetry, organization, and variety of colours [Rayner 1998; Oliva *et al.* 2004].

Therefore, we wanted to examine the effect that the amount of space that exists around different areas or the way information is grouped to make users' interaction easier has on user perception. That is, the more edges, different paragraphs and sections are used to represent a page, the more visually complex it becomes. The second pilot evaluation was designed to examine two things: the effect that the number of edges has on users' rankings on visual complexity and how users rank screenshots of Web pages with respect to their visual complexity. This section describes the first part of this second pilot study for which the same previously examined ten pages were manipulated in two ways: with less and with more space/edges. The participants had to then compare the transformed pages and decide which one is more visually complex.

4.2.1 Methodology

Using the pairwise comparison technique described in Section 4.1.1 users' perception of visual complexity was examined in relation with the number of edges used to represent the same information on a Web page. As explained later, the source code of the ten Web pages was manipulated to create two different versions of each Web page. Therefore, a set of thirty pages was collected where each Web page was represented in three versions: *Original*, *Less Edges*, *More Edges*. Due to the time limitation previously explained (Section 4.1.1), a non-balanced paired-comparison design was executed, where participants were asked to compare only the Original *vs* Less Edges and Original *vs* More Edge versions of each of the ten Web pages. That is, 20 possible comparisons had to be made by every participant in order for each original page to be compared with both modified versions of the same Web page. In addition, in order to avoid design problems identified during the previous study, all participants made the same comparisons giving a more robust set of data.

4.2.1.1 Preparation

The ten Web pages used in the previous pilot study were also used in this study. Table 4.1 lists these Web pages along with the ID used for each page. Screenshots were taken for these pages (see Appendix C) and their source code was saved the same day and close to the time the screenshots were taken. For each of these page, the source code was manipulated in two ways: removing and adding edges.

In this way, two sets of Web pages are created: *Less Edges* and *More Edges*.

In order to create the **Less Edges** version of each page the following actions were taken to keep the content of the page the same but change the spacing around most of the sections/areas of the page:

- Cellpadding within tables and borders around tables were set to zero,
- Background colours were changed to white,
- Removed any list styling in order to merge various lists into one with the same style. Any extra space between the list items was decreased,
- Removed
 elements, and merged paragraphs to have continuous text and decrease any extra space,
- Merged tables and removed empty rows and columns that help to create space and alignment,

To create the **More Edges** version of each page the following actions were taken:

- Increase the table borders to be visible,
- Added rows and columns to create more space by usually adding *<tr><td>
</td></tr>*,
- Added breaks (*
) and sections (<div>
</div>*) within various sections to make them look as if they are different sections,
- Split existing tables to make sections more invisible and with the additional borders created more sections.
- Used more invisible images previously used within a page to create more space between various sections,
- Added background colours to create more visible areas

The above changes were made in the HTML file of the page and in any associated CSS files. When all page versions were created screenshots were taken of the manipulated pages. All 30 screenshots are shown in Appendix E.2.

The 20 possible comparisons were put on a PowerPoint slide show. A set of 11 slide shows were created by using counterbalancing and randomization of the order of comparisons to avoid any bias effects such as practice and boredom.

4.2.1.2 Participants and Procedure

The study took place as part of visit days by prospective students and their parents to the School of Computer Science. Eighty-one prospective undergraduate students and their parents volunteered to participate, 58 male and 23 female. Fifty-six (70%) participants aged between 16 and 25, 16 (19%) between 46 and 65 and the rest between 26 and 45. All of them had normal or corrected vision and 73 participants had English as their native language. Two participants were color-blind and due to the study's type their data were dropped. Most of the participants (90%) reported that they use the Internet daily or a few times a week with 64% of them using it for business/work, 90% for email/chat, 72% for special interests, and 66% for online purchases.

Participants were sat on a similar set-up as during the previous pilot study which is described in Section 4.1.1.1. The scripts read for introducing the experiment and the materials used for providing consent and the answer sheet can be found in Appendix E.

4.2.1.3 Materials

The pages participants viewed were the same ten Web pages used for the first pilot evaluaion. Screenshots of the ten (10) original images are shown in Appendix C and their transformed versions (with Less and More Edges) in Appendix E.2 and Table 4.1 lists the Web pages used for the experiment and their complexity level.

4.2.2 Results

Similarly with the first pilot evaluation, for the data analysis, we used -3, -2, -1, 0, 1, 2, 3 for the responses. Table E.1 lists the comparison data given by each participant. Terminology of various statistical terms used throughout the analysis can be found in Appendix A.

First, we ran a one-sample t-test, the results of which gave an initial feedback for further analysis. The mean value of each column of Table E.1 show how on average participants compared their complexity. A zero mean value means that the participants overall believe that the two pages do not differ. Instead of using these mean values for creating our ranking sequence, we first test, for each column, a null hypothesis that its true mean is zero. The more a sample (observed) mean differs from zero and the more consistent the individual responses are, the more significant its difference from zero is. To compare the means and test the null hypothesis (that all images have the same level of visual complexity) we ran a one-sample t-test [Field 2005] on all comparisons.

Table 4.4 shows the means for all the pairs that are significant at $\leq .05$. That is, for every paired comparison where the t-test revealed a significance we list the mean value. Otherwise we set it to zero. Tables E.2 and E.3 list the one-sample descriptive statistics (mean values of all comparisons) and the t-test results respectively. The higher the number, the more complex a given page was overall judged and a zero value means that it was judged as neutral. Figure 4.3 shows the Web pages in order of visual complexity and also shows which pages were preassigned as simple or complex. We can notice that simple pages with less-edges received lower rankings than complex pages, with the Yahoo as an exception. Also, with the results from the data analysis on the more-edges rankings we can not make solid conclusions. The graph and table shows only that three pages were perceived having different level of complexity than the original, with a preassigned visually simple page (Google Search) receiving the highest ranking.

Table 4.4: Pilot IIa - The Sum of Mean Values for Comparisons Significant at $\leq$ .05 (based on Tables E.2 and E.3).
The higher the number, the more complex a given site was overall judged and a zero value means that it was judged as neutral.

Page ID	Page Name	LESS EDGES	MORE EDGES
P1	Amazon	0.864	0
P2	Annotea	0.506	0
P3	BBC	0.8148	0.3827
P4	Firefox	0.802	0
P5	Flickr	0	-0.370
P6	GoogleSearch	0.667	0.728
P7	MSN	1.235	0
P8	Orkut	0	0
P9	WAI	0.432	0
P10	Yahoo	0.444	0

To test the interaction effects between the number of edges as well as the pages used, we run a Repeated Measures General Linear Model (GLM or ANOVA) on a 3 x 10 within subjects design, where 3 there are types of edges: less, more and original and 10 the types of pages. Mauchly's Test indicated that the assumption of sphericity has been violated for the interaction effect of edges on pages, $\chi^2(170)$

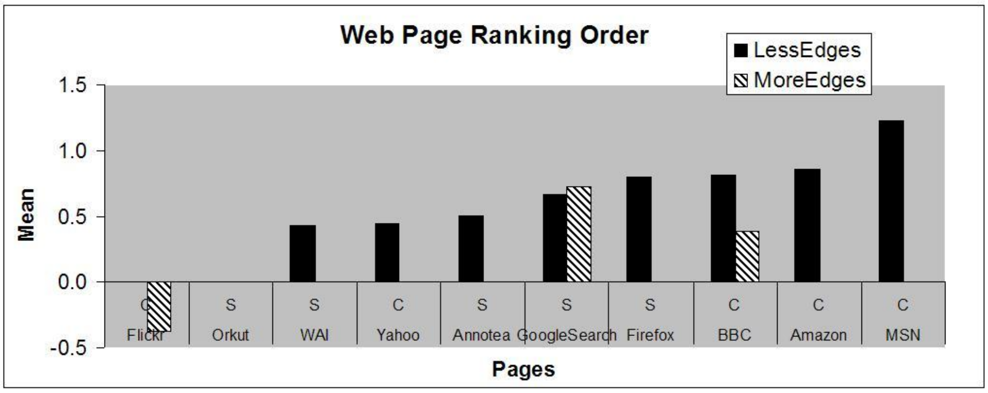


Figure 4.3: Pilot IIa - Ranking Order based on Complexity Mean Values

= 343.4, p<.001. Therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\epsilon = 0.79$ for the interaction effect). This analysis revealed that all effects are significant at p<.001, which denotes a possible relation between the number of edges and users' perception of visual complexity.

A significant main effect was determined on the number of edges on the ratings, with F(2, 160) = 22.48, p<.001. *Post Hoc* pairwise comparison analysis, using Sidak adjustment, revealed a significant main effect between the rankings given on pages with less-edges and more-edges and between the rankings of less-edges and original. Less-edges overall received higher ranking scores (Mean = 4.6) than more-edges (Mean = 4.08) and original pages (Mean = 4). This shows that pages with less-edges were perceived by the participants as more complex than the other type of pages. This could happened for various reasons. One reason could be the fact that pages with less-edges looked disorganised due to the lack of layout and structural elements, since they were removed, forced users to give higher rankings of complexity.

A significant main effect was determined on the type of pages, with F(9, 720) = 5.132, p<.001. *Post Hoc* pairwise comparison analysis, using Sidak adjustment, revealed a significant interaction effect between the following pages: P1 - P5, P3 - P5, P3 - P8, P5 - P6, P5 - P7, P7 - P8, P8 - P6. This result shows how significantly different were pages ranked by users which provides support towards our assumption between the effect of structure and visual complexity perception. A significant interaction effect was also determined of the number of edges on the type of pages (visually complex/simple), with F(14.134, 1130.72) = 3.46, p<.001.

4.2.2.1 Discussion

The above results revealed a possible relationship between the number of edges a page has and users' perception on visual complexity. However, the analysis could not reveal robust results on the direction of this relationship. This was a pairwise comparison that examined how the manipulation of this space can affect users' perception. The study showed that pages with less organisation and space between different sections received higher rankings of visual complexity than the original page or the ones with more space. By extending this evaluation into a better experimental design will help determine how the Visual Complexity Score of a Web page can be predicted by its structural organisation. As the authors in [Henderson *et al.* 1999; Rayner 1998] explain, the cognitive and semantic aspects of a stimulus affect visual and scene perception. The organisation of a page can be defined by the different sections or subjects, the tables or colours that group a set of information. Further studies need to be conducted to show whether less or more sections a page has the more complex it is perceived by users.

4.3 Web page Ranking - Pilot Evaluation IIb

The next stage towards defining complexity was a ranking study, where ten Web pages were rank by users based on their visual complexity perception. In this way, we tried to increase our data and collect further implicit information in order to understand how sighted users perceive Web page visual complexity. This study was the second part of the second pilot evaluation.

4.3.1 Methodology

In traditional ranking experiments, a complete factorial paired-comparison design is run. That is, each participant should compared all possible combinations between the ten pages, that is 45 comparisons. However, as with the Pilot Evaluation I, time limitation was a big constraint as the experiments were part of the department visit days. Therefore, we decided to run a basic ranking experiment were ten images are shown to participants and asked to be scored based on their visual complexity level. In this way, a ranking score was obtained for each Web page. Then, a correlation was examined between the two sets of results (pilot I and IIb). The experiment was run in groups of 4-10 participants. Each image was shown twice so that participants could change the assigned rank if they wanted to. Randomization and counterbalancing was used to obtain a different stimuli sequence for each group to avoid further biases.

The study was part of the Pilot Evaluation II, so the experiment set-up, participants and equipment are described in Sections 4.2.1.2 and 4.2.1.3. During this part of the experiment the participants asked to give a score for each image they saw from 1 to 10, with 1 being very visually simple and 10 as complex. The images were shown on a projector screen in front of them. The scripts read for introducing the experiment and the materials used for providing consent and the answer sheet can be found in Appendix E.1.

Table 4.5: Pilot IIb: Ranking Score Mean Values

Web Page ID	Web Page	Visual Complexity Level	Ranking Mean	Ranking Order
P4	Firefox	Simple	3.60	1
P5	Flickr	Complex	4.06	2
P8	Orkut	Simple	4.14	3
P9	WAI	Simple	4.79	4
P6	GoogleSearch	Simple	5.10	5
P3	BBC	Complex	5.22	6
P2	Annotea	Simple	5.48	7
P7	MSN	Complex	6.02	8
P10	Yahoo	Complex	6.26	9
P1	Amazon	Complex	6.89	10

4.3.2 Results

Participants were asked to give a score for each image shown based on their visual complexity by giving a number from 1 to 10, with 1 being the visually simplest and 10 the most complex. They were given the opportunity to change their ranks as the images were shown twice. Table E.4 lists the ranking data given by each participant. Table 4.5 lists the mean values of the ranks given for each page in ascending order along with the preassigned level of visual complexity. Figure 4.4 shows the Web pages in order of visual complexity and the preassigned level of visual complexity (simple or complex). One can notice that simple pages received lower rankings than complex pages, with the Flickr and Annotea as the exceptions. Flickr page was also an outlier in the Pilot Evaluation I which is an indication of the data validation and consistency. This could then mean that our initial framework was not valid for pages like Flickr that were short with only images and one purpose only (here signing in a portal). The familiarity could be the reason that the Annotea page received higher rankings as it was an academic and subject specific page. The participants were mostly below the undergraduate level which meant that they would never see such page before. However, the reason we introduced this page in the sample set was that it followed a very simple layout with a menu and main content hoping that the unfamiliarity with its subject would not affect the participants, something that appeared to bias the data in the end.

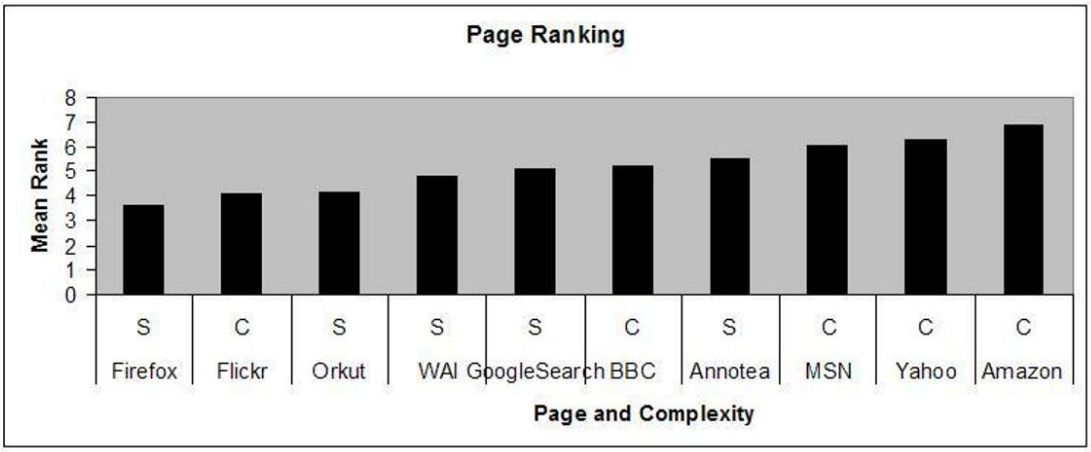


Figure 4.4: Pilot IIb - Ranking Order based on Complexity Mean Values

Table 4.6 lists the complexity mean that each page received from the both pilot studies (Pilot Evaluation I and IIb). Running a Spearman correlation test, r_s, a significant correlation was revealed between the two means data sets. A positive relationship was determined between the ranking mean and the paired-comparison mean values, $r_s = .68$, $p < .05$. This correlation shows that as the ranking mean increases the paired-comparison mean increases as well. The high value of the correlation ($> .5$) shows that the effect of this relationship is high and therefore a strong result.

Figure 4.5 shows the ranking order for both versions in paired-comparison's complexity mean ascending order. From both Table 4.6 and Figure 4.5 one can notice that pre-assigned visually simple pages were the ones that have significantly different rankings but are still grouped in both versions as simple pages.

Table 4.6: Complexity Means from the Pilot studies I and IIb

Web Page ID	Web Page	Visual Complexity Level	Ranking Mean	Pairwise Comparison Mean
P1	Amazon	Complex	6.89	0.61
P2	Annotea	Simple	5.48	-0.43
P3	BBC	Complex	5.22	0.50
P4	Firefox	Simple	3.60	-0.28
P5	Flickr	Complex	4.06	-0.45
P6	GoogleSearch	Simple	5.10	-0.45
P7	MSN	Complex	6.02	0.37
P8	Orkut	Simple	4.14	-0.53
P9	WAI	Simple	4.79	-0.34
P10	Yahoo	Complex	6.26	1.00

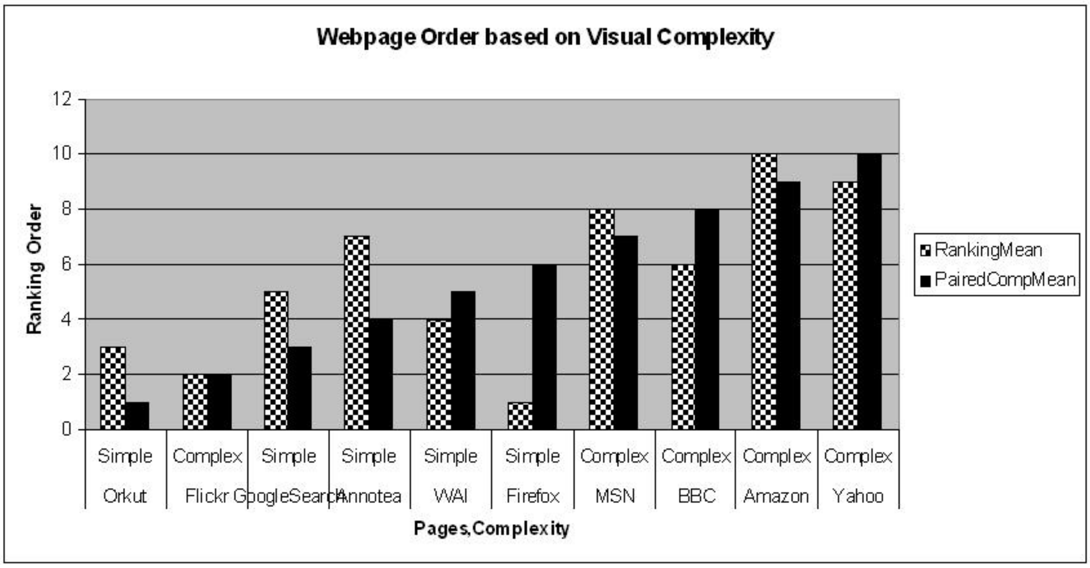


Figure 4.5: Ranking Order for Pilot I and IIb

4.3.2.1 Discussion

During this study, it was investigated how users rank Web pages based on their visual complexity. A page ranking was determined based on the visual complexity of the Web pages. The results were consistent with previous experiments on visual complexity showing that preassigned visually simple pages received lower rankings than visually complex pages. The complexity sequence is compared with results from a previous pilot experiment showing a significant correlation between the complexity means that each page received by both experiments.

This study acted as a draft evaluation were we wanted to examine if the perception of users based on page layout differences related with complexity perception. The results are not strong but provided as the basis into building further investigations to identify how structural elements affect page visual complexity.

4.4 Web page Online Comparison - Chunks

As the previous pilot study (IIb) showed, the number of different sections a page is organized into affects the visual complexity of the page. That is, the less *chunks* (sections/edges) a Web page is arranged into, the more visually complex it becomes. By extending the previous evaluation into a better experimental design we believe that can provide further support to determine how the Visual Complexity Score (VCS) of a Web page can be predicted by its structural organisation. As the authors in [Henderson *et al.* 1999; Rayner 1998] explain, the cognitive and semantic aspects of a stimulus affect visual and scene perception. The organisation of a page can be defined by the different sections or subjects, the tables or colours that group a set of information. This section describes an online pairwise experiment that extended the previous pilot study (Pilot IIa) and the effect of the page's layout on users' perception of visual complexity was examined.

4.4.1 Classification

A Web page's design layout is determined by the arrangement and characteristics of its structural elements. It is the "creative art of executing aesthetic or functional design" [Merriam-Webster 2006] and it is clear that Web site design can either facilitate or impede a user through the available resources. In order to understand how the visual elements can affect a user's cognition and visual perception, we focus on the comprehensive layout rendering of a Web page. A page is separated into various sections and subjects. This visual distinction is made with the use of colours, tables, lines and spacing. This section describes the steps that we followed to create the chunk rendering of a Web page in order to use them for the evaluation. This chunk rendering is a representation of the overall Web page layout without any visual elements.

4.4.1.1 Creating the Chunk Rendering

Each Web page has a distinct layout and information organization. In order to create a chunk rendering of a Web page's layout, a list of structural elements that distinguish and form that layout need to be identified. For our procedure and experiment, the set of variables used to identify the chunk rendering of the page are the background colours, headings and subsections, standalone images, and visible lines or borders. The original Web page images and their resulted chunk renderings generated following the classification described in this section can be found in Appendix F.

Background Colours The correct use of colours on a Web page aesthetically enhances the presentation of the page to group information and differentiate sections in a page that are placed close or are similar to each other. These sections include horizontal menus or sections that require user's input. Background colours are also used to group similar information, or to enhance user's reading. For our classification, we identify areas that are distinguished with background colours. For example, points 1 and 2 in Figure 4.6 show an example of how the background colour is used to distinguish a horizontal menu from an input form that are closely placed next to each other. The colours help users to instantly notice both of them.

Visible Lines/Borders Designers also use table borders to group relative information so its specific purpose can be easily noticed. For example, a list of links surrounded by lines creating a box can be easily identified by the user as a menu list. Points 3 and 4 in Figure 4.6 show an example of how the visible lines are used to box information and to differentiate it from the rest of the page. The lines can be in any colour, width or style and can be surrounded by any background colour (most commonly the white) or any other structural elements.

Headings In printed text, headings provide information about the subject that follows. Similarly with a Web page, headings are used to inform users what type of information follows. As in printed text as well, Web page headings are also used to separate different subjects in sections. These sections may include text, links, images and can be distinguished by background colours, borders or just white space. Points 5, 6 and 7 in Figure 4.6 show Web page areas that are separated by headings. It is important to notice that sections pointed out by their headings can be also within other sections. For example, as Figure 4.6 shows, point 5 is a subsection of point 3, both set apart from a heading.

Sections' Subsections Subsections are identified by subheadings, images, lists of links, text or paragraphs surrounded by white space. An example of subsections distinguished by images and list of links is shown as points 10, 11 and 12 in Figure 4.6. Point 5 can also be described as a subsection of point 3. A subsection often includes an image to separate out the subsection and as it will be described in the next section, these images do not define a different section but are as part of the subsections.

Standalone Images Logos, advertisements, animations or decorative images help a designer to organize a page and attract users' attention. For our classification, we point out images that stand alone on the Web page. These include, advertisements, logos or informative images and can be surrounded by borders, background colours or white space. Points 8 and 9 in Figure 4.6 show two examples for our image classification.

4.4.1.2 Organizational Elements

When the chunk rendering of a page is created and the overall layout of the page is revealed, organizational elements need to be identified for the evaluation. Figure 4.7a shows the chunk rendering generated when classification described above is applied on Figure 4.6. A variety of different sections are shown: sections that are only surrounded by white space, sections that are adjacent and share a side or sections that are subsections. Categorizing these groupings revealed a set of variables that will be used to determine how they affect users' perception of a Web page's visual complexity.

Box We define as box the area that is enclosed by four lines (that is, a rectangle). A box can be adjacent to other boxes, within another box or surrounded by white space. Boxes of 4.7a are shown in Figure 4.7b filled with grey colour.

Block A block is a box that is surrounded by white space only. A block can be a set of boxes that are adjacent to each other, or a box that is standalone and

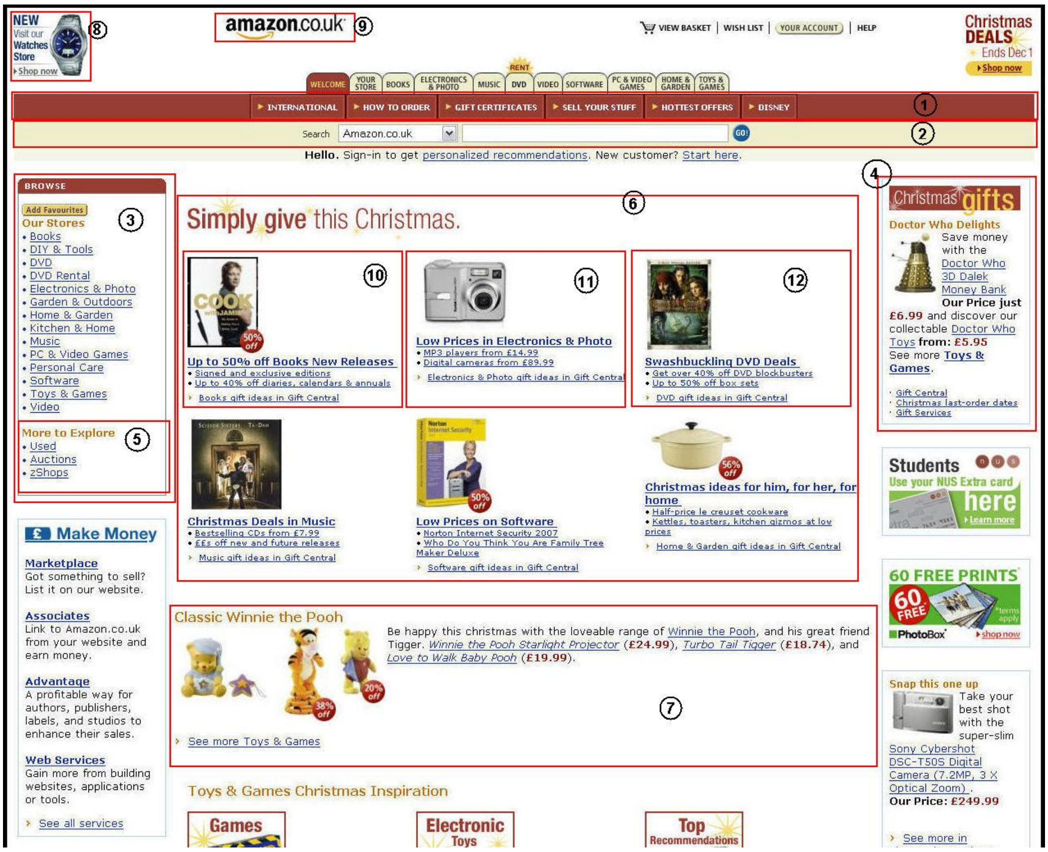


Figure 4.6: Example of how the classification of different structural elements helps to create the overall layout of a page

surrounded by white space. Figure 4.7b shows the blocks identified in rounded rectangles with black lines.

Comparing the number of blocks a page is organised into from the number of boxes we believe that can help us determine how clear and uncluttered the page is. For example, a page with a large number of boxes but a small number of blocks can be a page that has a significant number of different sections placed and grouped close to each other. This can be potentially a visually complex page because it will have a high amount of information but not enough spacing or clear layout to help the user to instantly identify the information that is looking for.

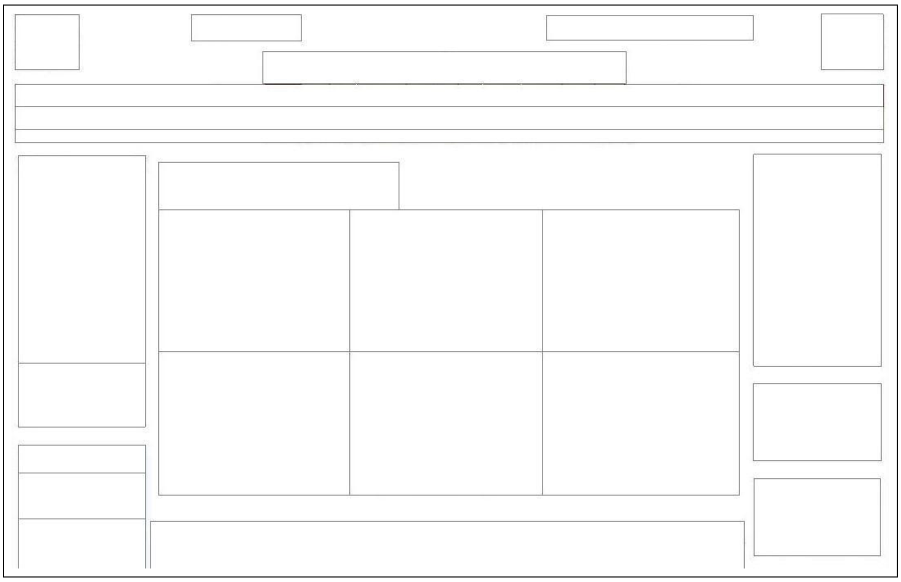
Corner A corner is the intersection of two boundary lines. For example, a box will have four corners, but if the box is adjacent to another box, then the total number of corners will be six. Figure 4.7b labels the corners as black filled circles. We assert that the more corners a page has, the more sections it is organised into and therefore the more visually complex it can be perceived by a user.

Top-Left-Corner (TLC) In order to understand how the number of blocks and the overall structure of the page affect user's perception we define the Top-Left-Corner (TLC) variable. TLC is a block's top left corner. If a box's left and top sides are not adjacent or have a common side with another box, then its TLC is also counted. Point 1 in Figure 4.7b shows a block with two TLCs because the top box has a top left corner's side not common with another box. TLCs are shown as curved dotted lines in Figure 4.7b. We believe that if the number of TLC is not equal to the number of blocks on a Web page, the overall layout of the page might have something different or unique. An example of this uniqueness is shown by point 1 in Figure 4.7b.

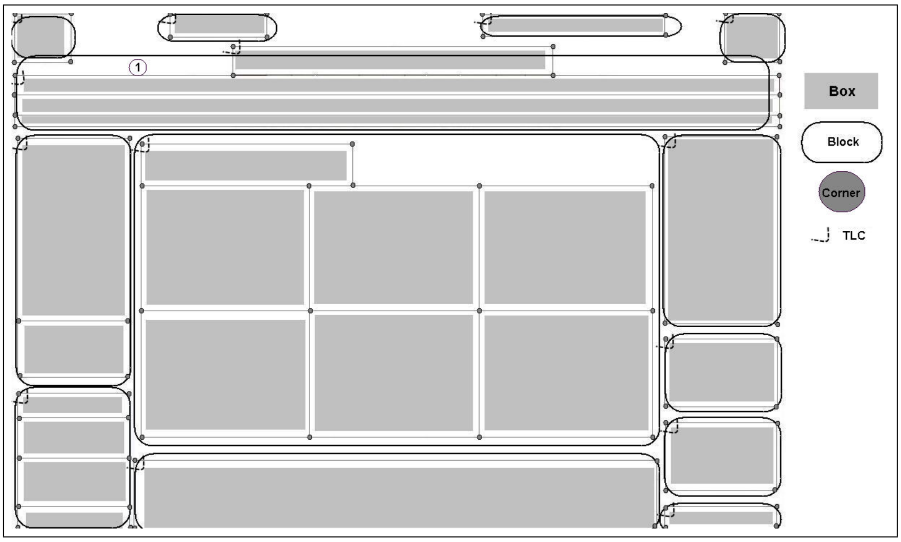
4.4.2 Hypotheses

We believe that the *chunk rendering* of a Web page is positively related with the level of visual complexity of the page. That is, the more sections a page is grouped in, the more visually complex it becomes for the user to perceive. By *chunk rendering*, we mean the number of different sections a page has. The set of hypotheses examined in this evaluation are:

- **H1**: The number of boxes a Web page is presented with is linearly related with the visual complexity level of the page. As the number of boxes



(a) Chunk Rendering



(b) Categorization

Figure 4.7: Chunk Example - Categorizing a Chunk Rendering

increases, the level of complexity of the page increases as well.

- **H2**: The number of blocks a Web page is sectioned in is linearly related with the visual complexity level of the page. As the number of blocks increases, the level of complexity of the page increases as well.

- **H3**: The number of corners a Web page is grouped in, is linearly related with the visual complexity level of the page. As the number of corners increases, the level of complexity of the page increases as well.

- **H4**: The number of top-left-corners a Web page is structured in is linearly related with the visual complexity level of the page. As the number of top-left-corner increases, the level of complexity of the page increases as well.

4.4.3 Methodology

For this experiment, we used the pairwise comparison technique which is described in Section 4.1.1. A set of ten Web pages were modified using the algorithms defined in Section 4.4.1 to create their respective blueprints (chunk renderings). Screenshots of the original images and their chunk renderings can be found in Appendix F.2. A randomization algorithm was implemented to assign a random sequence of 45 image pairs. To avoid any order effect, the algorithm automatically produced a randomized sequence for each participant. To test for the boredom effect, the time taken to give each score was collected for every participant. In addition, in order to have score for comparing both AB and BA pair versions and have a short experiment for time control, two sets of 45 image pairs were created and incorporated into two different questionnaire versions, where the only difference was the images order of pairing they looked at. Participants were automatically assigned to one of those versions, and a balance between the number of Version A and Version B questionnaire appearance was always kept.

4.4.3.1 Participants

The study was available online so participants could access it in their own time and place. It was advertised through mailing lists and newsgroups. Ninety participants from around the world volunteered to take part in this evaluation, 63 (70%) male and 27 (30%) female. Twenty - nine (32%) participants were aged up to 25, 6 (7%) between 46 and 65 and the rest between 26 and 45. From them, 29 (32%) participants had English as their native language. One participant was colour-blind and due to the study's type the data were dropped. All of the participants reported that they use the Internet daily or a few times a week with 38% of them more than twenty hours, only 2% less than an hour, 21% for 1-5 hours, 17% for 6-10 hours and the rest between 11 and 20 hours. 97% of them described that they use the Web for business/work, 98% for email/chat, 92% for special interests, and 65% for online purchasing. Figures 4.8 and 4.9 show for how long users described they use the Internet and for what purpose. One can notice that most of the sample use the Internet daily and for more than 20 hours. Also, all participants reported to be familiar with all kinds of browsing (business/chatting/purchasing) which made our sample even more generalisable.

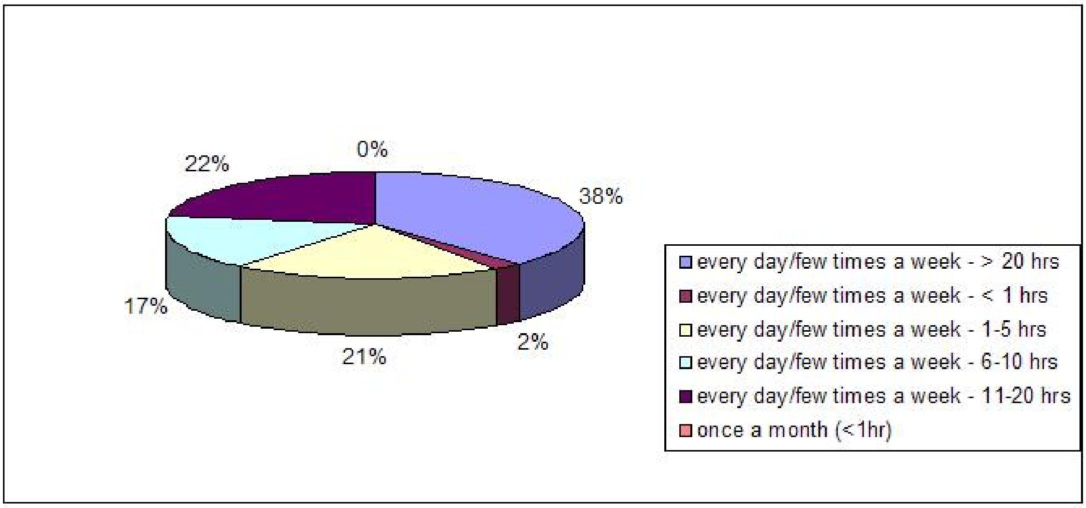


Figure 4.8: Participants' Internet Usage per Hours Daily or Weekly - (hours per week)

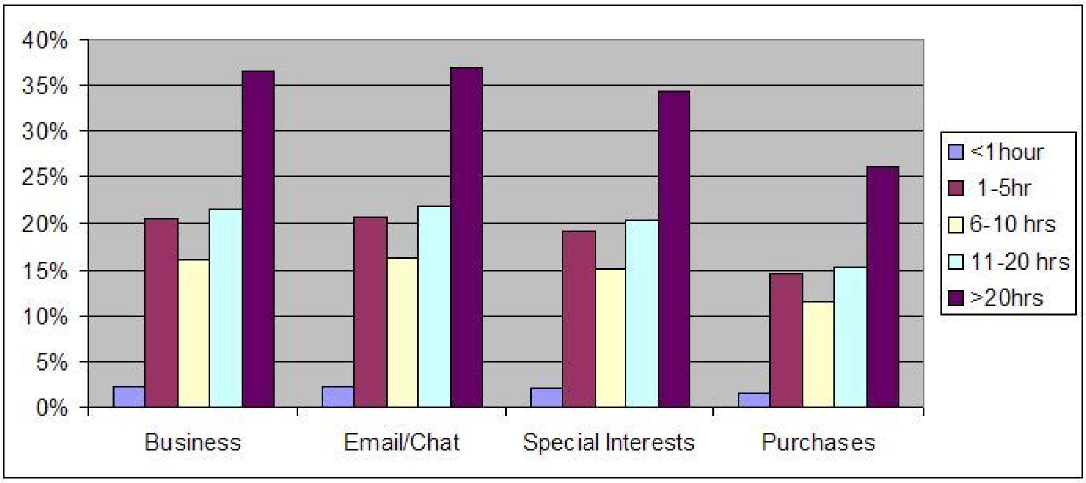


Figure 4.9: Participants' Purpose of Browsing per Hours

4.4.3.2 Procedure

Each participant could access the experiment in their own time. The study was hosted under the Human Centred Web Lab[1] website. Being now closed, the study is moved from its original place[2] to the ViCRAM Closed Studies[3] space for demonstration purposes only. The participants were introduced to the experiment with a summary of the experiment (see Figure F.1). The introduction also explained that the evaluation was anonymous and participants were free to withdrawn any time during the study by closing the browser. The participants were then asked to select the option 'Yes' in order to continue with the rest of the experiment, which acted as a consent in taking part in the evaluation.

The experiment consisted of two parts. During the first part, as Figure F.2 shows, participants had to fill a demographic questionnaire about their age, sex and Internet and browsing familiarity. The main study was conducted in the second part, in which after a short introduction (see Figure F.3) the participant looked at a pair of images and answered the comparison question on their own time (see Figure F.4). After looking at the set of the randomized 45 pairs, the participants were thanked for taking part and provided with more information about the experiment (see Figure F.5).

4.4.3.3 Materials

The ten Web pages used in the previous pilot studies were also used in this study. Table 4.1 lists these Web pages along with the ID used for each page throughout the study. Screenshots of the original and chunk rendering of each Web page can be found in Appendix F.2. In order to avoid any bias, the images were taken as a screenshot from the same monitor. Then the modification of the page concentrated on the top left corner of the page in an area of 500 x 300 pixels for every image. The top left area of the Web page was selected to control the sample set and evaluation as well as to investigate whether this small area can affect users' visual perception. The experiment was run using Perl under the Human Centred Web lab server and data would be automatically saved in the 'surveydata.txt' file. Both the data file and source code ('ranking.pl') can be downloaded from the associated experimental materials folder linked to the

[1] HCW Lab - http://hcw.cs.manchester.ac.uk
[2] http://hcw.cs.manchester.ac.uk/research/vicram/studies/comparison/
[3] http://hcw.cs.manchester.ac.uk/research/vicram/studies/closed/comparison/

technical report of this study (see page 265).

4.4.4 Results

The raw data collected during this evaluation can also be downloaded from the experimental folder linked to the technical report of this study (see page 265). The associated folder also contains Excel and SPSS[4] files that were created during the data analysis procedure. Terminology of various statistical terms used throughout the analysis can be found in Appendix A.

Table 4.7: Chunks Pairwise Comparison - Correlation Coefficients

Page/Image	ImageID	VersionA	VersionB	AverageMean	Complexity Order
Amazon UK	ID1	0.851	1.158	1.005	10
AnnoteaProject	ID2	-0.842	-1.08	-0.961	2
BBC UK	ID3	0.356	0.367	0.362	6
Firefox	ID4	-0.216	-0.493	-0.355	4
Flickr	ID5	-1.171	-1.242	-1.207	1
GoogleSearch	ID6	0.771	0.638	0.705	8
MSN	ID7	-0.004	-0.087	-0.046	5
Orkut	ID8	-0.796	-0.956	-0.876	3
WAI	ID9	0.709	1.104	0.907	9
Yahoo UK	ID10	0.342	0.591	0.467	7

4.4.4.1 Score Version Correlation

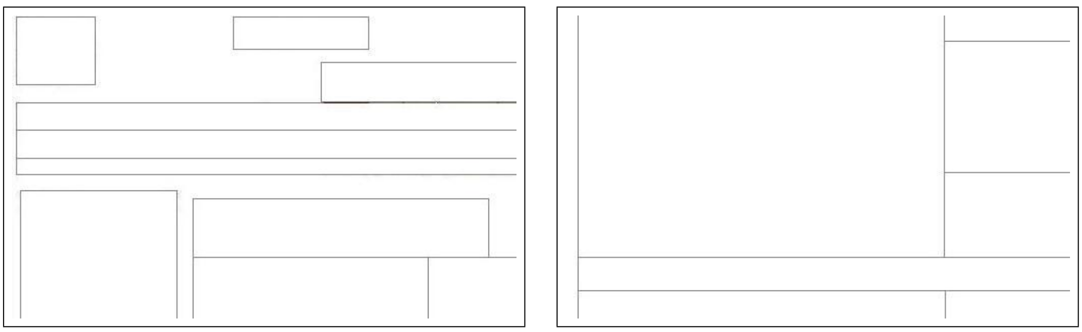
There were two versions of experiments during this study: versionA and versionB. In versionA images were paired as X-Y and in versionB as Y-X. For example, on a pair X-Y, image comparison was made for image X *vs* image Y, with X on the left side and Y on the right side, where image comparison of pair Y-X was made as image Y *vs* image X with image Y placed on the left and X on the right side. In order to merge the rankings from both versions, their score correlation needs to be examined. An independent samples t-test is run since there are two set of scores and different participants for each set. The test is performed on X-Y scores and on the negation of Y-X scores so that both score sets have the same direction of complexity. For example, a participant's (say pA) score for the comparison pair 6-7 of value -3 describes that image6 is more complex than image7. On the other hand, a participant's (say pB) score for the comparison pair 7-6 of value -3

[4] Originally: Statistical Package for the Social Sciences - Statistical Analysis Software, http://www.spss.com

describes that image7 is more complex than image6. So, in order to merge their data we need to negate pB's score and for the pair 67 assign the scores of -3 and 3 respectively. The correlation analysis revealed significant difference between the means of the two samples of the image pairs: 8-1, 2-9, 3-5, 6-7, 8-9, 8-10, 7-9, 6-9, 4-9, 2-9, 2-3, 7-1, 6-1, 2-1. This happened because users had to select a score from a 7-point scale which gives a large range of possible scores. Further analysis on the ranking order for each score version will give more information on the version correlation.

4.4.4.2 Ranking Order

Based on the scores that the participants provided for each version (versionA, versionB), three average score versions are calculated: MeanA, MeanB and AverageMean. A correlation analysis between the two versions is performed which revealed that scores of MeanA and MeanB were significantly correlated with a high effect of $r = .98$, $p < .0001$. Table 4.7 lists the mean values of complexity that each page was assigned with for the three versions. Figure 4.11 shows the ranking order based on the average complexity score of each image. The smaller the value the image has, the less complex it was ranked as by the participants. One can notice that image with ID5 (Flickr) is ranked as the visually simplest page, where image with ID1 (AmazonUK) as the most complex. Figures 4.10b and 4.10a show the two images respectively and one can instantly notice their difference on the layout and specifically from the number of chunks that are rendered. The last column of Table 4.7 shows the complexity order in which the images were ranked.



(a) Amazon - Visually Complex *(b) Flickr - Visually Simple*

Figure 4.10: Chunk Rendering Comparison: (a) Amazon *vs* (b) Flickr

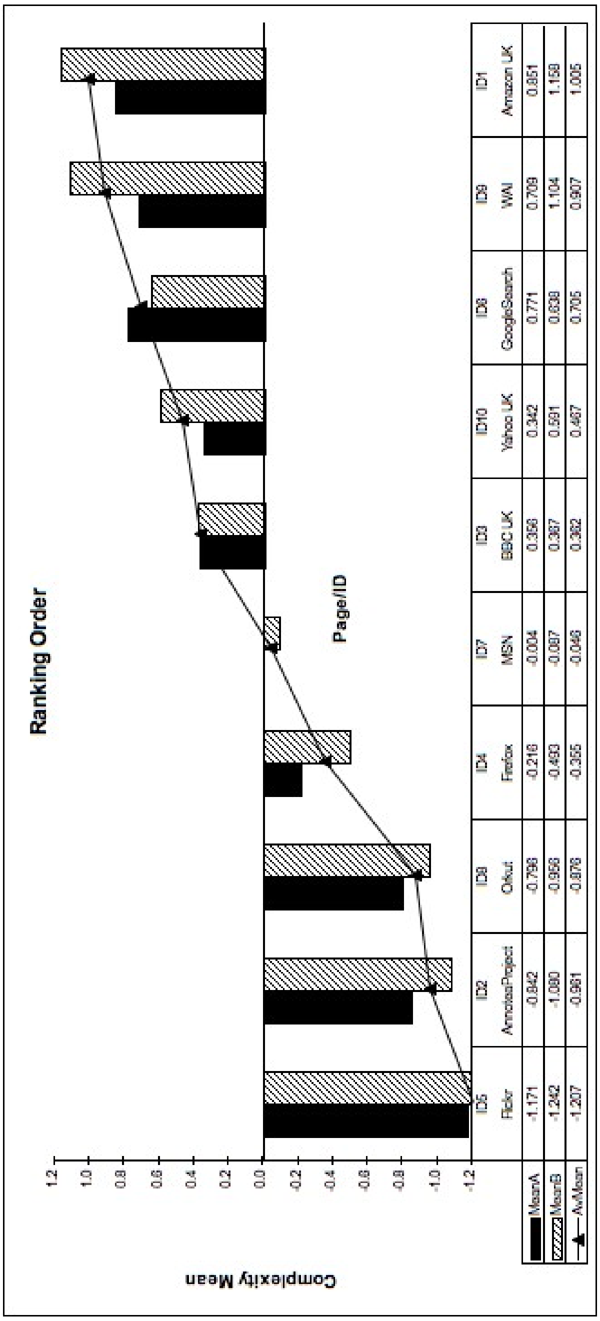


Figure 4.11: Ranking Order based on Complexity Scores

4.4.4.3 Relationship with Page Layout

Table 4.8 lists the number of Boxes, Blocks, Corners and TLCs each image's chunk rendering has. In order to test our hypotheses that the higher number of each organizational element a page has, the more visually complex it will be perceived by the user, a series of statistical analysis tests which include correlation and multiple regression were performed.

Table 4.8: Mean Complexity Score each page received by participants and number of Organizational Elements for Each Stimuli/Page based on classification

Page/Image	ImageID	MeanA	MeanB	AvMean	Box	Block	Corners	TLC
Amazon UK	ID1	0.851	1.158	1.005	10	5	21	6
AnnoteaProject	ID2	-0.842	-1.08	-0.961	5	3	12	3
BBC UK	ID3	0.356	0.367	0.362	11	10	26	10
Firefox	ID4	-0.216	-0.493	-0.355	7	3	7	3
Flickr	ID5	-1.171	-1.242	-1.207	7	1	6	1
GoogleSearch	ID6	0.771	0.638	0.705	9	5	22	5
MSN	ID7	-0.004	-0.087	-0.046	10	3	15	3
Orkut	ID8	-0.796	-0.956	-0.876	4	2	7	2
WAI	ID9	0.709	1.104	0.907	14	4	24	4
Yahoo UK	ID10	0.342	0.591	0.467	18	7	38	7

Correlation A bivariate correlation test was performed to examine the relationship between the mean scores and the number of the independent organizational elements: box, block, corner and TLC. The one-tail test was used for this analysis since there was a specific direction to the experiment's hypothesis. The hypothesis, as explained in Section 4.4.2 stated that the higher number of the organization elements a Web page has, the higher level of complexity the page will be assigned with, which shows a positive direction. The correlation analysis revealed that all independent variables were highly and significantly correlated with the mean scores. Table 4.9 lists the correlation coefficients for every variable and it is important to note that all correlations have a high effect (>.63).

Multiple Regression Using the AverageMean as the dependent variable and the number of boxes, blocks, corners and TLCs as the predictors a series of stepwise regression analysis was performed. The data analysis could only reveal individual linear models with Box, Corner and TLC as the predictors respectively. Table 4.10 lists each predicted model's Summary and Coefficients. The equations

Table 4.9: Correlation Coefficient between Means and Organizational Elements for Each Stimuli

Mean	Box	Block	Corner	TLC
MeanA	$.662^b$	$.633^b$	$.734^a$	$.669^b$
MeanB	$.740^a$	$.613^b$	$.777^a$	$.655^b$
AverageMean	$.703^b$	$.593^b$	$.745^a$	$.635^b$

a: p (1-tailed) < .01; b: p (1-tailed) < .05

produced by the data analysis that provide an estimation of the visual complexity score (VCS) of a Web page are:

$$VCS = -1.191 + 0.138(Box) \tag{4.1}$$

$$VCS = -0.941 + 0.060(Corner) \tag{4.2}$$

$$VCS = -0.741 + 0.196(TLC) \tag{4.3}$$

The ratio between the number of boxes and blocks was also calculated. An initial assumption was that when the number of boxes is larger than the number of blocks, could reveal a factor that affects visual complexity (see Section 4.4.1). However, analysis results showed that the complexity scores were not significantly correlated with this ratio.

In Section 4.4.1 it was also argued that when the number of TLC is larger than the number of blocks can be an additional indication of a complex layout. However, only image with ID1 did not have equal number of blocks and TLC and the test could not be executed. On the other hand, we can see from Figure 4.11 that the same image (ID1 - Amazon) was scored as the most visually complex image from the participants. This does not mean that the inequality between the number of blocks and TLCs is an indication of visual complexity but a factor that should be further examined in this context.

4.4.4.4 Discussion

This evaluation was an investigation on how the overall structure of an area of a Web page can affect users' visual perception by just looking at the chunk rendering of the page. A chunk rendering of a Web page's layout does not involve

Table 4.10: Model Summary and Coefficcient for Box, Corners and TLC

IndepVariable	**R**	R^2	R^2**adj**	**SE Est.**	**F-ratio**	**Durbin-Watson**	**Model**	**B**	**SE B**	β	**t**	**sig.**
Box	0.703	0.494	0.430	0.622	7.798	2.437	(Constant)	-1.191	0.509		-2.341	0.047
							Box	0.138	0.049	0.703	2.792	0.023
Corner	0.745	0.556	0.500	0.583	10.002	2.578	(Constant)	-0.941	0.383		-2.460	0.039
							Corner	0.060	0.019	0.745	3.163	0.013
TLC	0.635	0.403	0.328	0.675	5.401	2.203	(Constant)	-0.741	0.427		-1.734	0.121
							TLC	0.196	0.084	0.635	2.324	0.049

any aesthetic elements and the classification procedure was explained. A list of organizational elements were also identified that were used to investigate their relationship with participants' visual complexity scores. Results showed that by removing the visual attractions of a Web page and isolating the overall layout of the page can also affect users' perception and when used can estimate the visual complexity score of the page.

All of the experiment's hypotheses were supported (**H1 - H4**) and results demonstrated that the number of boxes, blocks, corners and top-left-corners a Web page's chunk rendering has, was significantly correlated with participants' visual complexity score of the page. Three linear models were also determined by the data analysis with the number of box, corner and TLC individually estimating the visual complexity score of the page.

4.5 Conclusions

This chapter described a series of pilot evaluations that were conducted to enhance our understanding of sighted users' visual perception. A ranking sequence of ten Web pages was obtained based on users perception about the page's visual complexity level. A possible relation was revealed between the preassigned complexity level and the participants' perception on visual complexity which provided further evidence that visual complexity depends on the density and diversity of Web page elements. Then, the following pilot evaluation showed that by increasing and decreasing the space that surrounds and identifies different sections within a page affects users' visual perception of the page which enhanced our assumption that the position of elements and overall layout of the page is another factor of visual complexity.

By extending the above pilot studies, a better designed and controlled experiment was performed to investigate how the overall layout of the page can predict the level of visual complexity of a Web page and how it can be used to define complexity. The evaluation showed that the number of boxes, blocks, corners and top-left-corner a Web page's chunk rendering has, was positively related with the visual complexity score of the page.

These series of evaluations provided fundamental information towards defining visual complexity based on users perception and page layout and work towards our second research question (Q2 - see page 23). We believe that, in order to provide a robust definition and prediction model of visual complexity a follow-up user evaluation should be designed. This should involve a larger set of Web pages than the one during these evaluations, create their chunk rendering and the number of their organizational elements and correlate user rankings with these variables. This chapter enabled us to understand the variables that affect visual complexity of Web pages as perceived by sighted users and future studies should evaluate the extend to which these variables affect complexity in order to determine a complexity prediction model.

Chapter 5

Visual Complexity and Aesthetics

Our previous work investigated the relationship between a user's perception and the visual complexity of a Web page. It was shown that visual complexity of Web pages depends on the presentation of the page's elements and by the density and diversity of the elements that are presented and overall layout of the page. The next stage towards modeling complexity was to determine the extend and weight that these factors have on visual complexity. As the need of a large set of stimuli and a more controlled design was also identified during the previous evaluations, a new study was designed in order to control various experimental problems such as familiarity and time of stimuli appearance to be able to expand our analysis, which this chapter describes.

In addition, during this study we wanted to explore users' perceptions of the aesthetic qualities of the Web pages. Human-computer interaction research is mostly emphasized on performance criteria, such as time to learn, error rate and time to complete a task [Butler 1996] and pays less attention to aesthetics. Lately, researchers have tried to understand how visual aesthetics can affect the viewer's perception [Hoffmann & Krauss 2004; Ivory & Megraw 2005; Zettl 1999; Lavie & Tractinsky 2004] but the relationship between the aesthetic presentation of a Web page and the user's interaction needs further exploration. By examining visual aesthetics as well we aim to understand how the user's perception is formulated by the visual appearance of a Web page derived by both the structural and aesthetic elements that describe a page.

5.1 Methodology

For this experiment, we asserted that the visual complexity of a Web page is related to the elements presented to the user through the structure of the document, by the quantity of each element that is used on the page and by the perception of its aesthetic properties. During this evaluation participants were asked to rank thirty Web pages based on their visual complexity and aesthetic appearance. We then related their rankings with the structural elements used to design each page. Raw data collected and stimuli used can be found at the technical report in (see page 265).

5.1.1 Design

The screenshots of thirty Web pages were the stimuli that all participants had to rank for visual complexity and aesthetic characteristics. A within-subjects experiment was designed to collect ranking scores for all images by all participants. The participants looked at each stimuli for exactly 7 seconds. The time was assigned for two reasons. First, from the eye-tracking study we conducted (described in Chapter 3) and from literature review it was shown that users can get an overall idea of the Web page within 5 seconds. Then, we ran a series of pilot studies within our lab to determine whether 10, 7 or 5 seconds was enough for a user to look at the pages and answer a set of questions. All users suggested that 7 seconds was enough for looking at an image and be able to remember it to answer the follow-up questions. To avoid any order effect, a randomized sequence of the thirty stimuli was assigned for each participant. In addition, the order was counterbalanced so each participant had to rank the same image twice but the second time was in reverse (counterbalanced) order. The time taken to give each score was collected for every participant to examine any unreasonable delays that might affected the user rankings.

5.1.2 Participants

The study was available online so participants could access it on their own time and place. It was advertised through mailing lists and newsgroups. Fifty-five participants from around the world volunteered to take part in this evaluation, 36 (65%) male and 19 (35%) female. Twenty (36%) participants were aged up to

25, 8 (15%) between 46 and older and the rest between 26 and 45. From them, 25 (45.5%) participants had English as their native language. Two participants were colour-blind and due to the study's type their data were dropped and not used for any analysis.

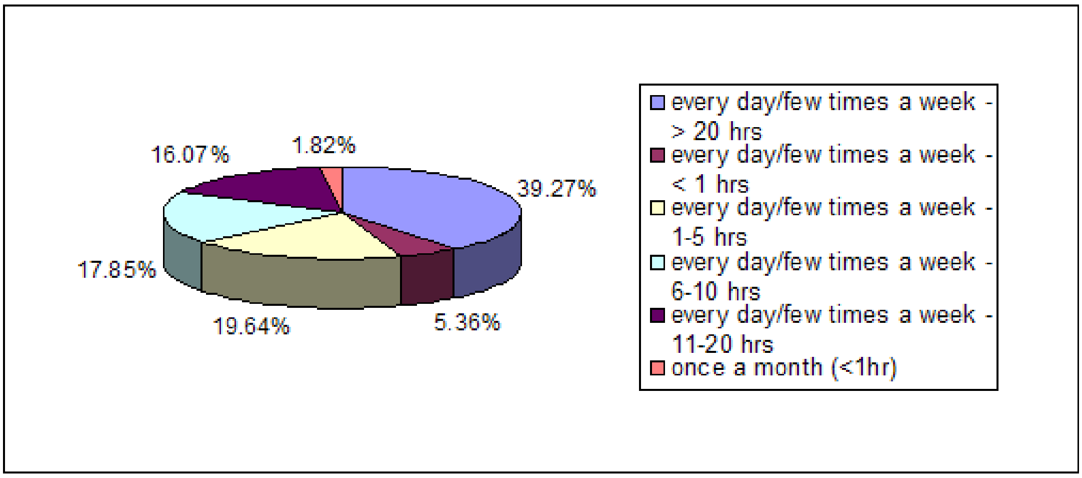


Figure 5.1: Participants' Internet Usage per Hours Weekly

Most of the participants, 98%, reported that they use the Internet daily or a few times a week with 40% of them more than twenty hours, only 5% less than an hour, 20% for 1-5 hours, 18% for 6-10 hours and the rest between 11 and 20 hours. 84% of them described that they use the Web for business/work, 87% for email/chat, 93% for special interests, and 75% for online purchasing. Figures 5.1 and 5.2 shows for how long users described they use the Internet and for what purpose. We can see that even if most of the sample reported to use the Internet daily and more than 20 hours, all users were familiar with all kinds of browsing (business/chatting/purchasing) which made our sample more generalisable.

5.1.3 Procedure

Each participant accessed the experiment in their own time. The study was hosted on the Human Centred Web Lab[1] website. Being closed, the study has been moved from its original place[2] to the ViCRAM Closed Studies[3] space for demonstration purposes only.

[1]HCW Lab - http://hcw.cs.manchester.ac.uk
[2]http://hcw.cs.manchester.ac.uk/research/vicram/studies/ranking/
[3]http://hcw.cs.manchester.ac.uk/research/vicram/studies/closed/ranking/

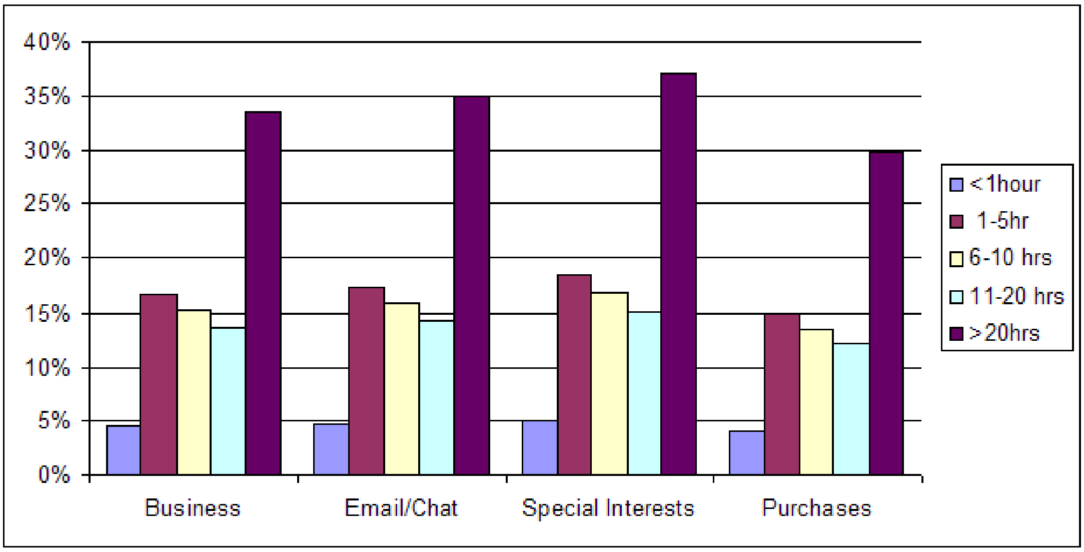


Figure 5.2: Participants' Purpose of Browsing per Hours

After reading about the experiment and consenting to take part (Figure G.1) the experiment consisted of three parts. The first part (Figure F.2) included demographic questions such as age, sex and browsing familiarity. The main study was conducted in the second part, in which after a short introduction about the procedure of the experiment (Figure G.2) the participant looked at an image for 7 seconds which then automatically changed to the ranking questions page. The participants (Figure G.3) had to give a ranking score for the stimuli they had just looked at with respect to its visual complexity and a set of five aesthetic characteristics. They also had to state whether they were familiar with the page. After looking at the randomized sequence of the 30 images, the participants had to look at the counterbalanced sequence of what they just saw and rank the images again. The third part of the evaluation (Figure G.4) was a set of feedback questions which asked participants about visual complexity. The evaluation was concluded with a 'Thank You' message (Figure F.5).

The aesthetic terms used in this evaluation were selected from a set of assessed terms examined in [Lavie & Tractinsky 2004]. In a series of studies using exploratory and confirmatory factor analyses, the authors in [Lavie & Tractinsky 2004] found that users' perceptions consist of two main dimensions, which they termed "classical aesthetics" and "expressive aesthetics". We selected five terms (and their antonyms) that were used in [Lavie & Tractinsky 2004] and can describe the aesthetic dimensions of the page: (1) Cluttered - Clean, (2) Boring -

Interesting, (3) Disorganised - Organised, (4) Confusing - Clear and (5) Ugly - Beautiful. For the data analysis, the right hand side terms were used, that is: clean, interesting, organised, clear and beautiful (Figure G.3).

5.1.4 Materials

The pages participants viewed were selected to be representative of sectors such as public information, business, academic, entertainment, leisure, Web services including search engines, and personal home pages [Amitay *et al.* 2003]. Pages were similar to a subset of the twenty pages used during our previous empirical studies and were also part of the Alexa UK top 100 Websites[4]. Pages ranged from visually simple to complex based on our previous conclusions (see Chapter 4). Selecting visually simple pages was found to be quite difficult from the Alexa list, so two sites from the University of Manchester were selected as it was a familiar (and simple) page for most of our expected participants. Table 5.1 lists the thirty Web pages selected as the set of stimuli for the evaluation and screenshots can be found in Appendix G. In order to prevent pages from changing and avoid any bias in the data, screenshots were taken of all the pages on the same day (in close time), from the same monitor and all had the same size of 1260 x 885 pixels.

5.1.5 Variable Identification

In order to understand how the structural elements of a Web page can determine the level of visual complexity, each of the stimuli used for the evaluation was analyzed and the following variables were identified: menus, images, words, top-left-corners and links. These variables are defined as follows:

Menus - A series of horizontal or vertical list of links that are grouped together in an obvious way. That is, the list of links might be divided by a line, grouped in a box, or as a section surrounded by white background. Menus are usually horizontal on the top and bottom of the page, or vertical on the left hand side of the page.

Images - Any images on the page, including advertisements, animations, logos and decorative images.

Words - Text used to present any type of information on the page. For the analysis all words within the screenshot of the page were counted, including text

[4]Alexa UK Top 100: November 2007

Table 5.1: Pages used during Complexity and Aesthetics Experiment and their Visual Complexity level

ID	Page Name	Website URL	Visual Complexity
1	Amazon UK	http://amazon.co.uk/	Complex
2	AnnoteaProject	http://www.w3.org/2001/Annotea/	Simple
3	AutoTrader	http://www.autotrader.co.uk	Complex
4	BBC UK News	http://news.bbc.co.uk/	Complex
5	BBC UK	http://www.bbc.co.uk	Medium
6	BloggerPostHQ	http://blog.last.fm/2007/08/29/audio-fingerprinting-for-clean-metadata	Simple
7	BloggerPostDE	http://www.agenturblog.de/	Simple
8	Blogger Dashboard	http://www.blogger.com/home	Medium
9	Delicious	http://del.icio.us	Medium
10	Ebay	http://www.ebay.co.uk/	Complex
11	Firefox	http://www.mozilla.com/en-US/	Simple
12	Flickr	http://flickr.com/	Simple
13	GoogleSearch	http://www.google.co.uk/search?hl=en&q=manchester&btnG=Google+Search&meta=	Medium
14	GumTree	http://www.gumtree.com	Medium
15	IMDB	http://imdb.com/	Complex
16	InvisionFree	http://invisionfree.com/	Medium
17	Job Centre	http://www.jobcentreplus.gov.uk/JCP/index.html	Medium
18	MegaUpload	http://www.megaupload.com	Complex
19	MySpace	http://www.myspace.com/	Complex
20	Orkut	http://www.orkut.com	Simple
21	Rapidshare	http://www.rapidshare.com	Simple
22	Rightmove	http://www.rightmove.co.uk	Medium
23	StudentNet	http://www.studentnet.manchester.ac.uk/	Medium
24	StudentNet SelfService	http://www.studentnet.manchester.ac.uk/selfservice/	Simple
25	WAI	http://www.w3.org/WAI/	Medium
26	Wiki Result	http://en.wikipedia.org/wiki/Wiki	Medium
27	Wikipedia	http://wikipedia.org/	Simple
28	Yahoo UK	http://www.yahoo.co.uk	Complex
29	Yell	http://www.yell.com/ucs/HomePageAction.do	Simple
30	YouTube	http://youtube.com/	Complex

from menu lists and within images.

Links - The number of links on each page was counted based on the visibility of the link. This included links that are underlined, distinguished with colours (such as blue when there is a lot of text or pale colour when the page is simple), within a list as a menu or surrounded by different colours (most commonly white).

Top Left Corner (TLC) - A page is separated in various sections and subjects. This visual distinction is made with the use of colours, tables, lines and spacing. In order to identify the number of different sections a page is organised into, we create a page chunk rendering which is the representation of the overall Web page layout without any visual elements. Chapter 4.4 described in detail the procedure followed to create a chunk rendering of a page. During this experiment and for the rest of the project we decided to use only the TLC variable. The reason was that after examining the source code of Web pages in relation with the chunking classification we found that TLC could be determined algorithmically in a closer approximation than the rest of the variables. This would later enable us to programmatically determine the number of TLCs for each page which could be used in any models that we would reveal through this study.

5.1.6 Hypotheses

To support our aim, the evaluation was based on the following hypotheses:

H1 The number of menus, images, visible links, words and TLCs a Web page has, is positively related with the page's level of visual complexity.

H2 User perception of Web page aesthetic characteristics with respect to organization, clarity, cleanliness, interestingness and beautifulness, is linearly related with the number of menus, images, visible links, words and TLCs a Web page has.

H3 User perception of Web page visual complexity is negatively related with the aesthetic qualities it presents. That is, the more organised, clear, clean, interesting and beautiful a Web page is the less visually complex it is perceived as by the user.

5.2 Results

The data collected during this evaluation are available to be downloaded from the experimental folder linked to the technical report that describes this study and is available online (see page 265). Participants' raw data submitted during the study are in the 'surveydata.txt' file in the associate folder. The rest of the associated files are Excel and SPSS files used for the data analysis. Terminology of various statistical terms used throughout the analysis can be found in Appendix A.

5.2.1 Score Correlation

As discussed in Section 5.1.1, participants ranked each image twice in a counterbalanced sequence. A reliability test between each score was conducted in order to check for consistency of the participants' scores, which was achieved by running a bivariate correlation on all the scores. It is important to note that a correlation coefficient (ϱ) of larger than .3 has a medium effect, larger than .5 a large effect and the closer to 1 the better and stronger the correlation is. The correlation coefficient determines the strength of relationship between two variables. Whether a correlation is statistically significant is determined by the p value. If this value is less than .05 and the closer to 0 the better the correlation is. These are discussed further in [Field 2005].

5.2.1.1 Complexity Scores

The complexity ranking scores given for each stimulus the first time was found to be significantly correlated with the ranking scores given the second time, with Pearson correlation coefficients of $r > .48$ and significant at $p < .001$. Image with ID26-WikiResult page had the only smaller correlation value of $r > .32$ but still significant at $p < .01$. The small effect could be caused due to the user's possible familiarity with the page. This test showed that the complexity scores given by the participants the first time were significantly correlated and had large effect with the scores given on the same stimuli the second time. Also, a strong correlation was determined between the mean values of the scores provided for each page between each version of the evaluation and their mean (ScoreA, ScoreB and AverageScore) with $r > .97$, $p < .001$.

The time that each participant spent to give an answer about the visual complexity level of the image was also recorded. However, no significant correlation was determined between the reaction times and the complexity scores given for each page. We noticed that there were some outliers in the reaction time data. That is, some times participants spent even 5-10 minutes to give an answer. In those cases we removed those outliers and rerun the analysis which did not reveal any significance, concluding that the time a participant needed to give a score on the visual complexity level of a page was not significantly related with the page's given score of visual complexity.

5.2.1.2 Aesthetic Terms Scores

The scores given the first time for each page per Aesthetic term (Clean, Interesting, Organised, Clear and Beautiful) were found to be significantly correlated with the scores given on the second time with Pearson correlation coefficients of $r > .45$ at $p < .001$.

The significant correlation between the same scores along with the large values of the Pearson correlation showed that participants were consistent with their rankings. For the rest of the analysis and data report the mean values for the rankings are used unless stated otherwise.

5.2.2 Visual Complexity

Participants ranked the images based on their visual complexity by giving a number from 1 to 10, with 1 being the visually simplest and 10 the most complex. The term Scores is used for the visual complexity ranking score that each participant assigned to the images.

5.2.2.1 Ranking Order

Table 5.2 lists the mean values of the scores given for each page in ascending order of the complexity mean value along with the preassigned level of visual complexity. Figure 5.3 shows the Web pages in ascending order of the visual complexity score that were ranked along with the complexity level preassigned using the framework defined in Chapter 4 and in Michailidou [2005]. As the graph shows, pages preassinged as visually simple received lower rankings than the visually complex pages. This can be used as an initial validation of our framework and therefore our assumptions.

Table 5.2: Visual Complexity Scores

IMAGEID	Complexity	Mean A	Mode A	Mean B	Mode B	Average Score
20	S	1.47	0	1.75	1	1.61
24	S	2.07	1	2.09	3	2.08
29	S	2.18	2	2.22	1	2.20
27	S	2.35	1	2.85	2	2.60
12	S	3.11	4	3.00	3	3.05
11	S	3.18	2	3.00	1	3.09
17	M	3.47	2	3.05	2	3.26
6	S	3.31	1	3.53	2	3.42
2	S	3.07	1	3.78	2	3.43
13	M	3.56	2	4.00	2	3.78
23	M	3.84	2	4.00	4	3.92
7	S	4.20	2	3.73	4	3.96
21	S	4.16	2	3.84	2	4.00
22	M	4.02	4	4.35	4	4.18
26	M	4.42	2	4.51	6	4.46
8	M	4.64	7	4.85	5	4.75
1	C	4.60	5	5.58	8	5.09
10	C	5.36	6	5.45	7	5.41
16	M	5.33	7	5.55	6	5.44
30	C	5.67	7	5.27	6	5.47
5	M	5.69	8	5.55	6	5.62
9	M	5.71	7	5.55	8	5.63
14	M	5.69	4	5.84	6	5.76
4	C	5.78	8	5.84	6	5.81
25	M	6.00	6	6.00	8	6.00
15	C	5.65	6	6.53	7	6.09
28	C	6.15	5	6.27	7	6.21
19	C	6.05	5	6.44	8	6.25
18	C	6.60	8	6.27	7	6.44
3	C	6.93	7	6.78	7	6.85

5.2.2.2 Multiple Regression

In order to understand how the structural elements of a Web page can determine the level of visual complexity, each stimulus used for the evaluation was analyzed and the following variables were identified: menus, images, words, TLC and links (the variables are defined in Section 5.1.5). Table 5.3 lists the structural elements and their respective number for each page. Using as dependent variable the AverageScore and as predictors the number of menus, words, TLC and links in an enter regression method, a significant model emerged ($F_{5,24}$=12.098, p<.001) with a strong fit of $R^2{}_{adj}$=0.66. However, only the effect of TLC (t(24)=3.028, p <.01), words (t(24)=2.917, p <.005) and images (t(24)=2.234, p <.05) were significant, leaving menus and links not included in the model.

The value of R^2 describes how much of the variance in the Visual Complexity Score is accounted for by the regression model from the sample, where the $R^2{}_{adj}$

Table 5.3: Web Page Structural Elements

PageID	**Familiarity**	**Mean A**	**Mean B**	**Mean**	**Menus**	**Image**	**Words**	**TLC**	**Links**
1	1	4.60	5.58	5.09	4	14	334	28	63
2	0	3.07	3.78	3.43	1	1	464	11	46
3	0	6.93	6.78	6.85	3	13	357	22	58
4	1	5.78	5.84	5.81	2	13	530	29	84
5	1	5.69	5.55	5.62	1	7	353	20	112
6	0	3.31	3.53	3.42	1	1	345	8	8
7	0	4.20	3.73	3.96	0	5	281	4	10
8	0	4.64	4.85	4.75	2	1	331	10	43
9	0	5.71	5.55	5.63	1	8	341	17	44
10	1	5.36	5.45	5.41	3	6	285	17	50
11	1	3.18	3.00	3.09	2	7	200	8	19
12	0	3.11	3.00	3.05	0	5	130	11	22
13	1	3.56	4.00	3.78	1	3	401	16	44
14	0	5.69	5.84	5.76	2	2	360	20	31
15	1	5.65	6.53	6.09	4	7	789	19	96
16	0	5.33	5.55	5.44	2	0	400	11	31
17	0	3.47	3.05	3.26	1	2	116	10	19
18	0	6.60	6.27	6.44	1	24	125	21	9
19	1	6.05	6.44	6.25	4	13	272	17	56
20	0	1.47	1.75	1.61	1	1	8	4	6
21	0	4.16	3.84	4.00	1	2	271	4	12
22	0	4.02	4.35	4.18	5	7	240	15	34
23	1	3.84	4.00	3.92	2	8	246	17	32
24	1	2.07	2.09	2.08	3	1	175	7	24
25	0	6.00	6.00	6.00	2	1	550	17	42
26	1	4.42	4.51	4.46	6	2	450	14	76
27	1	2.35	2.85	2.60	1	7	180	6	110
28	1	6.15	6.27	6.21	5	35	370	19	83
29	0	2.18	2.22	2.20	2	6	85	5	12
30	1	5.67	5.27	5.47	2	17	267	17	42

how much variance in the dependent variable (here Visual Complexity Score) would be accounted for if the model had been derived from the population from which the sample was taken (here the Web pages) [Field 2005]. In other words, the $R^2{}_{adj}$ gives some idea of how well a model generalizes and ideally one would like its value to be the same, or very close to, the value of R^2. Using the mean values from ScoreB data the best model with the strongest significance in the coefficients emerged.

An initial equation that revealed from the study's data analysis using the enter regression method, which predicts the VCS of a page is:

$$VCS = 1.759 + 0.103(TLC) + 0.004(Words) + 0.061(Images) \qquad (5.1)$$

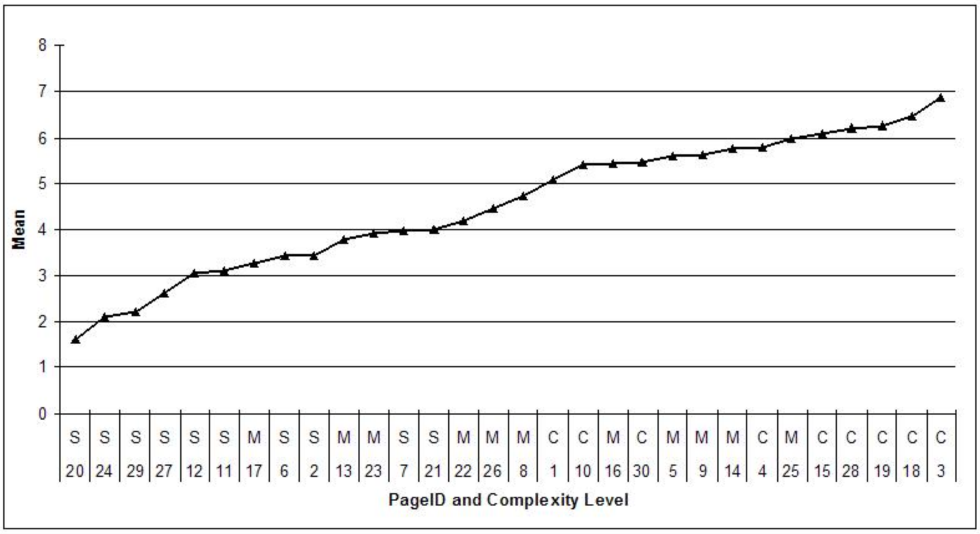


Figure 5.3: Visual Complexity Web Page Ranking Order

Then, using the significant variables as the predictors a second series of regression analysis was performed, with three different dependent variables: ScoreA, ScoreB and AverageScore. Even if these scores were highly and significantly correlated, testing what models each produced provided a more complete understanding. During our last evaluation described in Chapter 4.4 it was shown that the number of TLCs a page had, was significantly related with the visual complexity level of the page. Therefore, the TLC variable was used as the first variable in a stepwise regression analysis. Tables 5.4 and 5.5 show the Model's Summary and Coefficients that each analysis produced. Model 1 refers to a regression using only the TLC predictor and Model 2 to the regression model with all the significant predictors (TLC, Words and Images).

Table 5.4: Stepwise Regression - Visual Complexity Score (VCS) Model Summary

Dependent Var.	**Model**	**R**	R^2	**Adj.** R^2	**SE Est.**	ANOVA **F-ratio***	**Durbin-Watson**
AvScore	1	.774	0.599	0.585	0.938	41.809	
	2	.837	0.701	0.666	0.841	20.292	2.062
ScoreA	1	.738	0.545	0.528	1.010	33.487	
	2	.796	0.634	0.592	0.940	15.007	2.010
ScoreB	1	.798	0.637	0.624	0.896	49.169	
	2	.867	0.752	0.724	0.768	26.342	2.097

All Significant at $p < .0001$;
Model 1 Predictors: (Constant), TLC;
Model 2 Predictors: (Constant), TLC, words, images

Checking for multicollinearity during the data analysis is very important for

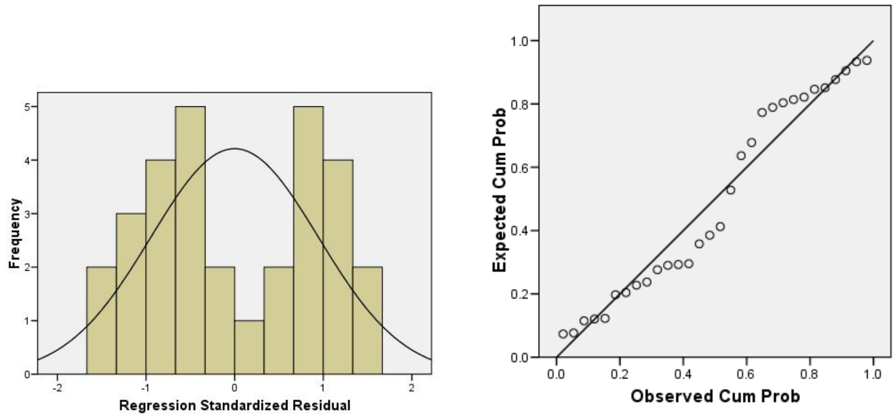
Table 5.5: Stepwise Regression - Coefficients

	Average Score				Score A				Score B			
	B	SE B	β	t	B	SE B	β	t	B	SE B	β	t
Model 1												
Constant	2.186	0.401		5.456	2.218	0.432		5.138	2.155	0.383		5.630
TLC	0.166	0.026	0.774	6.466[c]	0.160	0.028	0.738	5.787[c]	0.172	0.024	0.798	7.012[c]
Model 2												
Constant	1.743	0.399		4.366	1.835	0.446		4.111	1.651	0.365		4.526
TLC	0.097	0.033	0.452	2.925[a]	0.093	0.037	0.430	2.514[b]	0.101	0.030	0.468	3.328[a]
Words	0.053	0.026	0.279	2.067[b]	0.056	0.029	0.295	1.974[b]	0.049	0.023	0.259	2.110[b]
Images	0.003	0.001	0.361	2.743[b]	0.003	0.001	0.317	2.174[b]	0.004	0.001	0.401	3.344[a]

a. $p < .01$; b. $p < .05$; c. $p < .001$

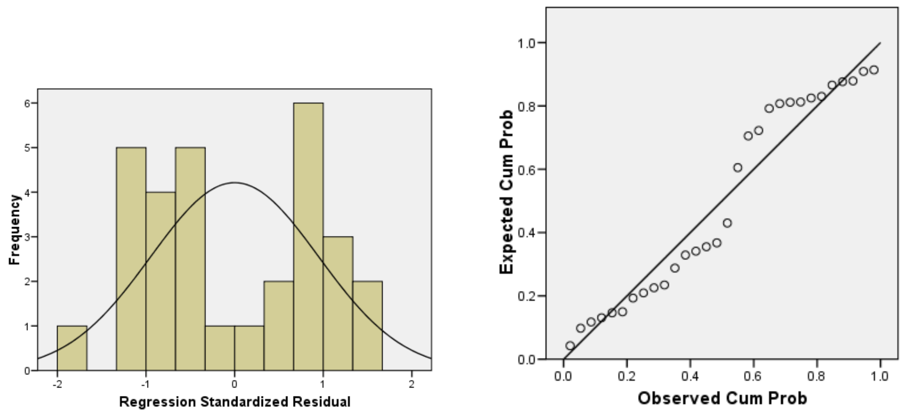
validating the effect of a regression model. Multicollinearity exists when there is a strong correlation between two or more predictors in a regression model [Field 2005]. Such high correlations cause problems when trying to draw inferences about the relative contribution of each predictor variable to the success of the model. A way of identifying multicollinearity is to scan a correlation matrix of all of the predictor variables and see if any correlate very highly ($> .8$). Another diagnostic is the variance inflation factor (VIF). The VIF indicates whether a predictor has a strong linear relationship with the other predictor(s). A rule-of-thumb is that the sum of all VIF values should be less than 10. If it is greater than 10, that means that the model's independent variables are highly correlated with one another. As a result, the variable with the large VIF should be removed from the analysis and perform a new regression. For this study's analysis, both collinearity matrix and VIF values fitted under the correct values. The model with the ScoreB as dependent variable produced the smallest values, which made it again a better (but not less significant) predictive model.

Residuals are known as the difference between the values of the outcome predicted by the model and the values of the outcome observed in the sample. These residuals effectively represent the error present in the model, usually caused by outliers [Field 2005]. A way of checking how well the regression model fits is to check the normality of the residuals (the standardized residuals) which can be done by looking at the histogram and normal probability plots also shown in Figures 5.4, 5.5 and 5.6. The histogram should look like a normal distribution (a bell-shaped curve) and Figure 5.6 seems to have the better fit of normality. The straight line in the probability plots represents a normal distribution and the points the observed residuals and in a perfectly normally distributed data set, all points will lie on the line. Again, the model using the ScoreB data set has the best normality within the rest sets (AverageScore and ScoreA).



(a) Histogram (b) Normal P-P

Figure 5.4: Average Score - Regression Model Normality Check

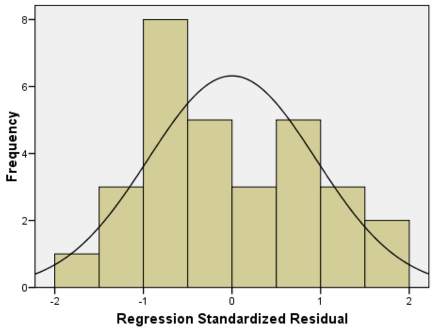


(a) Histogram (b) Normal P-P

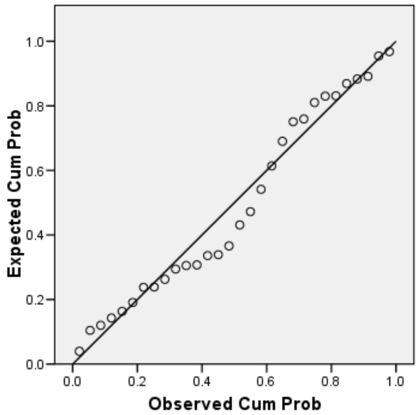
Figure 5.5: ScoreA - Regression Model Normality Check

After checking for the best fit of the data and their collinearity, using TLC as the first factor in a stepwise regression model, the analysis revealed three possible equation based on AverageScore, ScoreA and ScoreB respectively:

$$VCS_{av} = 1.743 + 0.097(TLC) + 0.053(Words) + 0.003(Images) \tag{5.2}$$



(a) Histogram



(b) Normal P-P

Figure 5.6: ScoreB - Regression Model Normality Check

$$VCS_A = 1.835 + 0.093(TLC) + 0.056(Words) + 0.003(Images) \quad (5.3)$$

$$VCS_B = 1.651 + 0.101(TLC) + 0.049(Words) + 0.004(Images) \quad (5.4)$$

Table 5.6: Correlation between Independent Variables (structural elements), Visual Complexity and Aesthetic Scores

Variables	Complexity	Clean	Interesting	Organised	Clear	Beautiful
Menus	.341	-.302	.265	-.016	-.095	-.048
Images	.510[a]	-.423[b]	.294	-.249	-.318	.047
Words	.576[c]	-.585[a]	.020	-.384[b]	-.451[b]	-.275
TLC	.774[c]	-.704[a]	.277	-.345	-.462[c]	-.200
Links	.417[b]	-.342	.519[c]	.004	-.095	.174
Familiarity	.116	.005	.733[a]	.382[b]	.285	.487[c]

a. $p < .01$; b. $p < .05$; c. $p < .001$

For the independent variables Menus and Links that were determined as not significant for the regression model a correlation was conducted between the variables and the three versions of the complexity scores. Visual complexity was found to be positively related with the number of menus and links a Web page has. The number of menus was significantly correlated with only the second

Table 5.7: Correlation between Independent Variable (structural elements) and Visual Complexity Scores (mean values)

Variables	Score A	Score B	Average Score
Menus	0.291	0.387^b	0.341
Images	0.515^a	0.498^a	0.510^a
Words	0.521^a	0.624^a	0.576^c
TLC	0.738^c	0.798^c	0.774^c
Links	0.362^b	0.465^a	0.417^b

a. $p < .01$; b. $p < .05$; c. $p < .001$

complexity score (ScoreB), r = .4 at $p < .05$. On the other hand, the visual complexity score was significantly correlated with the number of links a page has with all the three scores: ScoreA with r = .36, $p < .05$; ScoreB with r = .47, $p < .01$; AvScore with r = .42, $p < .05$. Menus and links were also significantly related with each other (r = .4 at $p < .05$). A value of r between .3 and .5 has a medium effect which might be the reason of excluded from the models by the regression test. The second column of Table 5.6 lists the correlations of all independent variables with the average complexity score and Table 5.7 lists the correlations with all the score versions. We notice that the TLC has the highest correlation with the visual complexity of a page. This is an indication on how the number of different sections a page is grouped into and the overall page layout affect user perception which validates our initial assumptions.

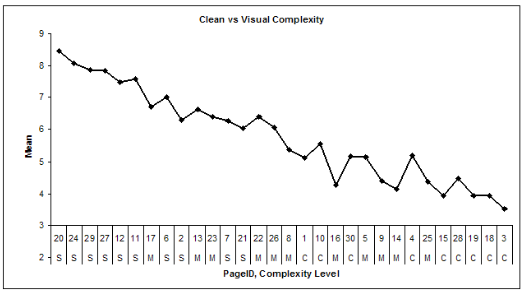
5.2.3 Aesthetics

The second part of the experiment was to ask participants to rate each page with respect to its aesthetic appearance. Figure G.3 shows the characteristic scales that each participant had to fill and so rank each stimuli. Participants' rankings started from the left hand side as 1 and the maximum was 10. For example, a very cluttered page was given a 1 and a very clean page a 10.

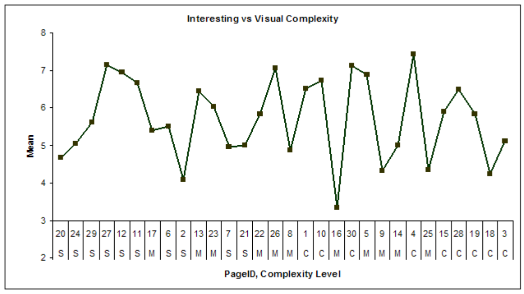
5.2.3.1 Correlation

Table G.12 lists the mean values of each characteristic per page. Again, the mean values from each score version are listed along with the mean value from both versions. As described in the previous sections, participants' scores were significantly and highly correlated but, again, for this analysis all three versions per characteristic are used.

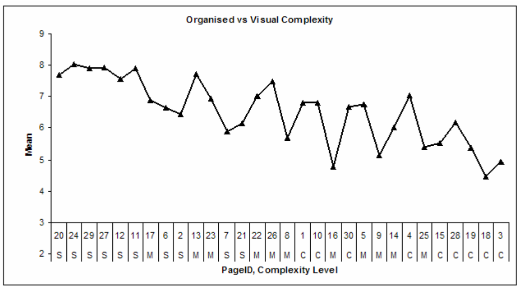
Figure 5.7 shows each aesthetic characteristic's average value with respect to the ranking order of the Web pages based on the Visual Complexity Score each received. Cleanness of the page is shown as negatively related with the complexity level of the page. That means that the less visually complex the page was ranked as, the more clean it was perceived by the participants. Organisation and clearness of the page were also negatively related with the visual complexity of the page but not as linear as clear was. Overall, the terms clear, clean, organised and beautiful have a negative slope which shows the negative relationship between the level of complexity. Interestingness did not reveal any relationship with the complexity level of the page.



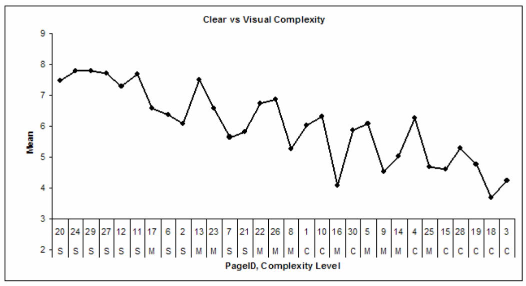
(a) Clean



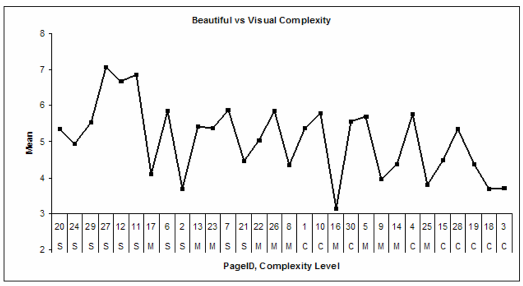
(b) Interesting



(c) Organised



(d) Clear



(e) Beautiful

Figure 5.7: Graphs of Aesthetic Characteristics with respect to Visual Complexity Ranking Order

Correlation analysis was also conducted between visual complexity and the aesthetic characteristics. A negative and of high significance relation ($r > -.5$ at $p < .01$) between the visual complexity score of the page and the following aesthetic characteristics was determined: clean, organised, clear and beautiful. That means that the less visually complex the page was ranked as, the more clean, organised, clear and beautiful it was perceived by the participants. Interestingness did not reveal any relationship with the complexity level of the page. Table 5.8 shows the correlation coefficients for all the terms between the visual complexity score and between each other. Between all of the characteristics, clearness had the higher relation with visual complexity, $r = -.85$ at $p < .001$. Interestingness was not related with the visual complexity which might be caused by participants' familiarity with the page, which was the reason of including a familiarity question. Indeed, the last column of Table 5.8 shows that participants' familiarity with the page was highly and significantly related with their interestingness rankings ($r = .733$).

Table 5.8: Correlation between Aesthetic Characteristics and Visual Complexity Scores

	Complexity	Clean	Interesting	Organised	Clear	Beautiful	Familiarity
Complexity	1	-.974[c]	-.036	-.775[c]	-.849[c]	-.463[a]	.116
Clean	-.974[c]	1	.214	.861[c]	.923[c]	.616[c]	.005
Interesting	-.036	.214	1	.594[c]	.507[c]	.801[c]	.733[c]
Organised	-.775[c]	.861[c]	.594[c]	1	.986[c]	.768[c]	.382[b]
Clear	-.849[c]	.923[c]	.507[a]	.986[c]	1	.743[c]	.285
Beautiful	-.463[a]	.616[c]	.801[c]	.768[c]	.743[c]	1	.487[a]
Familiarity	.116	.005	.733[c]	.382[b]	.285	.487[a]	1

a. $p < .01$; b. $p < .05$; c. $p < .001$

5.2.3.2 Regression Analysis

Using the structural and familiarity variables listed in Table 5.3 a series of analysis using stepwise regression were conducted. By identifying which variables play an important role in each model, each of the characteristic's regression analysis was repeated. It is important to note here that the familiarity with a page was the only variable that was included in all the estimation models, something that did not happen with the visual complexity score models. Regression models were identified for: clean using TLC, words, familiarity and images; interesting using familiarity, links and words; organised using familiarity, words and images;

clear using TLC, familiarity, words and images; and beautiful using familiarity and TLC. Tables G.1 - G.10 list the regression models for each characteristic respectively, which are described in more detail below.

For a good model, one should look at the value of R^2 and Adj. R^2 and check how close their values are with each other. A value of R that is higher than .5 has a large effect and hence the model is more effective. Also, the larger the F value returned by the analysis, the better the model it is (if it is significant at $p < .05$). In addition, the Durbin-Watson value should be as close to 2 as possible and in general between 1 and 3. As described in previous sections, multicollinearity is really important in regression analysis and looking at the VIF values and correlation coefficients between the independent values are some of the ways of checking for it. Throughout all this experiment's analysis, the first score that participants gave derived the worst (but some significant) models and are not discussed further. However, the regression results can be found in the associated folder of the study's technical report (see page 265).

Using familiarity in these prediction models results in subjective and less stronger models because they require the user's input. Therefore, the stepwise regression test was repeated without the familiarity variable for every characteristic to get a better idea of how the aesthetic characteristics of a page can be predicted by its structural elements only.

Clean

Table G.1 lists the values of the effect that each model returned. Looking at the last model (Model 4) of each score version, it is noted that R^2 and F-ratio values were close to each other. Both versions' models were significant with Durbin-Watson value close to 2. CleanB's values were slightly but not more significant better. So, taking the coefficient values of the model 4 (B-values) for CleanB and AverageClean score versions (see Table G.2) the equations that the data predict the cleannes of the page are:

$$Clean_B = 8.407 - 0.095(TLC) - 0.004(Words) + 1.111(Familiarity) - 0.059(Images) \quad (5.5)$$

$$Clean_{av.} = 8.342 - 0.094(TLC) - 0.004(Words) + 1.069(Familiarity) - 0.060(Images) \quad (5.6)$$

Running a stepwise regression without including the familiarity variable, a model revealed that included the TLC and words variables by both CleanB and AverageClean score values. For CleanA score values a linear model revealed with TLC its predictor. These models are also listed as Model 1 and 2 in Tables G.1 and G.2. So the new equations that do not require user's input to predict the cleannes of a Web page based on this study's data are:

$$Clean_B = 8.374 - 0.115(TLC) - 0.003(Words) \quad (5.7)$$

$$Clean_{av.} = 8.307 - 0.115(TLC) - 0.003(Words) \quad (5.8)$$

Interesting

Table G.3 lists the values of the effect that each of the model that predict interestingness of a page returned based on the data. Looking at the last model of each score version, it is noted that R^2 and F-ratio values were close to each other with Durbin-Watson value not very close to 2 but still valid. Taking the coefficient values of the model 1 (B-values) for each score version that the data analysis revealed (see Table G.4), the new equations that estimate the level of interestingness of a page are:

$$Interesting_B = 5.060 + 1.545(Familiarity) \quad (5.9)$$

$$Interesting_{av.} = 4.955 + 1.560(Familiarity) \quad (5.10)$$

Running a stepwise regression without including the familiarity variable a model is returned with links and words the predictors for both InterestingB and AverageInteresting score version values. These models are also listed as Model 2 and 3 in Tables G.3 and G.4. The new equations that based on the experiment's analysis estimate the interestingness level of the page are:

$$Interesting_B = 5.536 + 0.026(Links) - 0.003(Words) \quad (5.11)$$

$$Interesting_{av.} = 5.342 + 0.026(Links) - 0.003(Words) \tag{5.12}$$

Organized

Table G.5 lists the values of the effect that the organized prediction models revealed. Again, OrganisedB's models gave slightly larger values but not more significant than AvOrganised's model. Taking the coefficient values of the model 3 (B-values) for each score version the following models (see Table G.6) are returned:

$$Organised_B = 7.595 - 0.004(Words) + 1.406(Familiarity) - 0.067(Images) \tag{5.13}$$

$$Organised_{av.} = 7.459 - 0.004(Words) + 1.354(Familiarity) - 0.064(Images) \tag{5.14}$$

Running a stepwise regression test without the familiarity variable, linear models are return for both OrganisedB and AvOrganised score version with the words variable their predictor. These models are also listed as Model 1 in Tables G.5 and G.6. The new linear equations that based on the experiment's data can predict the organization of the page based on its structural elements only are:

$$Organised_B = 7.423 - 0.003(Words) \tag{5.15}$$

$$Organised_{av.} = 7.297 - 0.003(Words) \tag{5.16}$$

Clear

Table G.7 lists the values of the effect that each of the clearness model shows. ClearB values' model are not all the same as AverageClear's one. This is because AverageClear values are affected by the ClearA scores. However, the models that were produced using ClearA did not have large effect ($R < .5$) and were not listed here. Taking the coefficient values of the model 6 (B-values) for each score version the following models (see Table G.8) are formulated:

$$Clear_B = 7.595 + 1.406(Words) - 0.004(Familiarity) - 0.067(Images) \quad (5.17)$$

$$Clear_{av.} = 7.365 - 0.005(Words) + 1.417(Familiarity) - 0.082(Images) \quad (5.18)$$

Running a stepwise regression analysis without the familiarity variable a linear model that uses only Words in ClearB and TLC in AvClear is produced. These models are also listed as Model 1 and 3 respectively in Tables G.7 and G.8 and their respective equations are:

$$Clear_B = 7.423 - 0.003(Words) \quad (5.19)$$

$$Clear_{av.} = 7.172 - 0.082(TLC) \quad (5.20)$$

One can notice that equations 5.15 and 5.19 are exactly the same even if they predict different aesthetic characteristics. This could happen for various reasons, such as the possibility that the equations do not correctly predict the aesthetic characteristic of the page or that participants associated the organization of the page with the clearness to be closely defined.

Beautiful

Table G.9 lists the values of the effect that each model produced. BeautifulA or BeautifulB's models produce models that are not as effective as the Average-Beautiful's ones and are not described here. Taking the coefficient values of the model 2 (B-values) from the AverageBeautiful version scores (see Table G.10), the equation that based on the data analysis predict the beautifulness of a Web page is:

$$Beautiful_{av.} = 5.358 + 1.267(Familiarity) - 0.064(TLC) \quad (5.21)$$

Running a stepwise regression test without the familiarity variable did not reveal any model. This could mean that the perception of beautifulness is dependent on the user's familiarity with the Web page rather than its appearance and structural layout.

5.2.3.3 Correlation

In order to investigate the relationship between the variables that could not fit in any of the models, a correlation analysis was performed between the aesthetic terms and the structural elements. Table G.11 lists all independent variables and their respective correlation coefficient between them and the aesthetic characteristics. The variables that were highly and significantly related with the aesthetics terms were also the ones that were included in the regression models described in the previous section. However, it is important to note that Cleanness was the only characteristic that was significantly correlated with the most of the structural elements examined. Also, Beautifulness was not significantly related with any of the structural elements.

5.2.4 Feedback Comments

The last part of the evaluation was introduced to directly ask participants' opinion of what a visually simple and complex page is. First, the participants were asked whether they agree or disagree with a series of statements shown in Figure G.4. As Figure 5.8 shows, most of the participants agreed that they "can easily find information on a visually complex page" but they strongly agreed that they "can easily find information on a visually simple Web page". The majority of them also agreed with the statement that they tend to "spend more time interacting with a visually complex page than a simpler one", but more data will be needed to collected in order to explain the reason of this. We assume that this happens due to the amount of information a visually complex page has, which is also shown from their agreement on the fact that a visually complex page has a lot of information to choose from.

It is important to note that most of the participants disagreed with the statement that "a visually simple page is boring". However, the analysis showed that familiarity with a page was the most related factor in rating a page as interesting. Also, most participants were neutral on the statement that "a visually complex page is disorganized". However, the previous analysis showed that complexity of the page was significantly and negatively related with the organization of a page.

Participants were also asked to define in their own words a visually simple and

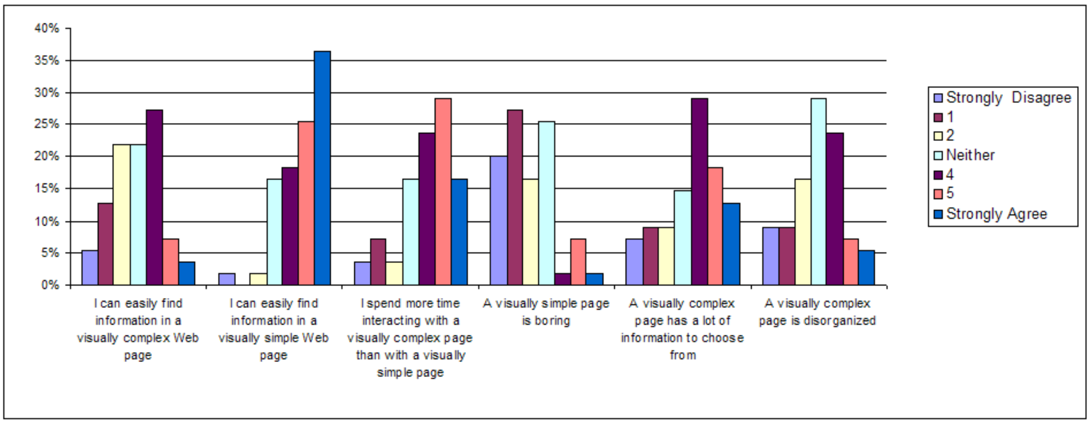


Figure 5.8: Participants Agreement on Feedback Statements

complex page. In summary, participants described a visually simple page as one with pale colours, clean, can easily find what they are looking for, covers a small screen area, has limited set of images and links. Overall, participants described a visually simple page as organised, with limited amount of different subjects and whose purpose can easily be understood. Some of the participants' comments are:

> *"It quickly becomes clear where I need to navigate based on the purpose of my visit to the website. A simple page gives me a quick impression of what the website is about / what I can achieve through the website. A simple page has a good mixture of images and text so that I am quickly guided to the specific thing i was looking for without being blinded by details."*

> *"A page that doesn't have too much on it so that it's easy to find what you want; doesn't have large blocks of text; don't have to scroll around too much to find something and where key menus or sections are clearly defined"*

> *"Uniform visual elements all treated similarly or purely textual layout"*

On the other hand, participants described a visually complex page as one with a lot of information and categories, not uniform in size and shape, has a large number of images, different colours, buttons, animations and generates an overall distraction. Some of the participants' comments are:

> *"Too much on the page especially if there are poor boundaries between sections or if everything is written in the same way. One with large blocks of text that mean scrolling down for ages until you find what you want. A page where you can get to the same place several ways making it difficult to navigate the site properly. One which uses lots of different formatting techniques e.g. different menustabs, boxes, different fonts and colours etc"*

> *"Complicated layout, lots of graphics, lots of information."*

> *"I have to visually scan lots of different parts of the page and read lots of (especially small) text to discover where I need to navigate. Complexity is also linked to whether I need to disambiguate between areas of the page for which it is unclear if they will be able to serve the purpose of my visit to the web site."*

5.3 Discussion

The quantitative and qualitative analysis of this evaluation produced a set of conclusions and initial prediction models that can help expand our research and understand better the relationship between structural elements and users' perception of a Web page visual complexity. Table 5.9 lists the models derived through the data analysis for both Visual Complexity and the Aesthetics terms. The results of this evaluation provided robust information on how users tend to perceive visual appearance of a Web page and how designers can check whether their page will be perceived as complex, interesting, clean, organised, beautiful or clear by their clients. Our hypotheses statements are supported from the results with some limitations on the structural elements of the page.

First, **H1** stated that the number of menus, images, visible links and words a Web page has and sections it is grouped into (TLC), is positively related with the page's level of visual complexity. Only the first part of this statement was not supported by the data. That is, the experiment revealed a positive, significant and robust relationship between visual complexity of the page and its number of images, visible links, words and TLCs but not with the number of menus. In addition to their strong correlation, prediction models were identified using the number of images, words and TLCs as the variables in the equations.

Table 5.9: Prediction Models for Visual Complexity and Aesthetic Characteristics

ID		Equation	R^2
1	VisualComplexity	= 1.743 + 0.097 (TLC) + 0.053 (Words) + 0.003 (Images)	0.7
2	Clean	= 8.342 - 0.094 (TLC) - 0.004 (Words) + 1.069 (Familiarity) - 0.060 (Images)	0.72
		= 8.307 - 0.115 (TLC) - 0.003 (Words)	0.57
3	Interesting	= 4.955 + 1.560 (Familiarity)	0.54
		= 5.342 + 0.026 (Links) - 0.003 (Words)	0.37
4	Organisation	= 7.459 -0.004 (Words) + 1.354 (Familiarity) - 0.064 (Images)	0.57
		= 7.297 - 0.003 (Words)	0.15
5	Clear	= 7.365 -0.005 (Words) + 1.417 (Familiarity) - 0.082 (Images)	0.61
		= 7.172 -0.082 (TLC)	0.21
6	Beautiful	= 5.358 + 1.267 (Familiarity) - 0.064 (TLC)	0.4

TLC, Words and Images are unit variables; Familiarity is a binary variable

This conclusion supported our initial investigation into visual complexity of a Web page and the previous observation made during the pilot evaluation (Chapter 4) that provided the basis for understanding how the visual appearance of a Web page can affect users' perception. These conclusions supported our second research question of this thesis (Q2 - page 23) by providing significant correlations between Web page structural elements and visual complexity user perception. In addition, the results can act as an initial step into modeling how the visual design, and especially the combination of visual elements, of a Web page can influence a user's first impression which will lead into supporting our third research question (Q3 - page 23). Further investigation needs to be done to examine a larger set of structural elements and how their visual appearance, position and density affects user perception, usability and accessibility. The current models, though, can still be used to examine the effect of prediction.

The second hypothesis, **H2**, states that user's perception of a Web page's aesthetic characteristics is related with the number of menus, images, visible links, words and TLCs on a Web page. The evaluation supported our hypothesis for all the structural elements except for the number of menus. The results showed that structural aesthetics (clean, clear and organised) are highly related with the number of images, words and different sections (TLC) a page has. This reflects how the density and position of various elements on a page can influence user's perception of the page design.

It was also shown that user's familiarity with a page has the highest effect on perceiving a page as interesting and beautiful. These conclusions can help designers evaluate a page with respect to its perceived aesthetics which can give a sense of credibility or enhance usability.

H3 states that user perception of a Web page's visual complexity is related with the aesthetic qualities it presents. The experiment revealed that only interestingness was not related with the complexity level. That is, data analysis showed that the more organised, clear, clean and beautiful a page is, the visually simpler the page was perceived by the participants.

The results of this study demonstrate that it is possible to model user perception based on Web page structural layout. This study should act as the basis for an extended investigation into the relationship of user perception on visual complexity and aesthetics with all possible structural elements and their attributes (such as colour and size). As the data supported, the overall structural layout of a page was the most important factor in predicting user impressions. That is, the TLC variable is mainly influenced on the overall structure of the page, the number of different sections and how they are organised.

Previous research is performed on how the aesthetic perception of a page influence user impression on credibility [Robins & Holmes 2008] and therefore interaction. We assert that future work should include an examination into how user aesthetic perception affects task completion and browsing behaviour on Web pages. This can be achieved by investigating user eye movement on Web pages and how their viewing behaviour changes based on the aesthetic appearance of the page.

5.4 Conclusions

This chapter described an online ranking evaluation that investigated how users perceive visual complexity and aesthetic appearance of Web pages. Limited previous research and work showed the importance of examining visual aesthetics on Web pages and how they manipulate user perception of page credibility. Our previous studies (Chapter 4) introduced a framework that identifies visually simple and complex pages based on a set of structural elements. This study extended our previous work by examining the relation between visual complexity, aesthetic characteristics and structural layout of the page. Models were derived that can predict the visual complexity level of the page, organisation, clearness, cleanliness, beautifulness and interestingess of the page.

The results showed that visual complexity of a page is negatively related with user perception of how organised, clear, clean and beautiful a page looks. The number of links, images, words and sections (TLC) were used to derive models that predict how a user will perceive a page's aesthetic and complexity appearance. We suggest that results should be used by designers to keep the balance between aesthetic appearance of a Web page and its visual complexity. When this balance is achieved Web pages can still be aesthetically attractive but also usable and not overloaded with information for the users. The next step into this project is to use the Visual Complexity prediction model into programmatically determine the level of complexity and areas of visual complexity for Web pages. By implementing this model we can examine any deficiencies and the need for evaluating further page variables. In addition, the aesthetic characteristics prediction models are not further used because their purpose was to get a more spherical perspective of user perceptions. The results derived interesting models and we identified the need for a further detailed research which at the time of this thesis is already under process through the EiVAA[5] project.

[5]EiVAA Project - http://hcw.cs.manchester.ac.uk/research/eivaa

Chapter 6

Visual Complexity Algorithm

It is shown throughout the literature that in order to ensure effective implementation of the Web Content Accessibility Guidelines, it is necessary to have effective testing methods. As explained in Chapter 2, evaluation of the accessibility of a Website is more complex than simply validating the markup of the page. This is because the implementation of accessibility requirements is not only a matter of using valid markup, but also the appropriateness of equivalent alternatives, and organization of information on a Website [Brewer 2004b]. Effective evaluation requires a "combination of automated and manual testing by someone with expert knowledge of WCAG and of relevant testing tools and approaches" [Brewer 2004b]. Hence, a combination of automated evaluation tools, design guidelines and user testing is a better approach for validating Website accessibility.

In addition, by understanding how disable users access the Web, developers can become more expert on identifying accessibility problems [Asakawa & Itoh 1998]. However, it is impractical for Web designers to attempt to experience the same situation as impaired users [Takagi *et al.* 2004]. As Takagi *et al.* [2004] explain, disability simulation can help Web designers to experience a similar experience to having disabilities as they will be able to imagine, recognize and understand the disabled Web access experience.

Our previous work investigated the relationship between users' perception and visual complexity of a Web page and as Chapter 5 describes, prediction models were derived that can estimate the level of visual complexity of a Web page based on the number of words, images and sections a page has. We believe that these models can be used along with disability simulation tools, such as the Accessibility Designer (aDesigner) [Takagi *et al.* 2004], so that a Web page can be automatically analyzed with respect to its visual complexity.

This chapter describes the development of the ViCRAM tool, which is an implementation based on aDesginer. The complexity algorithm, derived from the complexity prediction models described in Chapter 5, provided two features. The first was to provide designers with a complexity score that determines the page's level of visual complexity. The second was to provide an overlay heatmap that highlights the areas that are most visually complex on the page. A technical evaluation of the tool was executed that revealed outstanding issues that helped to improve the tool but also that the complexity score prediction is significantly correlated with the user scores.

6.1 Accessibility Tools Framework (ACTF)

"The Accessibility Tools Framework (ACTF)[1] is an incubation phase project that is a subproject of the Eclipse Technology Project. ACTF serves as an extensible infrastructure upon which developers can build a variety of utilities that help to evaluate and enhance the accessibility of applications and content for people with disabilities. A collection of example utilities are provided which were created on top of the framework. These include compliance validation tools, assistive technology simulation applications, usability visualization tools, unit-testing utilities, and alternative accessible interfaces for applications. The ACTF componentry and the utilities are integrated into a single tooling environment on top of the Eclipse framework. The framework components function cooperatively with each other and with other Eclipse projects to provide a comprehensive development environment for creating accessible applications and content."

6.1.1 aDesigner

The Accessibility Designer (aDesigner) tool is one of the features included in the ACTF project. This tool provides a way to evaluate a page at a glance by developing a visualization feature that tries to supplement or solve accessibility problems with the current guidelines checkers. aDesigner has capabilities to visualize blind users' usability by using colors and gradations. The visualization function allows Web designers to grasp weak points in their pages, and to recognize how

[1]ACTF Project Home - http://www.eclipse.org/actf/

accessible or inaccessible their pages are [Takagi *et al.* 2004].

The aDesigner API was chosen as the basis to implement the ViCRAM tool because aDesigner uses HTML parsers and has visualization capabilities that match our objectives. One component of aDesigner was the simpleVisualizer. We extended this component to implement the ViCRAM heatmapping and complexity score functionality.

6.2 The ViCRAM Tool

Research Question 3 in Chapter 1 stated that Web page layout can be used to predict the visual complexity of a Web page. During the user evaluations conducted throughout this project it was shown that visual complexity is affected by the page structural elements and overall layout. Prediction models were derived (see Chapter 5) that, based on specific page elements, can provide an approximation of the visual complexity of a page. In order to answer this research question we first needed to implement these models to automatically provide the complexity information. This implementation consisted of two parts. The first part was an algorithm that determined the level of visual complexity of a given page (Complexity Algorithm). The second part was an algorithm that visualized a page by highlighting areas based on their visual complexity level (Visualization Algorithm).

Both algorithms were based on the complexity model previously derived by the user evaluations:

$$VisualComplexity = 1.743 + 0.097(TLC) + 0.053(Words) + 0.003(Images) \quad (6.1)$$

Figure 6.1 shows how the two components (Complexity and Visualization Algorithms) fitted in the implementation. An overview of the stages that the implementation of the ViCRAM consists are (interface details can be seen in Figure 6.6):

1. The user loads a Web page into the ViCRAM ‘Browser view’ (top pane).

2. After the user presses the ‘Visualize’ button (in the middle pane) the HTML source code of the target page is sent to the HTML parser that converts the HTML into a DOM tree structure and style information. A screenshot of the page is also taken.

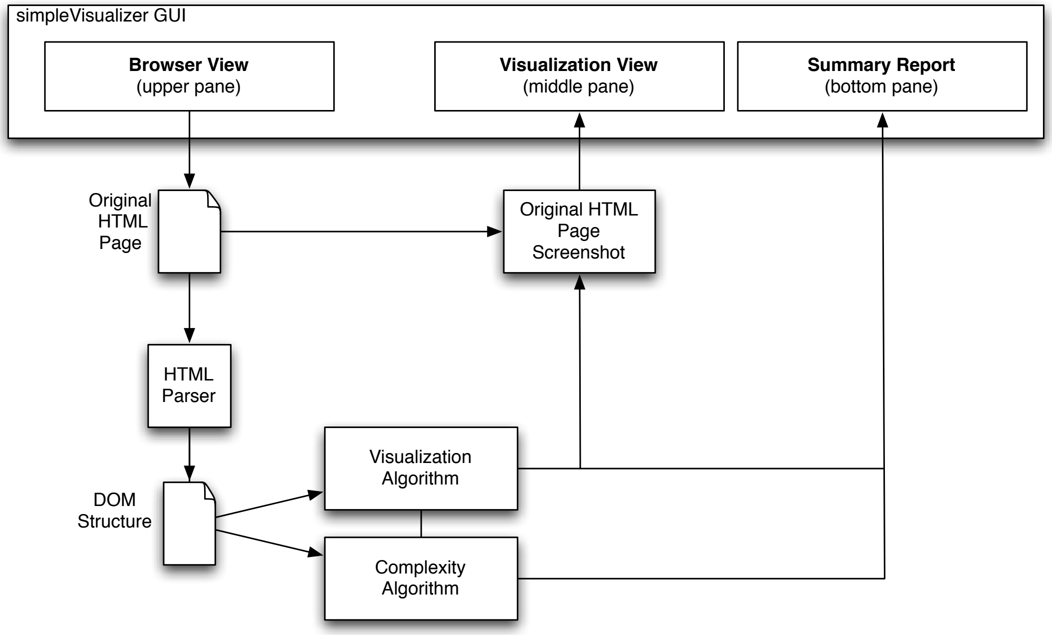


Figure 6.1: ViCRAM Algorithm Implementation Diagram

3. The Complexity algorithm analyzes the structure and style of the page and provides an output text with information regarding the complexity the page.

4. The Visualization algorithm, using also methods from the Complexity algorithm, analyzes the complexity of the page with repsect to grids and colours. It provides visualization data and output information about the data.

5. The visualization data are overlayed on the screenshot of the Web page and the 'Visualization view' (middle pane) shows the heatmap of the analyzed page.

6. The 'Summary Report' view (bottom pane) lists the information given by both the Complexity and Visualization analysis.

6.2.1 Complexity Algorithm

As Equation 6.1 shows, the model is based on the following structural elements of Web pages: TLC, Words and Images. Therefore, the algorithm needed to determine the number of each of these variables for each page. The ACTF provides various APIs that were used during the implementation. One of those APIs is the Document Object Model (DOM) HTML parser that it defines the logical structure of documents and the way a document is accessed and manipulated. When the parser loads a document, it scans the document, looks for elements, attributes, text, and so on and constructs a hierarchical tree based on those items. When the parser encounters an element in the source document, it creates a new node in the tree containing a representation of that element. With this parser the above variables can be determined.

The Complexity algorithm for the ViCRAM tool consists of two important methods and a set of help methods shown in Figure 6.2 as a pseudocode. The first method, `countElements(node)`, is a recursive method that performs DOM analysis. As the pseudocode shows (lines 11 - 20), the method counts the page elements by recursively going through the node using the DOM parser. Some counters are not used in the final equation but are used as part of determining the overall structure of the page used in the TLC count.

The second method, `countTLC(node)`, calculates the number of TLCs a page is grouped into. The TLC definition and classification procedure was described in Section 4.4. Heuristics were derived in order to apply this classification so the TLC count can be automatically determined. The heuristics, as Figure 6.2 shows in lines 25 - 44, use the DOM parser and the style information of the node to determine if a node will be counted as TLC. One is counted when the node is presented as a block with visible border, has headings, and is a table used for data or for layout. A common characteristic is that a TLC is also presented as a block element or a table. In addition, a node can be identified as a TLC only once. There are some exceptions as we explain later but are mainly casused due to the source code of the examined page.

Figure 6.3 shows an example of the TLC identification. The algorithm has highlighted the counted TLCs with a green box for demonstration purposes only. As the example shows, not all the TLCs can be found from the algorithm. Point ① on Figure 6.3 should be identified as one TLC. However, the tabbed horizontal menus were coded as images displayed as inline which did not match the heuristics (rules). In addition, some are counted more than once. Point ② shows an example of a double counted TLC as these were a `<div>` element and a `<h2>` element. Both were identified as TLC due to their block display attribute but also due to an erroneous identification. When the `<div>` element was identified as TLC,

```
Initialize variables;
String calculate(){
 initialize DOM
 reset necessary variables;
 countElements(node);
 countTLC(node);
 recurse through rest DOM;
 calculate VCS based on models;
 return VCS;
}
countElements(Node node){
 find type of the node;
 if Document_node
  extract DocumentElement
 if Element_node
  for all elements in node
   count elements (links, images, tables etc);
 if Text_node
  calculate wordCount;
}
countTLC(Node node){
 find type of the node;
 if Document_node
  extract DocumentElement
 if Element_node{
  Variables that are determined:
  singleChild = node has no children;
  blockChild = display style is a block or table;
  dataTable = if the table has a caption or a theading
  isLayout = if the table has one row OR one column
  isTLC = a flag that the node is already identified as a TLC;
  if ((display=block && blockChild==false)||(singleChild==true && isTLC==false)) =>TLC++;
  if (node is <div> element && has a visible border) => TLC++;
  if (node is a <div> element, has visible border and isLayout==false) =>TLC++;
   else if (node is <h1> or <h2>)  =>TLC++; && headingTLC;
   else if (node is <h3> && headingTLC==false) =>TLC++;
   else if (node is <h4> && headintTLC==false) =>TLC++;
   else if (node is a table && has visible border){
        need to make sure if the table is used for data or layout
          if the table has only visible border =>TLC++;
          if dataTable==true && isTLC==false =>TLC++;
          else if dataTalbe==false && isLayout==true =>TLC++;
          else if isLayout == false && blockChilNodes==true =>TLC++;
          else if node is a <div> =>TLC++;
   }
  }
}
```

Figure 6.2: Complexity Algorithm Pseudocode

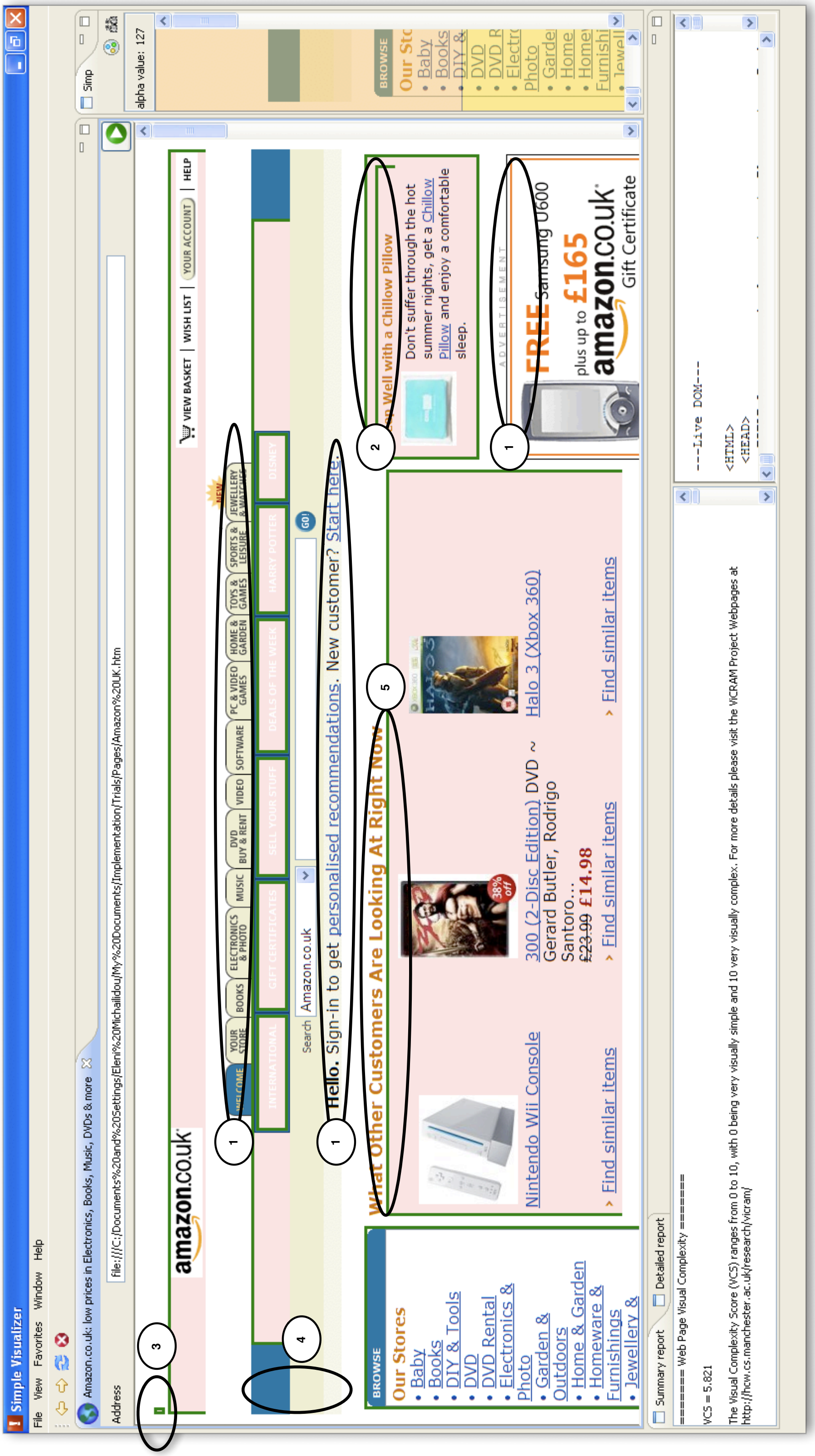


Figure 6.3: ViCRAM - TLC identification example

the `<h2>` tag should not be counted as well. Furthermore, sometimes a TLC is counted when the node is a hidden element. Invisible elements such as the one shown in point ③ are usually used for layout purposes or for use on later stages such as specific announcements (new products, news). The element on point ③ is an image place holder which is probably getting visible on temporary cases.

The duplication, missing and hidden element TLCs occur on those pages that use incorrect or non-standard markup and cannot be validated and therefore more heuristics should be derived. For example, a further check could be implemented to establish when a background colour change occurs (point ④). If that happens then it is a sign that a TLC exists if the associated node is not already identified as a TLC already. When these TLCs are identified, the missing TLC (point ①) could be included in the background colour change TLC. The reason this was not implemented was only due to the project's time limitations.

The main reason for TLC failures is believed to be due to non-compatible to standards coding. For example, a heading element `<h1>` was coded by increasing the font size rather than using the appropriate tag and the algorithm could not then recognize its existence. However, this is dealt with by counting the next TLC within that node (point ⑤), but this could not always happen.

After the page's structural element count (words and images) and layout (TLC) are found the visual complexity score (VCS) is calculated based on Equation 6.1. The 'Summary Report' tab of the ViCRAM interface returns the complexity information of the page. The tool reports a score of VCS from 0 to 10. During our testings on a set of a hundred Web pages from Alexa Top-100, we found that most pages ranged from 0 - 100, with very few generating complexity higher than 100 which was due to long pages with a lot of text only. So, the reported VCS is divided by 10 times of the calculated score. If the calculated score was more than 100, and therefore the VCS is more than 10, then the tool returns '10**'. The '**' is added next to the score along with an explanation that the examined page is very complex and exceeds the higher limit. This approach was decided after the results and conclusions from the technical evaluation described in Section 6.3 in addition with this observed score range. Also, the 0-10 range was always used as the user rankings scale and was found to be understandable and acceptable from the participants during the user evaluations, something that made this range more appropriate and consistent for this project.

6.2.2 Visualization Algorithm

The second aim of this project was to provide an overlay heatmap that mimics a user's visual complexity perception by noting the areas that are most visually complex. In order to achieve this, a Visualization algorithm, was designed, which divides a Web page in a hundred (100) grids and using the complexity algorithm it calculates the VCS for each grid. Then, it colours each grid by assigning a colour based on the ratio of the VCS of the grid and the overall page. The higher the VCS of the grid is, the closer to red its assigned colour is.

```
int[][] findElements(){
Get the overall VCS of the page
Complexity.calculate();
initialize DOM, GridVariablesArray;
reset necessary variables;
create Grid Objects to store data for each grid;
report information;
}
findTypeLoc(Node node){
find type of node;
find location of elements;
assign elements by incrementing appropriate counter in grid;
}
findGrid(Node node){
//helper method
find row&column that the node's top left point belongs into;
}
findTLCLoc(Node node){
mirrors the countTLC() method from Complexity;
determine the grid the TLC belongs into;
}
determineColour(){
set colour array;
find starting colour of page:
        find VCS of page;
        if VCS=0-3.5=>pageColour:first of the simple range;
        else if VCS=3.5-6.5=>pageColour:first of the medium range;
        else if VCS=6.5-10=>pageColour:first of the complex range;
find number of grids per color:
        find location of the pageColour on the colour array (colorIndex);
         numGridsPerColor = (numGrids/(color.length - ColorIndex - 1));
}
colour(int gRow, int gCol, int colour){
colour the rows and columns with given colour;
}
drawGridLines(){
draw lines of the grids on overlay heatmap;
}
```

Figure 6.4: Visualization Algorithm Pseudocode

Figure 6.4 shows the structure of the algorithm's methods and Figure 6.5 a screenshot of the visualization of the previously demonstrated page. As the visualization pseudocode shows, the algorithm divides each page in grids. The number of grids are static but can be easily changed in the code. During our implementations we set the visualization to have 10 rows and 10 columns and hence, 100 grids.

For each node in the page, a procedure similar to the complexity algorithm is followed. That is the number of elements that are used in the VCS model (TLC, images and word count) are identified and for each identified element, its associate node location is determined. By finding the top-left coordinate, the corresponding variable is assigned to the appropriate grid. For example, an image may span in two or more grids but the algorithm increments the image count of the grid that the image's top-left coordinate has on the source code.

In order to find the number of words each grid has, a specific procedure is followed. The word count for the examined node is calculated along with the number of grids that the node spans horizontally and vertically by using its location coordinates and width and height dimensions. The average word count is then calculated for each grid that the examined text node covers. It is important to note again that the algorithm does not count the words on images.

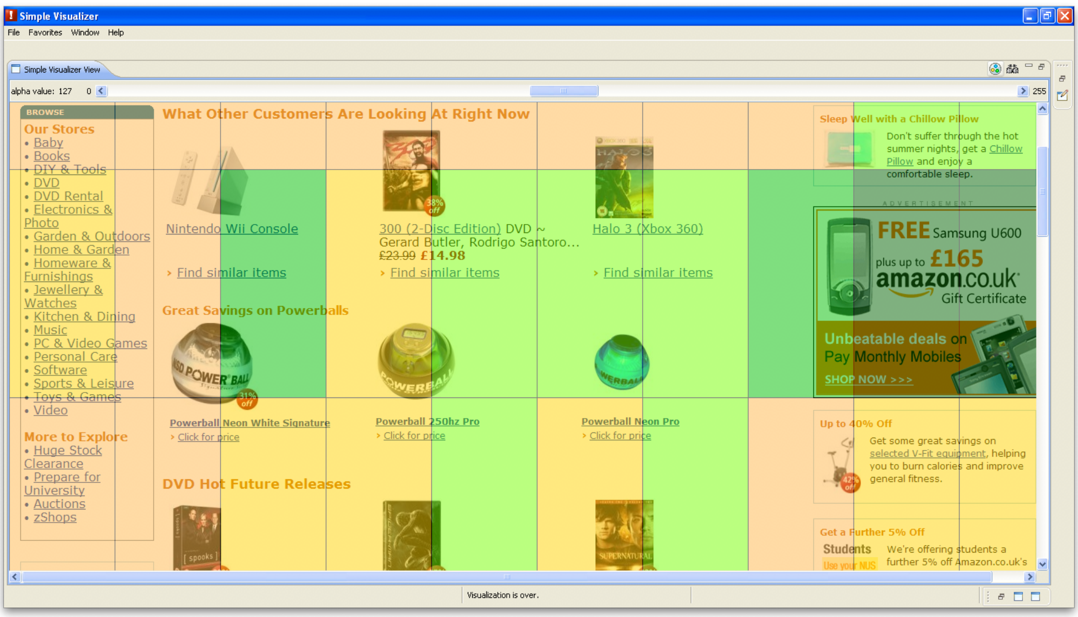


Figure 6.5: ViCRAM - Heatmap example of a complex page

The number of TLCs each grid has is calculated based on the complexity algorithm. When a TLC is indentified the Visualization Algorithm finds the grid that the top-left coordinate of the node is placed into and then it increments the

TLC count of that grid.

Each grid has four variables: TLC, images, wordCount and VCS. The algorithm calculates the VCS of each grid based on the other three variables. Then, based on the page's VCS it assigns a starting colour picked from the colour array which is used as the first colour. As the 'Summary Report' tab shows in Figure 6.3 the examined page had a VCS of 5.821 which ranges in the medium complexity range. Then, the algorithm determines the page overall colour to be the orange as this is in the medium range of the colour array. The colour array can be changed but for our implementation we use the following array: Red - Orange - Yellow - YellowGreen - Green - DarkGreen. The colours were picked to simulate the colours of a heatmap similar to the eye tracking heatmap output. The grids are then coloured starting from the pageColour to the last colour of the array (DarkGreen). Therefore, the more to the Red colour, the more visually complex the grid is. This approach enables one to instantly differentiate a visually complex from a simple page. A simple ranged page will not have red or orange colours that an overall complex page will have. It is important to note that the number of grids is static and further implementation of the algorithm could include an automatic generation of the number of grids based on the height and width of the page.

The number of grids that each colour is used for, is calculated based on the following formula:

$$gridsPerColour = (numberOfGrids - minVCScore)/(colourArrayLength - pageColour) \tag{6.2}$$

where *numberOfGrids* is the total number of grids the page is divided into, *minVCScore* is the total number of grids with the minimum value of Visual Complexity Score (when all variables are zero), *colourArrayLength* is the number of colours used and *pageColour* the page's overall colour which is based on the page's VCS.

The 'Summary Report' tab of Figure 6.6 shows the rest of the ViCRAM's output which describes the heatmap. It provides an explanation of the heatmap colouring and then, for each grid lists its respective VCS and associated variables (TLC, images, words).

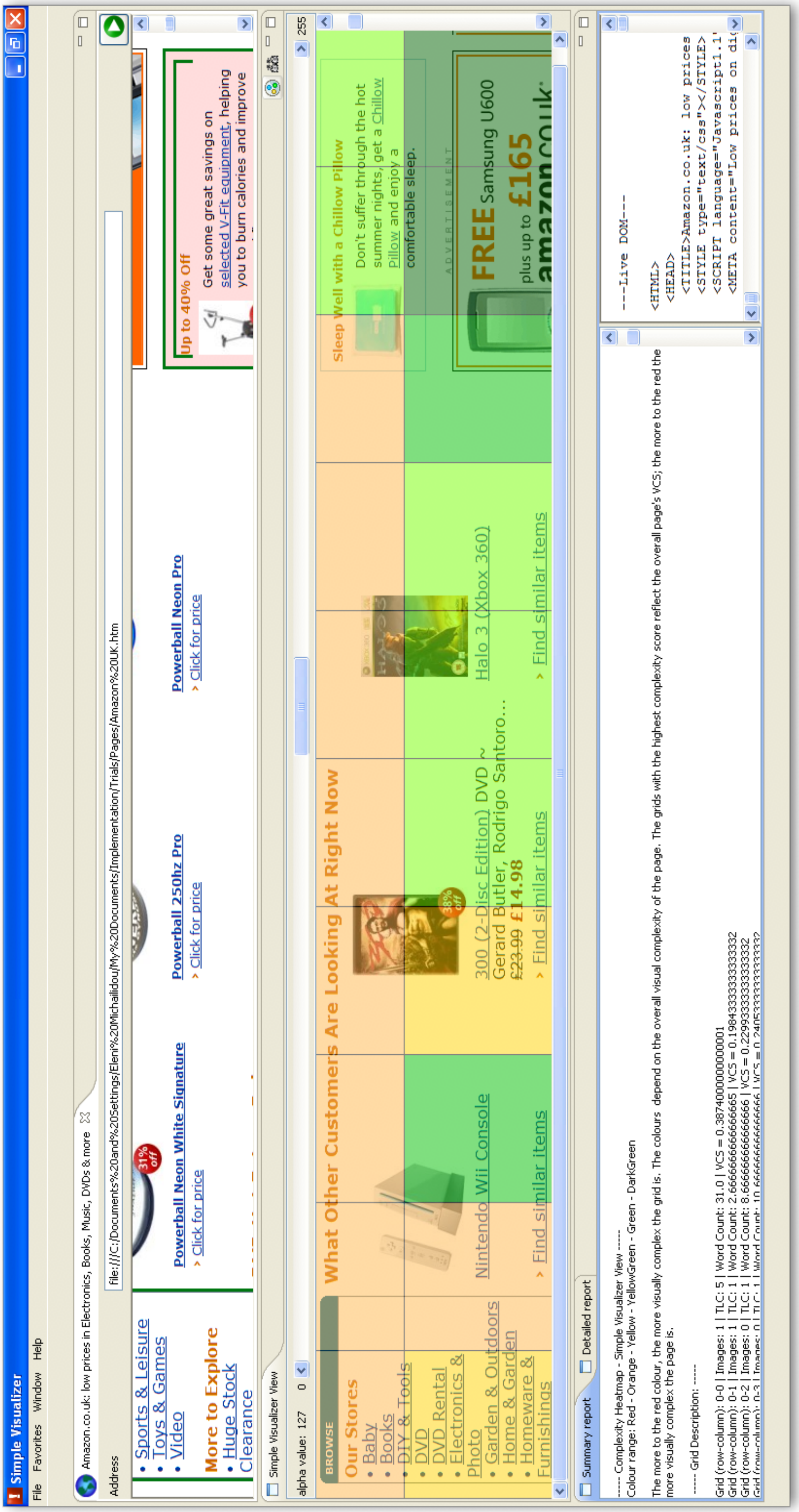


Figure 6.6: ViCRAM - Visualization Report

6.3 Technical Evaluation

In order to examine the effectiveness of the tool a technical evaluation was performed before running a user evaluation. In this way, any problems were identified and the algorithm edited before assessing the output with users.

Thirty Web pages were used during the technical evaluation. The Web pages were the same ones used for the Complexity and Aesthetics user evaluation described in Chapter 5. Table 6.1 lists the Web pages and the scores received during the previous evaluation (listed under 'Study'). Screenshots of the images can be found in Appendix G. The source code of these pages were saved at the same time the screenshots were taken.

Each page was loaded in the ViCRAM Browser view and the 'visualize' procedure was launched. The results with respect to the visual complexity score were taken from the 'Summary Report' view pane. The complexity score reported for this study was the value given before any adjustments to fit in the 0 - 10 score range. Table 6.1 also lists the values the tool returned and are listed under the 'Tool' column. The empty cell in the Tool columns are for pages that either could not provide a score for reasons such as missing scripts or lengthy pages that caused the code to crash.

6.3.1 Results

The prediction level of the tool was examined by running a correlation analysis between the Study and Tool scores for the complexity and aesthetic terms listed in Table 6.1. The table lists the Spearman's ρ correlation coefficient and whether the test was significant at $p < .05$.

As the analysis revealed, the algorithm significantly predicted the user rankings. The correlation coefficient was determined to be $\rho = .559$ at $p < .01$. This correlation has a high effect as it is higher than .5 [Field 2005] which can reflect the level of effectiveness of the tool. Figure 6.7 shows the rankings that users gave on average for each page (during the previous user evaluation) and the tool generated scores in ascending order. One can notice the positive correlation between each score type, with some exceptions discussed below, which can be seen by both data's trendlines that are both on the same positive trend.

The tool rankings for the graph were divided by 10 for two reasons. The first reason was for graph practicality so the correlation between user and tool

Table 6.1: ViCRAM Tool Technical Evaluation Data

PageID	PageName	Visual Complexity Score	
		Study	Tool
1	Amazon UK	5.09	55.59
2	AnnoteaProject	3.43	43.52
3	AutoTrader	6.86	62.39
4	BBC UK News	5.81	42.11
5	BBC UK	5.62	28.92
6	BloggerPostHQ	3.42	.
7	BloggerPostDE	3.96	67.21
8	Blogger Dashboard	4.75	37.43
9	Delicious	5.63	25.77
10	Ebay	5.41	27.54
11	Firefox	3.09	14.75
12	Flickr	3.06	7.08
13	GoogleSearch	3.78	32.82
14	GumTree	5.76	25
15	IMDB	6.09	56.82
16	InvisionFree	5.44	22.62
17	Job Centre	3.26	8.11
18	MegaUpload	6.44	.
19	MySpace	6.25	27.72
20	Orkut	1.61	5.78
21	Rapidshare	4	15.52
22	Rightmove	4.18	16.56
23	StudentNet	3.92	15.53
24	StudentNet SelfService	2.08	11.7
25	WAI	6	66.77
26	Wiki Result	4.46	.
27	Wikipedia	2.6	32.47
28	Yahoo UK	6.21	30.42
29	Yell	2.2	22.9
30	YouTube	5.47	105.07
Spearman's rho - Correlation Coefficient		**.559**^{b}	

a. $p < .05$; b. $p < .01$;

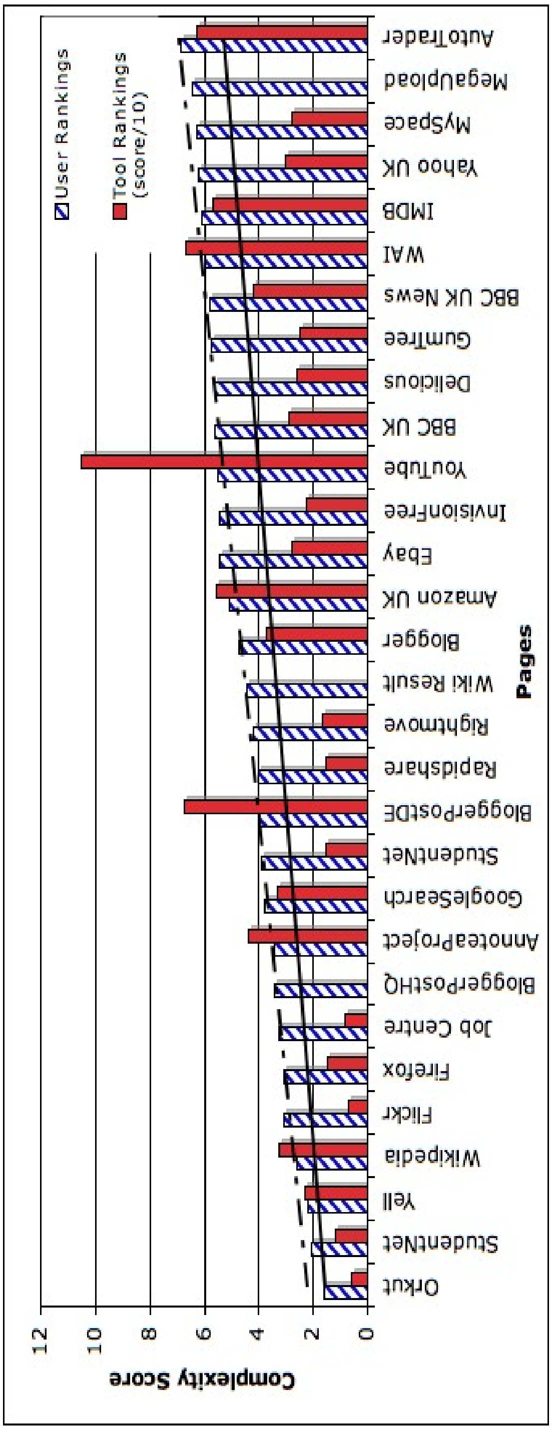


Figure 6.7: User and Tool Rankings

rankings was easy to identify. The second and main reason was that the ranks were less than 100 and along with the previously observed score range described in page 145 of 0 -100, it was decided to divide the score for presentation and consistency purposes only as this range was the one used throughout this project, both for reporting but mainly during the user evaluations.

As explained above, the tool crashed for three pages for which no score was provided. These pages were very long (see PageID-6 and PageID-26) and mainly dynamic (PageID-18). In order to solve the problem with the lengthy pages, we tried to limit the algorithm to run only on the top half of the page. After re-running each page through the tool, collecting the scores and running a correlation analysis, we found that the complexity analysis for some pages was limited a lot and the scores generated were changed to a point that the complexity correlation between user and tool rankings was not significant. An attempt to overrun the crashing problem led us to realize that internal API classes had to be changed, classes that we could not manipulate. The aDesigner developers assisted in this case and corrected specific classes. After their correction, we updated the code and the crashing problem was solved.

6.4 Discussion and Conclusions

Being a prototype, this tool[2] could be further improved if some issues discussed previously were resolved. For example, the algorithm could be improved to determine the TLC count more efficiently. This could happen if the Web page is coded using correct or standard markup and can be validated but also if more heuristics are applied. One of these heuristics should be with respect to background color-change. When in connected block level elements a background colour occurs then a TLC should be counted if that node is not already identified as one.

Another heuristic could be with respect to hidden elements. The algorithm does not check if an element can be seen or not by the user. Therefore, if one of those element fits under one of the rules implemented then the algorithm identifies that as a TLC which increases the visual complexity.

The algorithm could be also advanced by taking into account further weight factors for each of the variables (image, word count, TLC). For example, image size, purpose of use and whether is a dynamic image are three different factors

[2]Tool Installation instructions are given in Appendix H

for the image variable. A big and dynamic image causes more complexity than small and decorative ones. In order, though, to examine the effect of these factors further evaluation studies need to be performed especially using eye tracking technologies. Similarily with images, word count can have further weight factors. For example, words from headings, lists, menus or image captions can have different effects between each other and from paragraph text word count. Determining the correlation between visual attention and complexity perception could provide models that when implemented on top of the tool can advance the existing ViCRAM tool.

The visualization algorithm can also be improved by providing to the user the option of defining the number of grids the page's visualization will have. In addition, the grids could regenerated when the user will resize the browser and the image size will change. Currently, these facilities are not provided but are feasible issues. The image count and coordinate identification in the visualization algorithm could improve as well, when followed the wordCount procedure. That is, if the algorithm identifies the number of grids (horizontally and vertically) that each counted image spans into, the associated grids could increase their image count. Then the visualization (heatmap) of the page will be more specific on the complexity of each area/grid.

Additionally, more studies could be used to examine the effect of more page elements on visual complexity and their attributes. These could be links, menus, input forms and their respective colour contrast, font sizes and positions. In order to achieve this, studies with larger set of evaluated pages (in the range of 100s) need to be executed. Participants' sample number need to be also more than 100 in order to provide a more complete data set for such a large variable number.

This chapter described the implementation of the visual complexity framework into the ViCRAM interface. Using the prediction models derived through the previous user evaluations the algorithm predicts the visual complexity of a Web page which points towards the third research aim of this thesis (Q3 - see page 23) on analysing Web page layout to automatically predict visual complexity level of pages. An algorithm was also implemented that can create an overlay heatmap of a given Web page by noting the areas of most visual complexity. A technical evaluation was performed which revealed that, even with the algorithm limitations discussed above, the tool can provide complexity information that significantly relates with the information gathered by participants during the user evaluations performed throughout this project. A user evaluation should be conducted as a next step in order to determine the level of prediction of the algorithm and therefore support our final research aim.

Chapter 7

User Evaluation of the Complexity Algorithm

This chapter describes the user evaluation of the visual complexity algorithm that was implemented into the ViCRAM tool. In order for the tool to be efficiently used by users and designers, the algorithm must provide realistic information. To achieve this, users were asked to rank Web pages based on their visual complexity. These rankings were then compared with the complexity information that the ViCRAM tool provides for the same pages. The user evaluation demonstrated that the models previously predicted (see Chapter 5) were appropriately implemented into the ViCRAM tool and therefore the tool provides information about the visual complexity of Web pages that is significantly correlated with users' perception.

7.1 Methodology

As explained in previous chapters, the visual complexity level of a Web page relates to its design. Each Web page design is differentiated by page composition, layout, formatting and overall characteristics, such as tables, images, links, and colours. We assert that the visual complexity of a Web page can be automatically predicted and correlates highly with users' scores. During this evaluation users were asked to rank thirty Web pages based on their visual complexity appearance. The user rankings were then compared with the rankings automatically generated for each page by the tool.

The hypothesis for this experiment was therefore:

H1 Visual Complexity rankings generated automatically by the tool are significantly and positively related to users' rankings. The tool can be then used to predict users' visual complexity perception of a Web page.

The design and procedure of this evaluation was identical to the previous online ranking experiment described in Chapter 5 order to control the validation of the tool as much as possible. Sections 5.1.1 and 5.1.3 describe the design and details of the procedure followed for the evaluation.

7.1.1 Participants

The study was conducted online to allow participants to access it in their own time and place. The study was advertised through mailing lists and newsgroups, such as CHI-WEB[1] and DBWorld[2]. One hundred and four participants from around the world volunteered to take part in this evaluation, 53 (51%) male and 51 (49%) female. Fifteen (14%) participants were aged up to 25, 24 (23%) between 46 and 65 and the rest between 26 and 45. From them, 67 (64%) participants had English as their native language. No participants were colour-blind. One data set from a participant was dropped as they reported that they had a problem loading the images with their browser.

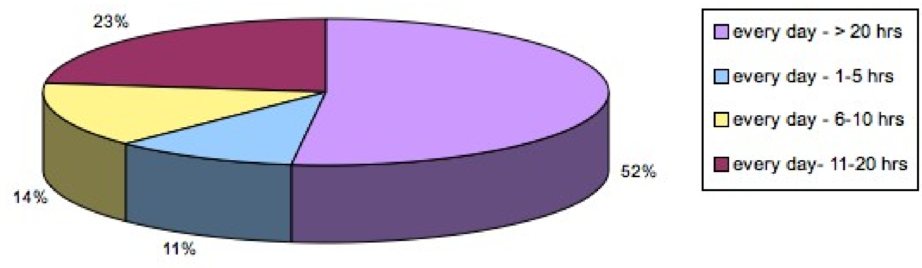


Figure 7.1: Participants' Internet Usage in Hours per Week

All the participants reported that they used the Internet daily with 52% using the Internet more than twenty hours per week, nobody less than an hour,

[1]CHI-WEB, http://sigchi.org/web/
[2]DBworld, http://www.cs.wisc.edu/dbworld/

11% for 1-5 hours, 14% for 6-10 hours and the rest between 11 and 20 hours. Figure 7.1 shows these results. 97% of them described that they use the Web for business/work, 95% for email/chat, 96% for special interests, and 87% for on-line purchasing. A breakthrough of Internet usage is shown in Figure 7.2. Even though most of the sample used the Internet daily for more than 20 hours a week, all users were familiar with all kinds of browsing (business/chatting/purchasing) which makes the sample more generalizable.

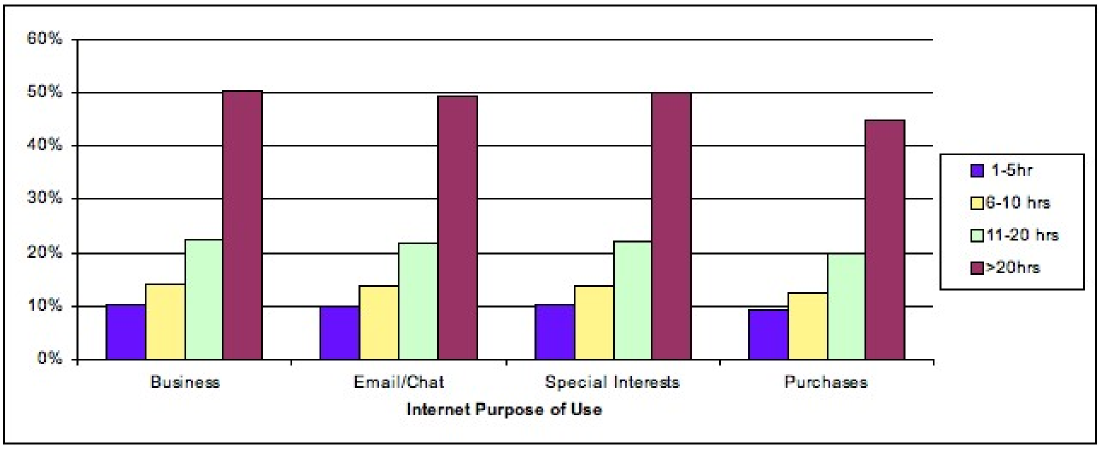


Figure 7.2: Participants' Purpose of Browsing per Hours

7.1.2 Materials

The pages participants viewed were selected to be representative of sectors such as public information, business, academic, entertainment, leisure, Web services including search engines, and personal home pages [Amitay *et al.* 2003]. Pages were similar to the set of pages used during the previous evaluations performed throughout the project, and part of the Alexa UK top 100 Websites[3] to provide a representative sample of the sites people actually browse. Out of the thirty pages selected from Alexa top 100, eight of them were the same URL and had the same visual layout (but not content) as eight pages originally used in the previous user evaluation (Chapter 5). The similarity was due to the fact that seven of these eight pages were also on the Alexa top 100 list and one of them (Annotea - PageID27) was selected due to its simple design.

Table 7.1 lists the thirty Web pages selected as the set of stimuli for the evaluation and screenshots can be found in Appendix J. In order to avoid any

[3]Alexa UK Top 100 - English Language: 14 January 2009

bias, the images were taken as a screenshot from the same monitor, within the same day and all the stimuli had the same size of 1111 x 800 pixels. In addition, the source code with any associate files was saved for each Web page during that period.

7.2 Results

Participants ranked the images based on their visual complexity by giving a number from 1 to 10, with 1 being the visually simplest and 10 the most complex. The data collected during this evaluation are available to download from the experimental folder linked to the technical report (see page 265). The term Scores is used for the visual complexity ranking score that each participant assigned to the images; ScoreA and ScoreB the scores collected during the first and second time participants looked at the images respectively. Terminology of various statistical terms used throughout the analysis can be found in Appendix A.

7.2.1 User Scores Correlation

As discussed in Section 7.1, participants ranked each image twice in a counterbalanced sequence. A reliability test between each score was conducted in order to check for consistency of the participants' scores, which was achieved by running a bivariate correlation on all the scores. It is important to note that a correlation coefficient (ϱ) of larger than .3 has a medium effect, larger than .5 a large effect and the closer to 1 the better and stronger the correlation is. The correlation coefficient determines the strength of relationship between two variables. Whether a correlation is statistically significant is determined by the p value. If this value is less than .05 and the closer to 0 the better the correlation is. These are discussed further in [Field 2005].

The complexity ranking scores given for each stimuli the first time is significantly correlated with the ranking scores given the second time, with Spearman's rho (ϱ) correlation coefficients of $\varrho > .5$ and significant at $p < .0001$. Images with ID 16 and 21 had a smaller correlation value of $\varrho > .43$ but are still significant at $p < .0001$. The medium effect (.43) could be caused by the user's familiarity with the page or by users changing some of their rankings during the second viewing after they had formed an opinion with respect to complexity based on all thirty images. No significant correlation was determined between the user's familiarity

Table 7.1: Pages used in the Experiment and the Visual Complexity Scores generated by the tool (Alg Score) and by the participants (Score A, Score B and Averageg Score)

ID	Page Name	Page URL	Alg Score	Score A	Score B	Average Score
1	4Shared	http://www.4shared.com/	1.45	3.67	4.04	3.86
2	Amazon UK	http://www.amazon.co.uk/	8.40	5.93	6.01	5.97
3	AOL UK	http://www.aol.co.uk/?src=www.aol.com	7.16	6.45	6.78	6.62
4	Badongo	http://www.badongo.com/	1.77	4.65	4.88	4.77
5	CNN International	http://edition.cnn.com/	8.84	5.18	5.76	5.47
6	Comcast	http://www.comcast.net/a/	9.55	6.35	6.46	6.4
7	DevianArt	http://www.deviantart.com/	6.67	4.89	4.96	4.93
8	Digg	http://digg.com/	7.65	5.44	5.63	5.54
9	Download	http://www.download.com/windows/?tag=TOCleftColumn.0	4.11	5.85	5.83	5.84
10	Ebay UK	http://www.ebay.co.uk/	2.15	5.36	5.37	5.36
11	ESPN	http://www.espn.go.com/	6.01	6.98	7.22	7.1
12	Filefactory	http://www.filefactory.com/	3.81	4.7	4.62	4.66
13	Flickr	http://www.flickr.com/	0.79	3.17	3.32	3.25
14	Fotolog	http://www.fotolog.com/	2.42	5.65	5.81	5.73
15	GameSpot	http://uk.gamespot.com/	8.42	6.28	6.9	6.59
16	GO	http://go.com/	3.18	6.34	6.4	6.37
17	Google Search (Use Initial)	http://www.google.com/search?hl=en&q=manchester	3.03	2.91	2.93	2.92
18	Image Venue	http://www.imagevenue.com/	1.19	3.13	3.15	3.14
19	IMDB	http://www.imdb.com/	4.45	6.68	6.84	6.76
20	Live Search Manchester	http://search.live.com/results.aspx?q=manchester&form=QBLH	3.13	3.63	3.82	3.72
21	Photobucket	http://photobucket.com/	2.51	5.42	5.64	5.53
22	Rapidshare	http://www.rapidshare.com/	0.83	1.3	1.53	1.41
23	Reference	http://www.reference.com/	0.83	3.43	3.85	3.64
24	WordPress	http://wordpress.com/	1.88	4.6	4.64	4.62
25	Yahoo	http://www.yahoo.com/	3.07	6.87	6.81	6.84
26	zShare	http://www.zshare.net/	1.45	3.09	3.19	3.14
27	Annotea	http://www.w3.org/2001/Annotea/	4.34	2.98	3.57	3.27
28	GumTree	http://www.gumtree.com/	3.32	5.31	5.4	5.36
29	Invision Free	http://invisionfree.com/	2.47	4.95	5.11	5.03
30	Job Centre Plus	http://www.jobcentreplus.gov.uk/JCP/index.html	0.90	3.92	4.1	4.01

and the complexity scores they assigned to each page. This leaves the reason for formulating a better opinion on complexity during the second round of rankings more possible. This is further discussed in Section 7.2.2.

The fact that all the correlation values were larger than .4 at $p < .0001$ shows that the complexity scores given by the participants the first time are significantly correlated and have high effect with the scores given on the same stimuli the second time. Also, a strong correlation was determined between the mean values of ScoreA, ScoreB and AverageScore with $\varrho > .983$ at $p < .0001$.

7.2.2 User Familiarity and Reaction Time

The time that each participant took to provide an answer about the visual complexity level of the image and whether they were familiar with the page were also recorded. No significant correlation was determined between the mean reaction times and the complexity scores given for each page or between the mean familiarity value and the complexity scores.

We noticed that there were some outliers in the reaction time data. That is, some participants spent up to 5-10 minutes to rank a page. In order to deal with such outliers, the median value for each page was found and any reaction time higher than 5 digits (the recorded time was in milliseconds) was removed from the data set and replaced with its respective median value. This is a common statistical analysis procedure also discussed in [Field 2005]. By replacing the outlier with the median value, the number of examined pages stays the same and does not decrease the overall mean value which is used for the analysis. The nature of the outlier needs to be checked before this action as in the case of user or design problem the outlier needs to be removed. This could be checked from the associated demographic data, text comments or reaction time for the score.

The average reaction time was then calculated for each page and used for the second run of the analysis. The new correlation analysis did not reveal any significance, concluding that the time a participant needed to provide a score on the visual complexity level of a page was not significantly related to the page's level of complexity even by removing the outliers.

A correlation analysis between complexity score, familiarity and reaction time for each page individually was also conducted. For some pages a significant relationship was revealed but the correlation coefficient was very low, signifying a low effect between the two variables (less than .3). Another observation from this analysis was that the pages that revealed significant correlation between ScoreA and FamiliarityA were not the same as the pages that revealed significant correlations between ScoreB and FamiliarityB. The same happened between ScoreA - ReactionTimeA and ScoreB - ReactionTimeB, ReactionTimeA - FamiliarityA and ReactionTimeB - FamiliarityB. This can add to the low effect of the significance of the above correlations. The only pages that had significant correlation between Scores and Reaction Time for both versions were pages with ID11 and ID27, but the low value of the correlation coefficient (around .2) does not allow any further discussion on a possible correlation to be made. Figure 7.3 shows the average user scores and familiarity value with each page ranked in complexity ascending order. We can see that the pages that users were more familiar with (any score above 5) were: ID17 - Google Search Results, ID13 - Flickr, ID10 - Ebay UK, ID5 - CNN, ID2 - Amazon, ID19 - IMDB and ID25 - Yahoo.

Figure 7.3 shows the complexity and familiarity scores for each page in user complexity score ascending order and Figure 7.4 shows both complexity scores in ascending order of the algorithm complexity score. By looking at both the graphs one can notice the complexity range (simple, medium, complex) that each score version (user and algorithm rank) assigns with each page. When we compare the users' rankings (Figure 7.3) with the scores that the tool assigns to each of the above pages (Figure 7.4) we notice that some pages fall in the same range as the users' rankings range (these are ID2, ID13 and ID19) and some in the closer range (these are ID10 and ID17). For example, page ID17 - Google Search was ranked by users as very simple but by the algorithm as medium and closest to the simple range which shows that the algorithm predicts the range close enough. On the other hand, the algorithm scores for pages ID25 and ID5 did not fit close with the users' scores. That is, users ranked ID5 - CNN as medium complexity and the algorithm ranked it as one of the most complex. The opposite occurred with ID25 - Yahoo.

Familiarity could have a low effect but this could have occurred mainly because the page has a lot of structural elements that are invisible to the user and since the algorithm is based on the structural elements and overall layout of the page only, it returns a higher score than the users' ranking. As it was discussed in Chapter 6, the invisible elements is one of the issues that needs further investigation.

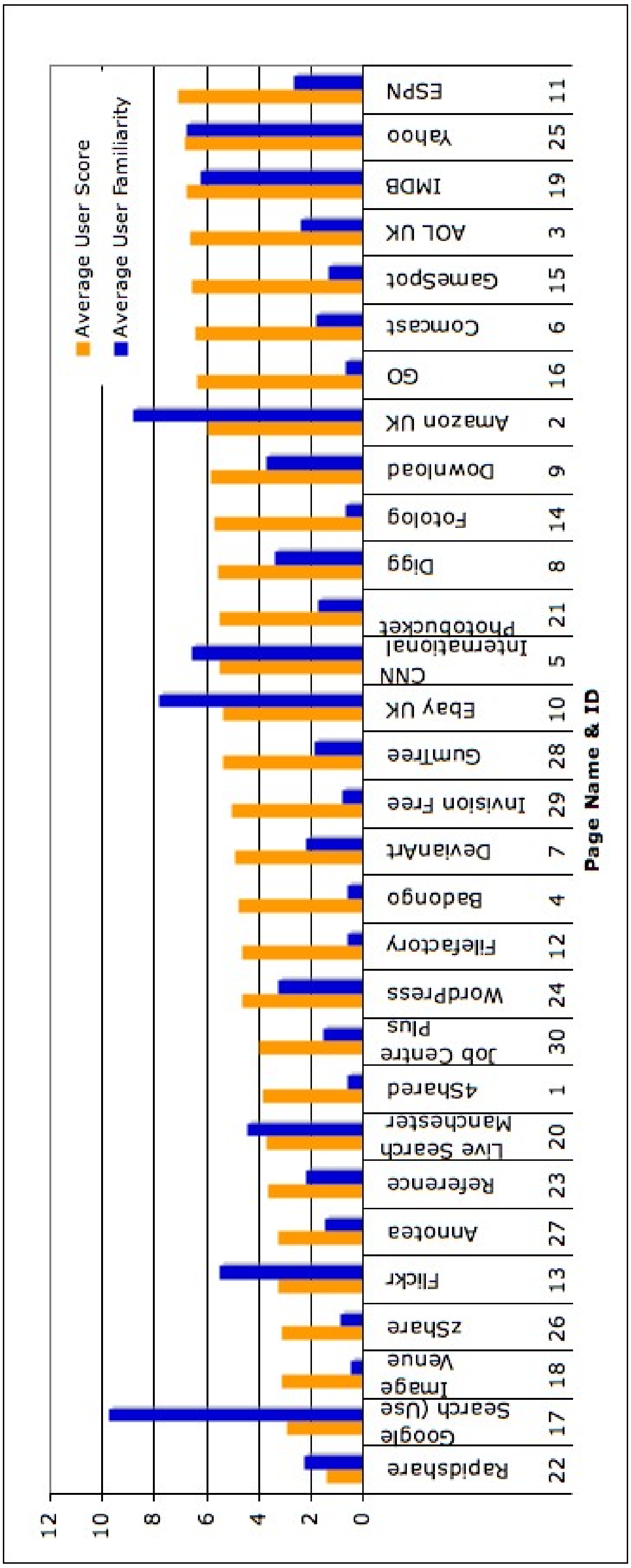


Figure 7.3: User Familiarity and Complexity Scores

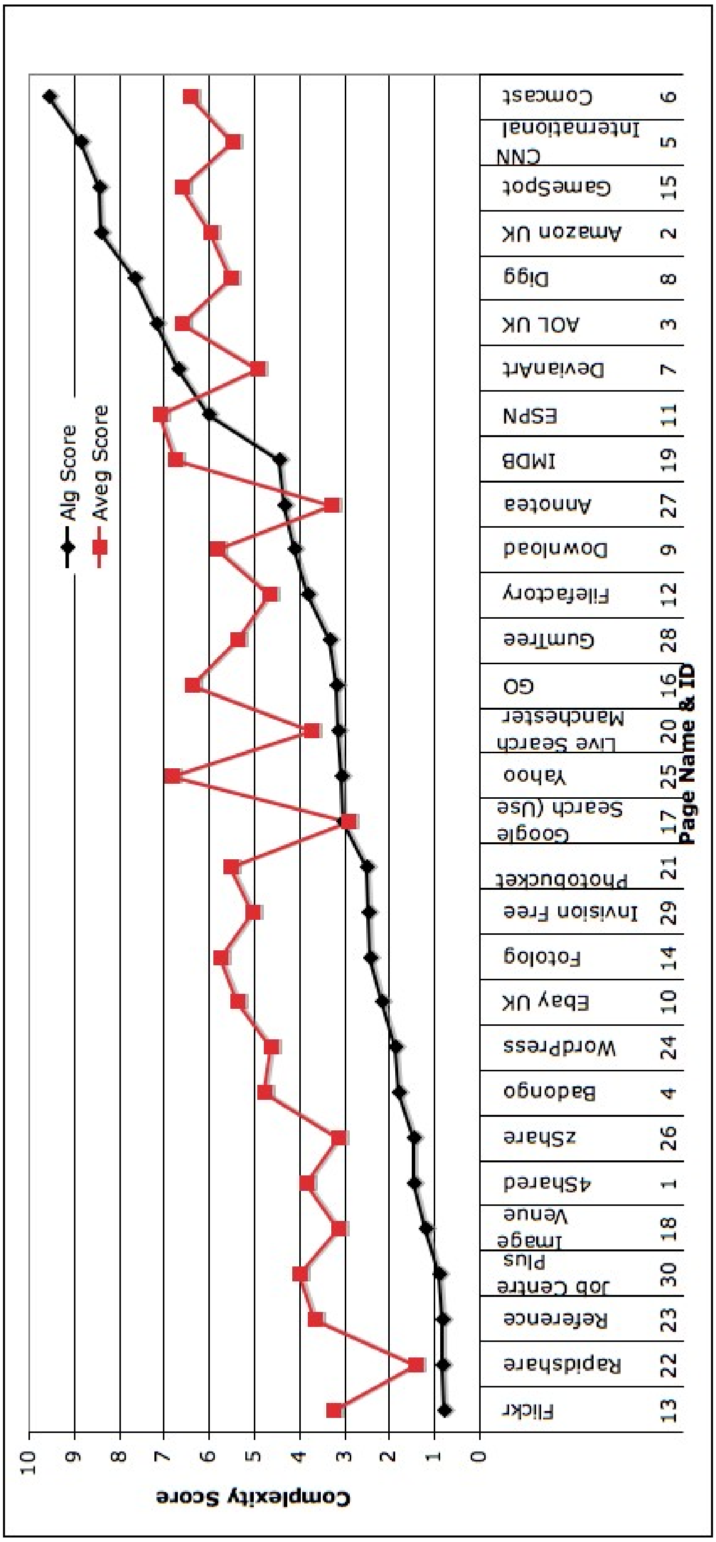


Figure 7.4: Visual Complexity Scores

7.2.3 Tool Significance

Table 7.1 lists the mean values of the ranks given for each page by the users (Averg. Score) along with the scores generated by the tool (AlgScore). Figure 7.4 shows the Web pages in ascending order of visual complexity based on the tool scores and Figure 7.5 the trend-line (generated by Excel) that the scores from both data sets have. Both graphs show a positive relation between the complexity scores that the participants gave and the scores from the algorithm.

In order to examine the level of relationship between the two data sets, a Bivariate correlation test was performed between the algorithm score and the three scores received by the participants (Score A, Score B and the Average Score). It is important to note that after examining the data we noticed that a few of the scores did not fit a normal distribution and we decided to examine the non-parametric correlations.

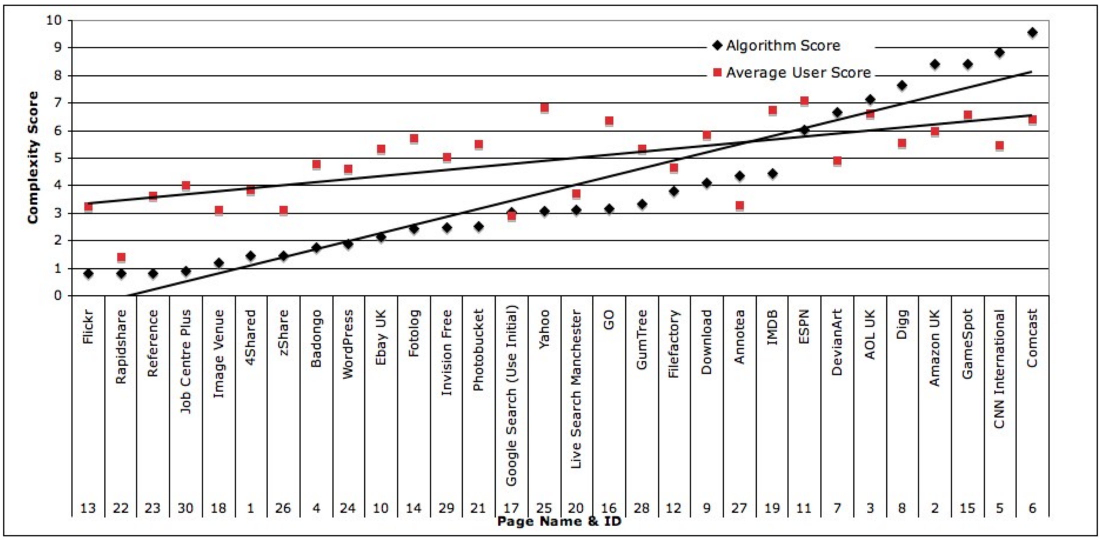


Figure 7.5: Visual Complexity Scores - Trendlines

Table 7.2 shows the Spearmans' ϱ correlation coefficients between each data set. The Algorithm Scores were found as significantly and positively correlated with the participants' scores at the high level of the range $\varrho > .64$ and $p < .0001$. It is interesting to note that the Average Score between the participant's rankings and the score that the automatic tool generated were significantly correlated at $\varrho = .68$, p=.0. These results show that the framework could predict at a high level the participants' perception with respect to the visual complexity level of a Web page, based on the structural elements and overall layout of the page.

Table 7.2: Spearman's ϱ Correlation between Tool and Participants Visual Complexity Scores

	Alg Score	**Score A**	**Score B**	**Average Score**
Alg Score		.644[a]	.689[a]	.679[a]
Score A	.644[a]		.983[a]	.994[a]
Score B	.689[a]	.983[a]		.994[a]
Average Score	.679[a]	.994[a]	.994[a]	

a. p < .0001

7.2.3.1 Linear Regression

Using the AvgScore as the dependent variable and the AlgScore as the predictor, a regression test was run. A significant model emerged (F=16.736, p<.001) with a medium effect fit of R^2=.374 and $R^2{}_{adj}$=.352. Two more regression analysis tests were performed using ScoreA and ScoreB as dependent variables respectively. Tables 7.3 and 7.4 show the Model's Summary and Coefficients that each analysis produced respectively.

Table 7.3: Stepwise Regression - Users' Visual Complexity Score Model Summary

Dependent Var.	**R**	R^2	$R^2{}_{adj}$	**Std.Error Est.**	ANOVA **F-ratio***	**Durbin-Watson**
AvgScore	.612	.374	.352	1.150	16.736	2.299
ScoreA	.591	.350	.326	1.178	15.056	2.288
ScoreB	.630	.396	.375	1.129	18.384	2.308

∗ All Significant at p < .0001;
Model Predictors: (Constant), AlgorithmScore;

The value of R^2 describes how much of the variance in the Users' Visual Complexity Score is accounted for by the regression model from the sample, where the Adjusted R^2 ($R^2{}_{adj}$) shows how much variance in the dependent variable (here Users' Visual Complexity Score) would be accounted for if the model had been derived from the population from which the sample was taken (here the Web pages) [Field 2005]. In other words, the $R^2{}_{adj}$ gives some idea of how well a model generalizes and ideally one would like its value to be the same, or very close to, the value of R^2.

For a good model, one should look at the value of R^2 and $R^2{}_{adj}$ and check how close their values are with each other. A value of R that is larger than 0.5 has a large effect and hence the model is more effective. Also, the larger the F value returned by the analysis, the better the model is (if it is significant at p <

Table 7.4: Linear Regression - Coefficients

	Average Score				Score A				Score B			
	B	SE B	β*	t	B	SE B	β*	t	B	SE B	β*	t
Model 1												
Constant	3.680	0.370		9.938	3.626	0.379		9.563	3.735	0.364		10.274
AlgScore	0.323	0.079	.612	4.091	0.314	0.081	.591	3.880	0.333	0.078	.630	4.288

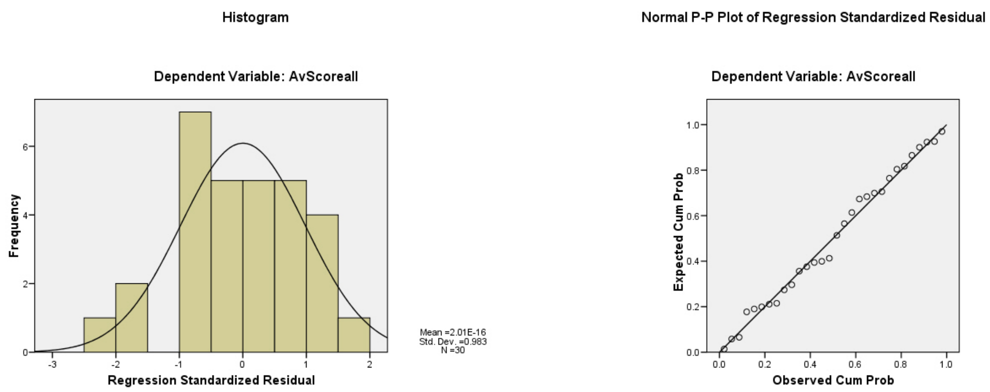
∗ All significant at p < .0001

.05). In addition, the Durbin-Watson value should be as close to 2 as possible and in general between 1 and 3. Throughout all the experimental analysis, the first score that participants gave derived the less effective, but still highly significant, models. Using the mean values from ScoreB data the best model with the strongest significance in the coefficients emerges. This could be because the users had seen all the pages during the first round and formulated an opinion about visual complexity.

The models emerged have a medium fit ($R^2 > .35$) but the overall effect of the model is high (R > .59). The R value shows how well the algorithm can predict the user's visual complexity score. Here, the model based on the AvgScore shows that the tool can account for 37.4% of the variation in users' visual complexity scores. This means that the tool can approximate users' perception on visual complexity by 37.4%. Even if this value would be considered between the medium effect range one should take into account the context of this model.

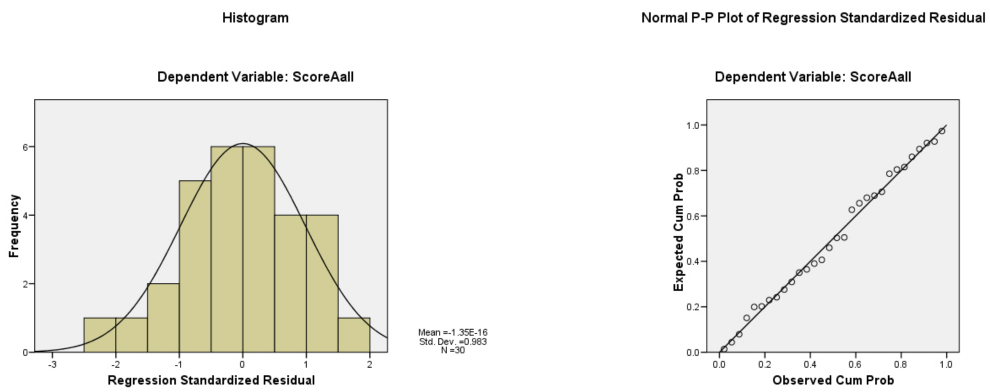
Predicting user perception of Web page complexity can have factors that are not yet considered from both sides of the equation: the user perception and the algorithm. On one hand, the user perception can be altered from the user's individual personality characteristics that are difficult to predict, user's culture and social trends. On the other hand, the algorithm, as discussed in Chapter 6 can take into account structural elements that are not yet evaluated and implemented, such as page type, font sized and colours, along with improving the existing algorithm.

Residuals are known as the difference between the values of the outcome predicted by the model and the values of the outcome observed in the sample. These residuals effectively represent the error present in the model, usually caused by outliers [Field 2005]. A way of checking how well the regression model fits is to check the normality of the residuals (the standardized residuals) which can be done by looking at the histogram and normal probability plots as shown in



(a) *Histogram* (b) *Normal P-P*

Figure 7.6: Average Score - Regression Model Normality Check

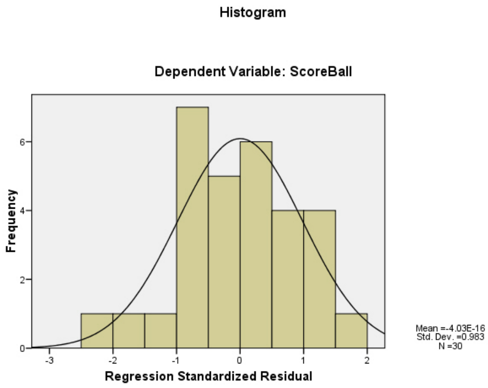


(a) *Histogram* (b) *Normal P-P*

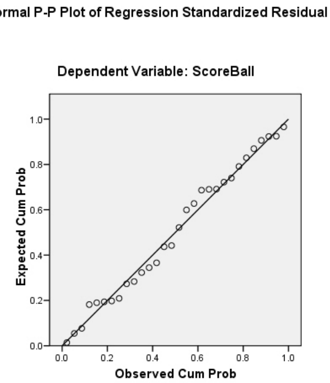
Figure 7.7: ScoreA - Regression Model Normality Check

Figures 7.6, 7.7 and 7.8. The histogram should look like a normal distribution (a bell-shaped curve) and the figures show that all histograms have a good fit of normality. The straight line in the probability plots represents a normal distribution and the points the observed residuals. In a perfectly normally distributed data set, all points will lay on the line. The model using the ScoreA data set had the best normality within the rest sets as most of the points lay on the line. However, all the probability plots had a very good fit.

After checking for the best fit of the data and their collinearity, using the Algorithm Score as the only predictor in a linear regression model, the analysis revealed three possible models based on AverageScore, ScoreA and ScoreB respectively:



(a) Histogram



(b) Normal P-P

Figure 7.8: ScoreB - Regression Model Normality Check

$$UserVCS_{av} = 3.680 + 0.323(AlgScore) \tag{7.1}$$

$$UserVCS_{A} = 3.626 + 0.314(AlgScore) \tag{7.2}$$

$$UserVCS_{B} = 3.735 + 0.333(AlgScore) \tag{7.3}$$

As discussed above, all data had a good fit (R^2) and high effect (R) of significance. We should note that using Familiarity in these prediction models results in subjective and less strong models because they require user input. In addition, Familiarity was not significantly correlated with users' rankings. Hence, it was decided not to include this variable in the regression analysis.

7.2.4 Feedback Comments

The last part of the evaluation was introduced to directly ask participants' opinion of what a visually simple and complex page is. First, the participants were asked whether they agreed or disagreed with a series of statements shown in Figure G.4. As Figure 7.9 shows, most of the participants disagreed that they "can easily find information on a visually complex page" but they strongly agreed that they "can easily find information on a visually simple Web page". The majority of them also agreed with the statement that they tend to "spend more time interacting with a visually complex page than a simpler one", but more data will be needed to be collected in order to explain the reason for this. We assume that this happens due to the amount of information a visually complex page has, which is also shown from their agreement on the fact that "a visually complex page has a lot of information to choose from".

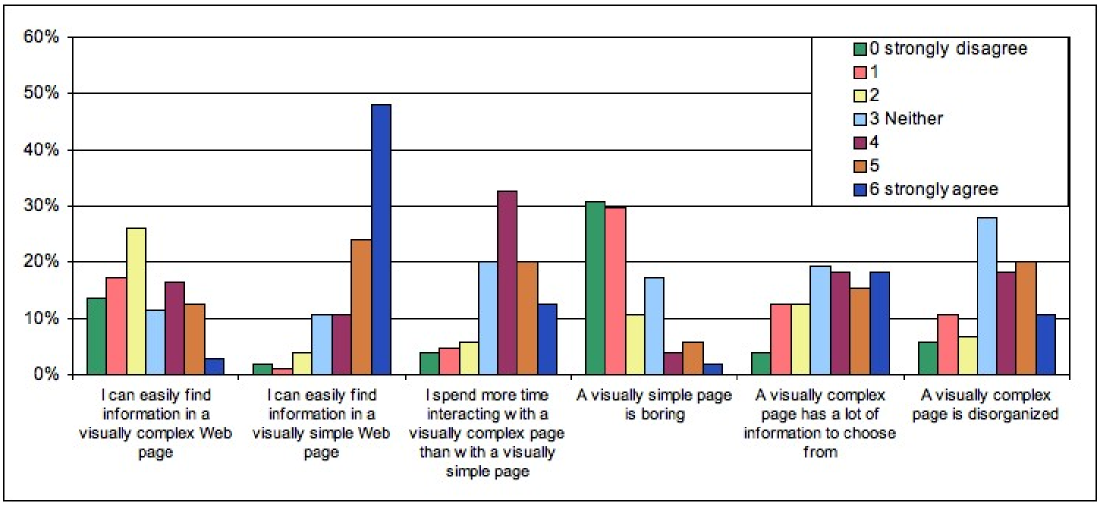


Figure 7.9: Participants Agreement on Feedback Statements

It is important to note that most of the participants disagreed with the statement that "a visually simple page is boring". However, our previous evaluations have shown that familiarity with a page is the most related factor in rating a page as interesting. Also, most participants were neutral on the statement that "a visually complex page is disorganized". On the other hand, the previous evaluation (Chapter 5) showed that complexity of the page was significantly and negatively related with the organization of a page.

Participants were also asked to define in their own words what a visually simple and visually complex page was. Comments were similar to those obtained from the previous evaluation (see Chapter 5). In summary, participants described a **visually simple** page as one with pale colours, clean, can easily find what they are looking for and has few images and links. They also described a visually simple page as organised, with clear layout, harmonious and muted colours and limited amount of different subjects and whose purpose can easily be understood. Some of the participants' comments include:

> *"The information hierarchies are easily distinguished and not blurred by an intrusive design. It does not necessarily refer to the amount of*

information but its organization on the page and how design enhances those categories. Reference.com has a lot of information but the layout makes it easy to scan and make decisions"

"Hierarchical structured information with visual focal points for the sites/service most important information"

"Guidance for the eye, sections "making sense" are recognizable, whitespace, calmness, less contrasts, peaceful"

On the other hand, participants described a **visually complex** page as one with a lot of information and categories, difficult to scan or skim text, not uniform in size and shape, has a large number of images, different colours, buttons, animations and generates an overall distraction. Some of the participants' comments include:

"lots of photos, images and colors; poor contrast between text and background, content not well grouped visually; too little white space to help define hierarchy and grouping. Layout seems haphazard; walls of text "

"Colors and layout compete. The hierarchy of information is uncertain, as if everything is demanding attention. It's not about having photos or images – it's about how the photos and image are used. Do they add value or compete for attention?"

"Eye has difficulty resting on any one focal point, not clear when focus should be and may be more likely to miss familiarization with all that site has to offer, but rather remember an image or a smaller subset of information"

A complete set of the users' comments can be found in the evaluation's technical report (see page **??**).

7.3 Discussion and Conclusions

The quantitative and qualitative analysis of the evaluation data has produced a prediction model that has formed the basis of an algorithm that can generate a complexity score for Web pages that will significantly match users' perception.

Our hypothesis, **H1**, stated that Visual Complexity rankings generated automatically by the tool are significantly and positively related with users' rankings. In other words, the tool can be used to predict users' visual complexity perception of a Web page at a significant and high correlation. The user evaluation provided data that support this hypothesis since it revealed positive and of high effect significance relationship between the users' visual complexity rankings of the page and the scores generated by the tool. In addition, a linear model was revealed that can estimate the users' ranks based on the algorithm value.

It is important to note that participants were consistent with their answers. Asking participants to see and rate each image twice initially raised issues about boredom effect. However, the results show that users' rankings were highly correlated and therefore reliable for both score versions. The time that they spent to give a score on the visual complexity level of a page was not significantly related with the page's level of complexity.

In addition, familiarity with the page did not reveal any significant correlation with the users' rankings or the algorithm scores but there were some exceptions. For example, the Yahoo page was ranked as very complex by the participants and most users were familiar with the page but the algorithm ranked it as one with a medium level of complexity. As discussed in Section 7.2.2 this could happen due to the fact that the algorithm does not account for familiarity. In addition, the algorithm is based on the structural elements of the page and it detects elements that the user can not always see as they are hidden in the underlying source code. For example, one should consider that the Yahoo page has an increased number of information and images. However, compared to the CNN page that was ranked as complex by the algorithm it is more organised and has less information.

By supporting our hypothesis, the prediction model (see Equation 6.1) can now be used to expand future research to better understand the relationship between structural elements and users' perception of a Web page based on the issues revealed above.

Even though we did not tested the tool with designer and visually impaired users, the tool can for now used by designers along with aDesigner to balance

Web page visual complexity with usability and accessibility as it was found to relate significantly and highly to user's rankings. A user prediction model was also generated (see Equation 7.1) through this user evaluation that can be used to 'correct' the algorithm score for the page's visual complexity as close to the user rank as possible; an evaluation that supported our last research objective (Q3 - see page 23).

Chapter 8

Conclusions and Future Work

Access to, and movement around, complex hypermedia environments, of which the Web is the most popular example, has long been considered an important and major issue in the Web design and usability field. With the rapid and constant advancement of technology, new ways are constantly being introduced to present information that leads to visually complex Web pages. Whilst it takes sighted users less than 5 seconds to visually glance through and grasp an overview of a Web page it can take visually impaired users much longer (in some cases over 2 minutes) to gain an overview of the page when it is read. This thesis addressed this interaction inequality and provided a framework that when followed can be used to design pages that are less visually complex and more accessible.

This chapter concludes the thesis by summarizing the ViCRAM project presented, appraising its contributions and significance, describing any outstanding issues and addresses future directions that could follow.

8.1 Thesis Overview

The visual complexity of Web pages has been described to affect the difficulty of using Web pages but also regarded as a subjective decision by the user [Harper *et al.* 2009]. Visually impaired users find many Web pages to be complex from an interaction standpoint. Indeed, studies have shown that if a page is too visually complicated visually impaired users will often not even try and interact with the content [Hoffos & Sharpless 1991; Faraday & Sutcliffe 1998] or give up after a short time [Lunn & Michailidou 2007; 2008]. Therefore, if it is possible to give a visually impaired user some notice of the expected complexity of the interaction required before link traversal then the time wasted on unproductive audio interaction, scanning, and glancing could be reduced. The project hypothesis was that by understanding sighted users' visual perception of Web page complexity we can understand the cognitive effort required for interaction with that page.

During this project, visual complexity was defined and information was provided in terms of complexity on how it can be used to benefit Web interaction and design. A series of user evaluations was conducted through ranking experiments, online questionnaires, user observations and eye tracking experiment. Results produced statistical models that based on the density and diversity of Web page structural elements (such as text, tables, links, and images) can predict sighted users' perception of Web page visual complexity. The framework was then implemented into the ACTF Eclipse framework by extending the aDesigner accessibility tool. For each Web page a complexity score that determines the page's level of visual complexity and an overlay heatmap that mimics a user's visual complexity perception by noting the areas that are most visually complex is generated.

A user evaluation was conducted that supported our hypothesis as the results revealed that the tool can significantly predict the level of visual complexity of a Web page. Therefore, users can have an initial perception of the visual layout of the page and designers can use this framework to balance Web page visual complexity with usability and accessibility. This is an important contribution to the Web accessibility area because by using visual complexity, an identifiable measure, as an implicit marker of cognitive load, Web pages can be designed that are easier to interact with, especially for visually impaired users.

8.2 Contributions

Throughout the literature it is shown that an accessible Website can be designed when designers and developers can understand how all users interact, their needs and technologies they use to access the information and when they follow the appropriate guidelines.

The literature examination coupled with a series of user evaluations showed a set of interesting facts: 1. the increased number of visual landmarks on a page can impede user navigation by increasing the time and effort to reach the information they seek; 2. the number of visual landmarks on a Web page affects the level of visual complexity in a positive manner, that is the more visual landmarks the more visually complex is perceived as by users; 3. visually impaired users, and specifically blind users, perform hypertext navigation in a similar fashion to sighted users with scanning navigation that corresponds to sighted users' eye movements, a finding that led us into investigating sighted user Web behaviour and perception to formulate our initial understanding of visual perception; 4. sighted user's eye movement revealed that gazing time increases and gaze order scatters as visual complexity of Web page increases as well.

It could be argued that these findings provided the basis into using eye tracking technologies and running evaluations with sighted users in order to understand user behaviour and perception for improving visually impaired user interaction. Even if robust conclusions could not be made, three key points were suggested that should be used by Web page designers and software developers to improve visually impaired users' Web experience. These points denote the areas that users tend to focus more and most, users' common browsing behaviour, menu design and location and how visual complexity affects Web browsing. These suggestions can be considered as guidelines for designing Web pages but also to transcode pages into simpler and more accessible ones.

The above findings supported part of the thesis Research Question 1 (Q1) stated in Chapter 1 on page 22. Specifically, the literature review along with the series of user evaluations showed that a relationship exists between visual complexity and Web interaction and between Web interaction and accessibility. Even though the relationship between visual complexity and accessibility could not be supported and further investigation needs to be conducted, during a formative technical accessibility evaluation [Mbipom 2008] results were taken and used in combination with this project to reveal a negative relationship between complexity and accessibility. That is, the more visually complex the page was ranked by users, the less accessible it was based on an automatic tool evaluation.

Mbipom [2008] describes a formative study conducted to investigate the relationship between visual aesthetics and Web accessibility, and also to determine if this relationship was significant enough to motivate further research. Technical accessibility evaluations were conducted for thirty Web pages that had previously been rated on their visual appearance by participants in this thesis (see Chapter 5) and by the Complexity tool (see Chapter 6.3).

Table 8.1 lists the complexity scores generated by the algorithm during the technical evaluation and the accessibility failures that the automatic tool generated during the study by Mbipom [2008]. In order to investigate towards this thesis first Research Question (Q1) a correlation test was run between these two data (except the pages that crashed during the algorithm run). The analysis revealed a positive and significant correlation ($r=.47$, $p<.05$) between the two data sets demonstrating and confirming our assumptions that visual complex pages and accessibility problems relate. These results do not act as the answer to this thesis research question (Q1) but as a support for future investigation based on these results. Due to time issues, it was not possible to run further accessibility tests but these results indicate that there is a relationship between visual complexity and Web accessibility. Issues such as accessibility automatic evaluation tools and visually impaired user evaluations have to be met before deriving more robust conclusions.

Design elements and techniques constantly change in order to create more attractive sites. Studies try to identify the design characteristics of the most effective sites, how these characteristics change over time and how users' cognitive load is affected in order to examine the effect of Web page presentation with user interaction. As explained above, visual complexity is found to affect user interaction. This project investigated how Web page presentation is related with visual complexity in order to be able to account for complexity.

By starting with the fact that Web page design is a combination of a large set of variables such as images, text, tables and links and vary in types such as colour, size, and position, the effect of user's complexity perception based on Web page presentation and explicitly the structural elements used to design a page was examined through user evaluations. This thesis demonstrated that visual complexity of Web pages depends on the presentation of the page's elements and by the density and diversity of the elements that are presented and overall layout of the page, a finding that supports this project's Research Question 2 (Q2) stated in Chapter 1 on page 23. The user evaluations conducted revealed a positive, significant and robust relationship between visual complexity of the page and its number of images, visible links, words and sections the page is grouped into (TLCs). In addition to their strong correlation, prediction models were identified using the number of images, words and TLCs as the variables in the equations.

Table 8.1: Algorithm Complexity Scores (see Section 6.3.1) and Accessibility Failure Score [Mbipom 2008]

PageID	Web Page	WCAG 1.0 Failures AAA	Algorithm Complexity Score
1	Amazon UK	8	5.56
2	AnnoteaProject	4	4.35
3	AutoTrader	6	6.24
4	BBC UK News	6	4.21
5	BBC UK	7	2.89
6	BloggerPostHQ	3	Crash
7	BloggerPostDE	5	6.72
8	Blogger Dashboard	5	3.74
9	Delicious	5	2.58
10	Ebay	7	2.75
11	Firefox	2	1.48
12	Flickr	5	0.71
13	GoogleSearch	7	3.28
14	GumTree	7	2.50
15	IMDB	8	5.68
16	InvisionFree	5	2.26
17	Job Centre	3	0.81
18	MegaUpload	7	Crash
19	MySpace	8	2.77
20	Orkut	5	0.58
21	Rapidshare	5	1.55
22	Rightmove	2	1.66
23	StudentNet	1	1.55
24	StudentNet SelfService	1	1.17
25	WAI	4	6.68
26	Wiki Result	5	Crash
27	Wikipedia	2	3.25
28	Yahoo UK	7	3.04
29	Yell	2	2.29
30	YouTube	8	10.51
rho - Correlation Coefficient		.47 $p<.05$	

Through the literature and continuing research in accessibility it is shown that disability simulation can help Web designers to imagine, recognize and understand the disabled Web access experience and developers can become more expert on identifying accessibility problems. This thesis also contributed into implementing a prototype simulation of visual complexity, the ViCRAM tool. It is not described as a disability simulator but when used along with the affiliated tools from the ACTF framework it can assist into designing less visually complex and more accessible Webpages.

The tool is developed based on the complexity prediction models so that it can automatically analyse a Web page with respect to its visual complexity. For each Web page, users and designers are provided with a complexity score that determines the page's level of visual complexity and an overaly heatmap that notes the areas that are most visually complex. A user and technical evaluation was performed and supported this project's third research question, Q3, stated in Chapter 1 on page 23 as a positive and of high effect significance relationship between the users' visual complexity rankings of the page and the scores generated by the tool was unveiled.

8.3 Outstanding Issues

Despite the contributions described above, there are some pending issues that even if they have been explored are not resolved or need further research and support.

Visual complexity and Web accessibility One of the unresolved issues of the project's Research Question 1 is the relationship between visual complexity and Web accessibility. The EiVAA project[1], using data from this project, examined this relationship. Technical accessibility evaluations were conducted on Web pages examined during a user evaluation from this project. The accessibility rankings were then analyzed in relation with the user complexity rankings collected in one of this project's user evaluation (Chapter 5). The analysis revealed a negative relationship indicating that as the visual complexity of Web page increases, the less accessible the page is ranked as. This, however, needs to be investigated further in order to examine false positives, through manual accessibility checks and user evaluations.

Heuristics As discussed in Chapter 6, being a prototype, the ViCRAM tool algorithms could be further improved if some issues were resolved. Specifically, the heuristics applied through the complexity algorithm could be enhanced to determine the TLC count more efficiently. Even if the TLC count depends on whether the page is coded following the standards, more heuristics could be applied in order to handle such cases. In addition, the heuristics could depend on the genre of the examined page. As demonstrated in Amitay *et al.* [2003] sites with a similar role and functionality

[1]EiVAA - http://hcw.cs.manchester.ac.uk/research/eivaa

exhibit similar structural patterns. Heuristics could be designed and applied for each genre.

Evaluation The user evaluation conducted focused on examining the complexity score provided by the tool. It is important to examine the effect that it has on designers as a tool. A user evaluation could be conducted on whether the tool can effectively assist designers in reducing the visual clutter on Web pages. By evaluating using the tool, the effect of the overlay heatmap can be also examined as it was not evaluated in this project.

8.4 Future Work

This research project provided a set of contributions but also has some outstanding issues that when attempted to resolved can offer new paths for future work.

Heuristics As discussed above, heuristics applied through the algorithms after resolving the above outstanding issues can be also improved both for the complexity and visualization algorithms. For the complexity algorithm heuristics could be derived that can efficiently determine the number of TLCs for each page depending on the page's type and level of underlying source code. The algorithm could be advanced by taking into account further weight factors for each of the variables (image, word count, TLC). For example, image size, purpose of use and whether is a dynamic image are three different factors for the image variable. A big and dynamic image causes more complexity than small and decorative ones. In order, though, to examine the effect of these factors further evaluation studies need to be performed especially using eye tracking technologies. Similarly with images, word count can have further weight factors. For example, words from headings, lists, menus or image captions can have different effects between each other and from paragraph text word count. Determining the correlation between visual attention and complexity perception could provide models that when implemented on top of the tool can advance the existing ViCRAM tool.

For the visualization algorithm the tool can be improved so that the user has the option of defining the number of grids the page's visualization will have. In addition, the grids could be regenerated when the users resizes the browser. Currently, these facilities are not provided but are feasible issues that could benefit the user when implemented.

User Evaluations Additional studies should be carried out, other than to determine further weight factors for improving the complexity heuristics. These should involve sighted and visually impaired users in order to determine the tool complexity prediction effectiveness for both type of users. For sighted users, eye tracking studies should be conducted in order to examine the prediction significance of the heatmaps generated by the eye tracking software and the heatmap generated based on the visualization algorithm. In this way a relationship would be examined between visual attention and complexity perception prediction of the tool.

Visually impaired users should be observed using the above sites and further qualitative data should be collected in order to examine the accessibility problems faced while interacting with visually complex and simple ranked pages. Audio output could be related with visual complexity and more credible conclusions will be made on whether pages that sighted users identify as visually complex are as complex for visually impaired users from accessibility perspective.

In addition, these user evaluations could be designed in a way to examine the effects of gender, age range, culture and social backgrounds on visual perception and interaction of Web pages.

Browser Plug-in The next stage for improving this research is by implementing the ViCRAM tool as a browser plug-in. Sighted and visually impaired users can be provided with the complexity score of each page they attempt to interact with. This will lead to further user evaluations into examining how effective is the complexity feedback provided to the user. The user evaluations would determine whether the complexity information influence interaction time and browsing behaviour.

Aesthetic Perception During the evaluation described in Chapter 5 user perception of visual aesthetics was also examined and prediction models derived. The tool implemented could also calculate the aesthetics terms but were not reported in the chapter as it was not one of the project's aims. Table I.1 lists the aesthetic values that the participants from the experiment described in Chapter 5 provided, the values the tool returned for the same pages set as described in Chapter 6.3 and their correlation coefficient. The only variables that were not significant are the Interesting and Clean scores. The interestingness level of the page, as it was discussed in Chapter 5, was highly correlated with the familiarity of the user which could be one of the reasons that the algorithm could not significantly predict the user Interesting rankings. In addition, a formative evaluation described by Mbipom [2008] showed that there is a linkage between visual aesthetics and Web accessibility which should be further investigated. User perception of Web page aesthetic characteristics with respect to clearness, cleanliness, organisation and overall aesthetic of a Web page is negatively and significantly related with the Web accessibility of the page measured by the WCAG 1.0 conformance levels. This study demonstrated that accessibility does not restrict aesthetically pleasing and attractive Web designs which acts as a support and enhances previous attempts into removing the misconception that accessibility restricts Web design creativity and aesthetics. Since the primary objective of the above study was to investigate if there was any significant relationship between visual aesthetic aspects of Web pages and their technical accessibility, external factors that could influence this relationship were not taken into account. These factors include terminology, designer expertise and user interaction. As it was discussed before, the aesthetics of Web pages need further investigation as the previous study acted as a pilot investigation into this area. The EiVAA project researches further the relationship between aesthetics, Web page perception and accessibility.

This thesis provided the basis for defining and understanding visual complexity of Web pages and sighted user perception which can assist the development of new hypotheses. A tool was presented that should assist designers to create more accessible Web pages and inform users with respect to the visual complexity level of the page they try to interact with. This research's contribution in the Web accessibility area can be enhanced by observing visually impaired users using the tool in order to examine the effect that the visual complexity prediction has on their interaction. However, as it was explained in the previous chapters, understanding the visual complexity, ergo the structural complexity, will have a direct benefit to blind users since an increased complexity for sighted users implies an increased complexity for blind users (see page 20). Hence, examining the relationship between visual complexity and the diversity and density of structural elements contained within a page can assist blind users to build their navigation model [Takagi *et al.* 2007].

The research presented in this thesis provided encouraging results into understanding the cognitive load required when interacting with Web and the basis into understanding the relationship between visually impaired and sighted user interaction. This project can be further advanced and one should consider this as supporting information into the foundation for enhancing Web accessibility, positive user experience and relating it with usability for all users.

Bibliography

AFB. *American Foundation for the Blind*. http://www.afb.org.

Amitay, Einat, Carmel, David, Darlow, Adam, Lempel, Ronny, & Soffer, Aya. 2003. The connectivity sonar: detecting site functionality by structural patterns. *Pages 38–47 of: HYPERTEXT '03: Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*. New York, NY, USA: ACM Press.

Asakawa, Chieko. 2005. What's the web like if you can't see it? *Pages 1–8 of: W4A '05: Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A)*. New York, NY, USA: ACM Press.

Asakawa, Chieko, & Itoh, Takashi. 1998. User interface of a Home Page Reader. *Pages 149–156 of: Assets '98: Proceedings of the third international ACM conference on Assistive technologies*. New York, NY, USA: ACM Press.

Asakawa, Chieko, & Takagi, Hironobu. 2000. Annotation-based transcoding for nonvisual web access. *Pages 172–179 of: Assets '00: Proceedings of the fourth international ACM conference on Assistive technologies*. New York, NY, USA: ACM Press.

Bechhofer, Sean, Harper, Simon, & Lunn, Darren. 2006. SADIe: Semantic Annotation for Accessibility. *Pages 101–115 of: ISWC '06: Proceedings of The 5th International Semantic Web Conference*. Springer Berlin / Heidelberg.

Brewer, Judy. 2004a (December). *How people with disabilities use the web*. http://www.w3.org/WAI/EO/Drafts/PWD-Use-Web/.

Brewer, Judy. 2004b. Web accessibility highlights and trends. *Pages 51–55 of: W4A '04: Proceedings of the 2004 international cross-disciplinary workshop on Web accessibility (W4A)*. New York, NY, USA: ACM Press.

BrookesTalk. 1998. http://www.brookes.ac.uk/schools/cms/research/speech/btalk.htm!

Brown, Andy, Jay, Caroline, & Harper, Simon. 2009. Audio presentation of auto-suggest lists. *Pages 58–61 of: W4A '09: Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A).* New York, NY, USA: ACM.

Brown, Silas S, & Robinson, Peter. 2001. A world wide web mediator for users with low vision. *CHI 2001: Workshop.*

Buhmann, Joachim M., Malik, Jitendra, & Perona, Pietro. 1999. Image Recognition: Visual Grouping, Recognition and Learning. **96**(25), 14203–14204.

Butler, Keith A. 1996. Usability engineering turns 10. *interactions*, **3**(1), 58–75.

Cater, Kirsten, & Chalmers, Alan. 2003. Maintaining Perceived Quality for Interactive Tasks. *Pages 231–239 of: SPIE-IS&T Electronic Imaging, Human Vision and Electronic Imaging VIII.* SPIE Vol. 5007.

Chen, Chaomei. 1997. Structuring and visualising the WWW by generalised similarity analysis. *Pages 177–186 of: HYPERTEXT '97: Proceedings of the eighth ACM conference on Hypertext.* New York, NY, USA: ACM.

Cove, J. F., & Walsh, B. C. 1988. Online text retrieval via browsing. *Info. Processing & Management*, **24**(1).

Coyne, Kara Pernice, & Nielsen, Jakob. 2001. *Beyond ALT Text: Making the Web Easy to Use for Users with Disabilities.* Tech. rept. Nielsen Norman Group.

Cutrell, Edward, & Guan, Zhiwei. 2007. What are you looking for?: an eye-tracking study of information usage in web search. *Pages 407–416 of: CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM.

DRC. 2004. *Disability Rights Commission. The Web: Access and Inclusion for Disabled People.* http://www.drc-gb.org/publicationsandreports/2.pdf.

Duchowski, Andrew T. 2002. *Eye Tracking Methodology: Theory and Practice.* Springer London Ltd; illustrated edition.

Education, BBC. *Betsie Home Page.* http://www.bbc.co.uk/education/betsie/.

EmacsSpeak. 2006. http://emacspeak.sourceforge.net/.

Fahle, M. 2004. Perceptual learning: A case for early selection. *Journal of Vision*, **4**(10), 879–890. http://journalofvision.org/4/10/4/.

Faraday, Pete, & Sutcliffe, Alistair. 1998. Making contact points between text and images. *Pages 29–37 of: MULTIMEDIA '98: Proceedings of the sixth ACM international conference on Multimedia*. New York, NY, USA: ACM Press.

Field, Andy. 2005. *Discovering Statistics Using SPSS*. SAGE Publications Ltd.

Fischer, S. 1996. *Multimedia Authoring*. Academic Press Professional, St. Louis, MO.

Freedman, M. 1984. Eye movement communication systems for computer control. *Pages 34–35 of: In Proceedings of the International Congress on Technology & Technology Exchange: Technology and the World Around Us*. IEEE Press.

Friedman, A. 1979. Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal for Experimental Psychology: General*, **108**, 316–355.

Furuta, Richard, Frank M. Shipman, III, Marshall, Catherine C., Brenner, Donald, & wei Hsieh, Hao. 1997. Hypertext paths and the World-Wide Web: experiences with Walden's Paths. *Pages 167–176 of: HYPERTEXT '97: Proceedings of the eighth ACM conference on Hypertext*. New York, NY, USA: ACM.

Germonprez, Matt, & Zigurs, Ilze. 2003. Causal factors for web site complexity. *Working Papers on Information Environments, Systems and Organizations*.

Goldberg, Joseph H., Stimson, Mark J., Lewenstein, Marion, Scott, Neil, & Wichansky, Anna M. 2002. Eye tracking in web search tasks: design implications. *Pages 51–58 of: ETRA '02: Proceedings of the symposium on Eye tracking research & applications*. New York, NY, USA: ACM Press.

Granka, Laura A., Joachims, Thorsten, & Gay, Geri. 2004. Eye-tracking analysis of user behavior in WWW search. *Pages 478–479 of: SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM Press.

GW-Micro. 2006. *GW Micro, Inc.* http://www.gwmicro.com/.

Hamker, Fred H. 2003. The reentry hypothesis: linking eye movements to visual perception. *Journal of Vision*, **3**(11), 808–816. http://journalofvision.org/3/11/14/.

Harper, Simon, & Bechhofer, Sean. 2005. Semantic triage for increase Web accessibility. *IBM Systems*, **44**(3), 637–648.

Harper, Simon, & Yesilada, Yeliz. 2007. editors, W4A '07: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A).

Harper, Simon, Goble, Carole, & Stevens, Robert. 2001. Web Mobility Guidelines for Visually Impaired Surfers. *Journal of Research and Preactice in Information Technology*, **33**(1), 30–41.

Harper, Simon, Yesilada, Yeliz, & Goble, Carole. 2004 (December). *Introduction - Special Issue on Accessible Hypermedia and Multimedia.*

Harper, Simon, Michailidou, Eleni, & Stevens, Robert. 2009. Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Trans. Appl. Percept.*, **6**(2), 1–18.

Heaps, C., & Handel, C. H. 1999. Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 299–320.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. 1999. The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing. *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 210–228.

Henter-Joyce. 2006. *Freedom Scientific Inc.* http://www.freedomscientific.com/.

Herbert, David A. 1988. *The Method of Paired Comparisons.* Oxford University Press, USA.

Heylighen, F. 1997. *The Growth of Structural and Functional Complexity during Evolution.* Springer, New York.

Hoffmann, Ralf, & Krauss, Kirstin. 2004. A critical evaluation of literature on visual aesthetics for the web. *Pages 205–209 of: 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.* , Republic of South Africa: South African Institute for Computer Scientists and Information Technologists.

Hoffos, S., & Sharpless, G. 1991. *The CD-I designers guide*. McGraw-Hill.

Horton, Sarah. 2006. Designing beneath the surface of the web. *Pages 1–5 of: W4A '06: Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A)*. New York, NY, USA: ACM.

IBM. *Ease of Use: Web design guidelines*. http://www-306.ibm.com/ibm/easy/eou_ext.nsf/publish/572.

IBM. 2006. *Home Page Reader*. http://www-306.ibm.com/able/index.html.

Ivory, Melody Y., & Hearst, Marti A. 2001. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.*, **33**(4), 470–516.

Ivory, Melody Y., & Megraw, Rodrick. 2005. Evolution of web site design patterns. *ACM Trans. Inf. Syst.*, **23**(4), 463–497.

Ivory, Melody Y., Sinha, Rashmi R., & Hearst, Marti A. 2000. Preliminary Findings on Quantitative Measures for Distinguishing Highly Rated Information-Centric Web Pages. *In: Proceedings of 6th Conference on Human Factors and the Web*.

Ivory, Melody Y., Sinha, Rashmi R., & Hearst, Marti A. 2001. Empirically validated web page design metrics. *Pages 53–60 of: CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM Press.

Ivory, Melody Y., Mankoff, Jennifer, & Le, Audrey. 2003. Using Automated Tools to Improve web site usage by users with diverse abilities. *IT&Society*, **1**(3), 195–236. http://www.ITandSociety.org.

Jacob, Robert J. K. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.*, **9**(2), 152–169.

Jacob, Robert J. K. 1995. *Eye tracking in advanced interface design.* New York, NY, USA: Oxford University Press, Inc.

Jay, Caroline, Stevens, Robert, Glencross, Mashhuda, & Chalmers, Alan. 2006. How people use presentation to search for a link: expanding the understanding of accessibility on the web. *Pages 113–120 of: W4A: Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A).* New York, NY, USA: ACM Press.

Jeffries, Robin, Miller, James R., Wharton, Cathleen, & Uyeda, Kathy. 1991. User interface evaluation in the real world: a comparison of four techniques. *Pages 119–124 of: CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press.

Josephson, Sheree, & Holmes, Michael E. 2002. Visual attention to repeated Internet images: testing the scanpath theory on the world wide web. *Pages 43–49 of: ETRA '02: Proceedings of the symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press.

Karat, Claire-Marie, Campbell, Robert, & Fiegel, Tarra. 1992. Comparison of empirical testing and walkthrough methods in user interface evaluation. *Pages 397–404 of: CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press.

Kolb, Helga, Fernandez, Eduardo, & Nelson, Ralph. 2005. *WebVision: The Organization of the Retina and Visual System.* John Moran Eye Center, University of Utah, USA.

Krauss, Kirstin. 2004. Visual aesthetics and its effect on communication intent: a theoretical study and website evaluation. *In: Proceedings of the Southern African Computer Lecturers Association (SACLA).*

Lavie, Talia, & Tractinsky, Noam. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum.-Comput. Stud.*, **60**(3), 269–298.

Lohse, G. L., & Johnson, E. J. 1996. A Comparison of Two Process Tracing Methods for Choice Tasks. *Organizational Behavior and Human Decision Processes*, **68**(1), 28–43.

Lunn, Darren. 2009 (November). *Towards Behaviour-Driven Transcoding of Web Content Through an Analysis of User Coping Strategies*. A Thesis Submitted to the University of Manchester for the Degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences, University of Manchester, School of Computer Science, Kilburn Building, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK.

Lunn, Darren, & Michailidou, Eleni. 2007 (July). *Observational Notes Acquired from Henshaws' Skillstep to Success Class: Observation Period 1*. Technical Report SADIe Technical Report 5. The University of Manchester, School of Computer Science. SADIe Technical Report 5.

Lunn, Darren, & Michailidou, Eleni. 2008 (February). *Observational Notes Acquired from Henshaws' Skillstep to Success Class: Observation Period 2*. Technical Report SADIe Technical Report 8. The University of Manchester, School of Computer Science. SADIe Technical Report 8.

Lunn, Darren, Bechhofer, Sean, & Harper, Simon. 2008. A user evaluation of the SADIe transcoder. *Pages 137–144 of: Assets '08: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. New York, NY, USA: ACM.

Mankoff, Jennifer, Fait, Holly, & Tran, Tu. 2005. Is your web page accessible?: a comparative study of methods for assessing web page accessibility for the blind. *Pages 41–50 of: CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM Press.

Marks, Linn. 1994. Integrative multimedia design. *Pages 383–384 of: CHI '94: Conference companion on Human factors in computing systems*. New York, NY, USA: ACM.

Mbipom, Grace. 2008 (August). *Examining the Relationship Between Visual Aesthetics and Web Accessibility: A Formative Study*.

McCarthy, John D., Sasse, Angela M., & Riegelsberger, Jens. 2003. Could I have the menu please? An eye tracking study of design conventions. *Pages 8–12 of: Proceedings of HCI2003*.

McConkie, George W., & Currie, Christopher B. 1996. Visual Stability across Saccades While Viewing Complex Pictures. *Journal of Experimental Psychology: Human Perception and Performance*, **22**(3), 563–581.

Merriam-Webster. 2006. *Merriam-Webster Online Dictionary.* (Accessed January 30, 2006).

Michailidou, Eleni. 2005. *Metrics of Visual Complexity.* M.Phil. thesis, Computer Science.

Nagao, Katashi, Shirai, Yoshinari, & Squire, Kevin. 2001. Semantic Annotation and Transcoding: Making Web Content More Accessible. *IEEE MultiMedia*, **8**(2), 69–81.

Nguyen, Kevin K., & Petty, Linda S. 1996. *Web browsing through Adaptive Technology: A Consumer Information Resource.* Adaptive Technology Resource Centre (ATRC), University of Toronto, Ontario, Canada. http://www.utoronto.ca/atrc/rd/library/papers/consumer.html.

Nielsen, Jacob. *Usability and Web Design.* http://www.useit.com.

Nielsen, Jakob. 1994. *Heuristic evaluation.* John Wiley and Sons.

Oliva, Aude. 2005. Gist of a Scene. *In Neurobiology of Attention*, 251–256.

Oliva, Aude, Mack, Michael L., Shrestha, Mochan, & Peeper, Angela. 2004. Identifying the perceptual dimensions of visual complexity of scenes. *In: Proceedings of the 26th Annual Meeting of the Cognitive Science Society.*

Outing, Steve, & Ruel, Laura. 2006. *Eyetrack III: What We Saw Through Their Eyes.* A Project of the Poynter Institute, http://www.poynterextra.org/eyetrack2004/index.htm (Accessed: 20 January 2006).

Paciello, Michael G. 2000. *Web Accessibility for People with Disabilities.* C M P Books.

Pan, Bing, Hembrooke, Helene A., Gay, Geri K., Granka, Laura A., Feusner, Matthew K., & Newman, Jill K. 2004. The determinants of web page viewing behavior: an eye-tracking study. *Pages 147–154 of: ETRA'2004: Proceedings* *of the Eye tracking research & applications symposium on Eye tracking research & applications*. New York, NY, USA: ACM Press.

Paul J Murphy, Anne L Duncan, Alastair J Glennie, Paul C Knox Paul J Murphy, Anne L Duncan, Alastair J Glennie, Paul C Knox Paul J Murphy, Anne L Duncan, Alastair J Glennie, & Knox, Paul C. 2001. The effect of scleral search coil lens wear on the eye. *British Journal of Ophthalmology*, 332–335.

Plessers, Peter, Casteleyn, Sven, Yesilada, Yeliz, Troyer, Olga De, Stevens, Robert, Harper, Simon, & Goble, Carole. 2005. Accessibility: a Web engineering approach. *Pages 353–362 of: WWW '05: Proceedings of the 14th international conference on World Wide Web*. New York, NY, USA: ACM Press.

Potter, M. C. 1976. Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 509–522.

Rayner, Keith. 1998. *Eye movements in reading and information processing: 20 years of research*. Psychological Bulletin, vol. 124, no. 3.

RetroAccess. 2002. *AccessEnable*. http://nwmail.washtenaw.cc.mi.us/ lmishler/Project3/products.html (Accessed February 2006).

Reuter, B. W., & Schenck, U. 1985. Data acquisition in visual screening of microscopical samples. *Pages 206–209 of: Proceedings of Image Science '85*. Acta Polytechnica Scandinavica, Applied Physics Series.

Richardson, Daniel C., & Spivey, Michael J. 2004. *Eye-Tracking: Research Areas and Applications*.

RNIB. 2006. *Royal National Institute of the Blind*. http://www.rnib.org.uk.

Robins, David, & Holmes, Jason. 2008. Aesthetics and credibility in web site design. *Information Processing and Management*, **44**(1), 386–399.

Russo, Edward J., & Leclerc, Francs. 1994. An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, **21**(2).

Sanders, M. G., Simmons, R. R., & Hoffman, M. A. 1979. Visual Workload of the Copilot /Navigator During Terrain Flight. *Human Factors*, **21**(2), 369–383.

Section508. 2006. http://www.section508.gov/.

Sullivan, Terry, & Matson, Rebecca. 2000. Barriers to use: usability and content accessibility on the Web's most popular sites. *Pages 139–144 of: CUU '00: Proceedings on the 2000 conference on Universal Usability*. New York, NY, USA: ACM Press.

Takagi, Hironobu, Asakawa, Chieko, Fukuda, Kentarou, & Maeda, Junji. 2004. Accessibility designer: visualizing usability for the blind. *Pages 177–184 of: Assets '04: Proceedings of the 6th international ACM SIGACCESS conference on Computers and accessibility*. New York, NY, USA: ACM Press.

Takagi, Hironobu, Saito, Shin, Fukuda, Kentarou, & Asakawa, Chieko. 2007. Analysis of navigability of Web applications for improving blind usability. *ACM Trans. Comput.-Hum. Interact.*, **14**(3), 13.

Tobii. 2006. *Tobii Eye Tracker*. http://www.tobii.com/.

Tongue, Thomas, & Inc., Imagiware. 1997. *DoctorHTML*. http://www2.imagiware.com/RxHTML/ (Accessed February 2006).

UsableNet-Inc. *Lift Machine*. http://www.usablenet.com (Accessed February 2006).

W3C. 1998. *The W3C CSS Validator*. http://jigsaw.w3.org/css-validator/README.html (Accessed February 2006).

W3C. 1999. *Web Content Accessibility Guidelines 1.0*. http://www.w3.org/TR/WAI-WEBCONTENT/.

W3C. 2000. *The W3C Markup Validation Service*. http://validator.w3.org/ (Accessed February 2006).

W3C. 2006. *World Wide Web Consortium*. http://www.w3.org.

WAI. 2005. *Evaluation, Repair, and Transformation Tools for Web Content Accessibility*. http://www.w3.org/WAI/ER/existingtools.html.

WAI. 2006. *Web Accessibility Initiative*. http://www.w3.org/WAI.

WatchFire. 2006 (February). *Bobby*. http://bobby.watchfire.com, Now Closed - Part of IBM Rational Software.

Whang, Yonghyun, Jung, Changwoo, Kim, Jihong, & Chung, Sungkwon. 2001. WebAlchemist: A Web Transcoding System for Mobile Web Access in Handheld Devices. vol. 4534.

Yarbus, Alfred L. 1967. *Eye Movements and Vision.* New York: Plenum Press.

Yesilada, Yeliz. 2005. *Annotation and Transformation of Web Pages to Improve Mobility for Visually Impaired Users.* Ph.D. thesis, The University of Manchester.

Yesilada, Yeliz, Stevens, Robert, & Goble, Carole. 2003a. A foundation for tool based mobility support for visually impaired web users. *Pages 422–430 of: WWW '03: Proceedings of the 12th international conference on World Wide Web.* New York, NY, USA: ACM Press.

Yesilada, Yeliz, Harper, Simon, Goble, Carole, & Stevens, Robert. 2003b. Ontology Based Semantic Annotation for Enhancing Mobility Support for Visually Impaired Web Users. *In: K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation.*

Yesilada, Yeliz, Harper, Simon, Goble, Carole, & Stevens, Robert. 2004. Screen Readers Cannot See (Ontology Based Semantic Annotation for Visually Impaired Web Travellers). *Pages 445–458 of:* Koch, Nora, Fraternali, Piero, & Wirsing, Martin (eds), *Web Engineering - 4th International Conference, ICWE 2004 Proceedings (LNCS 3140).* Springer.

Yesilada, Yeliz, Stevens, Robert, Harper, Simon, & Goble, Carole. 2007. Evaluating DANTE: Semantic Transcoding for Visually Disabled Users. *The special issue of TOCHI on Web Accessibility*, **14**(3), 0–0.

Yesilada, Yeliz, Jay, Caroline, Stevens, Robert, & Harper, Simon. 2008. Validating the Use and Role of Visual Elements of Web Pages in Navigation with an Eye-Tracking Study. *In: WWW '08.*

Zajicek, M., Powell, C., Reeves, C., & Griffiths, J. 1998. Web Browsing for the Visually Impaired. *Pages 161–169 of: CIFIP/ICCHP '98 6th Int. Conf. on Computers Helping People with Special Needs.*

Zajicek, Mary, Powell, Chris, & Reeves, Chris. 1999. Evaluation of a world wide web scanning interface for blind and visually impaired users. *Pages 980–984* *of: Proceedings of the HCI International '99 (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

Zettl, H. 1999. *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth Publishing Company, USA.

Appendix A

Glossary

The following glossary contains brief definitions of the statistical terms used throughout this thesis. The definitions were taken from Field [2005].

Adjusted R^2: a measure of the loss of predictive power of shrinkage in regression. The value tells us how much variance in the outcome would be accounted for if the model had been derived from the population from which the sample was taken.

Analysis of variance (ANOVA): a statistical procedure that uses the F-ratio to test the overall fit of a linear model. In experimental research this linear model tends to be defined in terms of group means and the resulting ANOVA is therefore an overall test of whether group means differ.

Boredom effect: refers to the possibility that performance in tasks may be influenced (the assumption is a negative influence) by boredom/lack of concentration if there are many tasks, or the task goes on for a long period of time.

Confounding variable: a variable (that we may or may not have measured) other than the predictor variables in which we're interested that potentially affects an outcome variable.

Durbin-Watson: tests for serial correlations between errors in regression models. Specifically, it tests whether adjacent residuals are correlated, which is useful in assessing the assumption of independent errors. The test statistic can vary between 0 and 4 with a value of 2 meaning that the residuals are uncorrelated. A value greater than 2 indicates a negative correlation between adjacent residuals, whereas a value below 2 indicates a positive correlation. The size of the Durbin-Watson statistic depends upon the number of predictors in the model and the number of observations. As a very conservative rule of thumb, values less than 1 or greater than 3 are definitely cause for concern; however, values closer to 2 may still be problematic depending on your sample and model.

F-ratio: A test statistic with a known probability distribution (F-distribution)

Huynh-Feldt correction: an estimate of the departure from sphericity. The maximum value is 1 (the data completely meet the assumption of sphericity). Values below this indicate departures from sphericity and are used to correct the degrees of freedom associated with the corresponding F-ratios by multiplying them by the value of the estimate. It is less conservative than the Greenhouse-Geisser estimate, but some say it is too liberal.

Independent t-test: a test using the t-statistic that establishes whether two means collected from independent samples differ significantly.

Mann-Whitney test: A non-parametric test that looks for differences between two independent samples. That is, it tests whether the populations from which two samples are drawn have the same location. It is functionally the same as Wilcoxon's rank-sum test, and both tests are non-parametric equivalents of the independent t-test.

Mauchly's Test: a test of the assumption of sphericity. If this test is significant then the assumption of sphericity has not been met and an appropriate correction must be applied to the degrees of freedom of the F-ratio in repeated-measures ANOVA. The test works by comparing the variance-covariance matrix of the data to an identity matrix, if the variance-covariance matrix is a scalar multiple of an identity matrix then sphericity is met.

Mean: A simple statistical model of the centre of a distribution of scores. A hypothetical estimate of the 'typical' score.

Mode: the most frequently occurring score in a set of data.

Multicollinearity: a situation in which two or more variables are very closely linearly related.

Multiple Regression: A regression analysis that is employed to account for (predict) the variance in an interval dependent, based on linear combinations of interval, dichotomous, or dummy independent variables. Multiple regression can establish that a set of independent variables explains a proportion of the variance in a dependent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables (by comparing beta weights). The multiple regression equation takes the form y = b1x1 + b2x2 + ... + bnxn + c. The b's are the regression coefficients, representing the amount the dependent variable y changes when the corresponding independent changes 1 unit. The c is the constant, where the regression line intercepts the y axis, representing the amount the dependent y will be when all the independent variables are 0. The standardized version of the b coefficients are the beta weights, and the ratio of the beta coefficients is the ratio of the relative predictive power of the independent variables.

Non-parametric tests: a family of statistical procedures that do not rely on the restrictive assumptions of parametric tests. In particular they do not assume that data come from a normal distribution.

Normal distribution: a probability distribution of a random variable that is known to have certain properties. It is perfectly symmetrical (has a skew of 0), and has a kurtosis of 0.

One-tailed test: a test of a directional hypothesis (a hypothesis with a stated direction of the relationship between the two variables).

Pairwise comparison: comparisons of pairs of means.

Parametric test: a test that requires data from one of the large catalogue of distributions that statisticians have described. Normally this term is used for parametric tests based on the normal distribution, which require four basic assumptions that must be met for the test to be accurate: normally distributed data (see normal distribution), homogeneity of variance, interval or ratio data, and independence.

Pearson's correlation coefficient: or Pearson's product-moment correlation coefficient, to give it its full name, is a standardized measure of the strength of relationship between two variables. It can take any value from .1 (as one variable changes, the other changes in the opposite direction by the same amount), through 0 (as one variable changes the other doesn't change at all), to +1 (as one variable changes, the other changes in the same direction by the same amount).

***Post Hoc* test:** a set of comparisons between group means that were not thought of before data were collected. Typically these tests involve comparing the means of all combinations of pairs of experimental conditions. To compensate for the number of tests conducted, each test uses a strict criterion for significance. As such, they tend to have less power than planned contrasts. They are usually used for exploratory work for which no firm hypotheses were available on which to base planned contrasts.

Practice effect: refers to the possibility that a participant's performance in a task may be influenced (positively or negatively) if they repeat the task because of familiarity with the experimental situation and/or the measures being used.

Repeated measures ANOVA: An analysis of variance conducted on any design in which the independent variable(s) have all been measured using the same participants in all conditions (in a within-subjects design).

Residuals: The difference between the value a model predicts and the value observed in the data on which the model is based. When the residual is calculated for each observation in a data set the resulting collection is referred to as the residuals.

Sidak adjustment: a slightly less conservative variant of a Bonferroni correction.

Spearman's correlation coefficient: a standardized measure of the strength of relationship between two variables that does not rely on the assumptions of a parametric test. It is Pearson's correlation coefficient performed on data that have been converted into ranked scores.

Sphericity: a less restrictive form of compound symmetry which assumes that the variances of the differences between data taken from the same participant (or other entity being tested) are equal. This assumption is most commonly found in repeated-measures ANOVA but applies only where there are more than two points of data from the same participant.

Standard deviation: an estimate of the average variability (spread) of a set of data measured in the same units of measurement as the original data. It is the square root of the variance.

Standard Error (SE): the standard deviation of the sampling distribution of a statistic. For a given statistic (e.g. the mean) it tells us how much variability there is in this statistic across samples from the same population. Large values, therefore, indicate that a statistic from a given sample may not be an accurate reflection of the population from which the sample came.

Stepwise regression: a method of multiple regression in which variables are entered into the model based on a statistical criterion (the semi-partial correlation with the outcome variable). Once a new variable is entered into the model, all variables in the model are assessed to see whether they should be removed.

t-Statistic: Student's t is a test statistic with a known probability distribution (the t-distribution). In the context of regression it is used to test whether a regression coefficient b is significantly different from zero; in the context of experimental work it is used to test whether the differences between two means are significantly different from zero (see also independent t-test and dependent t-test).

Variance inflation factor (VIF): a measure of multicollinearity. The VIF indicates whether a predictor has a strong linear relationship with the other predictor(s). A value of 10 is a good value at which to worry, where other statisticians suggest that if the average VIF is greater than 1, then multicollinearity may be biasing the regression model.

Weights: a number by which something (usually a variable in statistics) is multiplied. The weight assigned to a variable determines the influence that variable has within a mathematical equation: large weights give the variable a lot of influence.

Wilcoxon signed-rank test: a non-parametric test that looks for differences between two related samples. It is the non-parametric equivalent of the related t-test.

Within-subjects design: An experimental design in which different conditions are applied on the same subjects and the data are related. For example, in psychology, this would mean that the same people take part in all experimental conditions.

Appendix B

Eye Tracking Experiment

Associated materials during the eye tracking experiment data analysis described in Chapter 3.

B.1 Hotspot and Gaze Plot Analysis

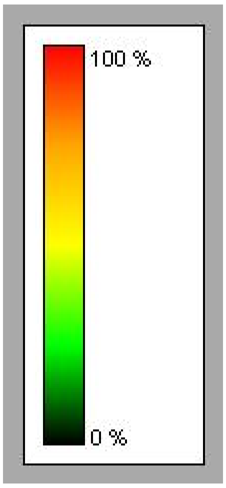
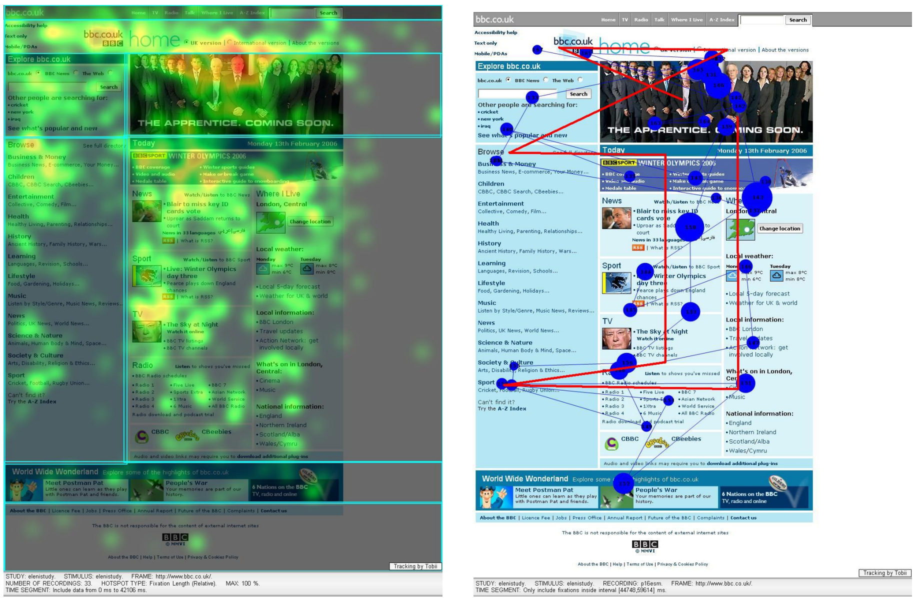


Figure B.1: Heatmap Legend

Hotspot analysis is a way of visualizing the gaze behaviour of an entire group of recordings by representing the areas of the screen receiving either more fixations or receiving the longest dwell time. A hotspot plot consists of the stimuli as background image and a hotspot mask superimposed on top of this. The hotspot mask consists of a black background with high-lighted around points where participants have been looking. The closer the highlights are to red the more fixations occurred in an area of the web page, with intensity decreasing with movement down the spectrum, as Figure B.1 shows.

The gaze plots that the software provides for each subject, displays a static view of the gaze data for each image of the stimuli and is a useful tool when visualizing scan paths. Gaze plots can also provide feedback on the order of fixation because it gives numbers for each fixation in order. This section gives the hotspot analysis for each web page that was used during the experiment, including the page's elements that were viewed first, the most, least or not at all.

B.1.1 BBC UK



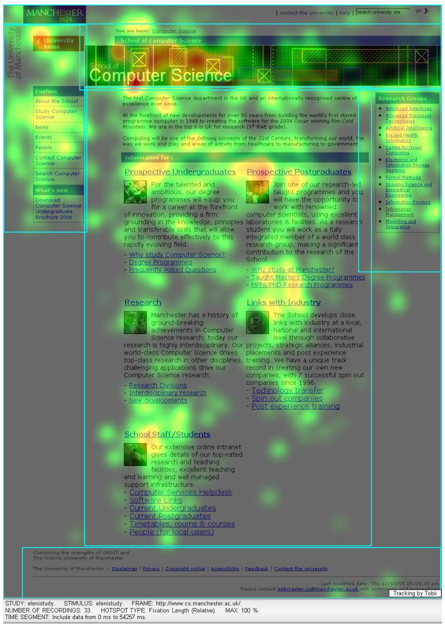
(a) Hotspot *(b) Gaze Plot*

Figure B.2: Eye Tracking - BBC

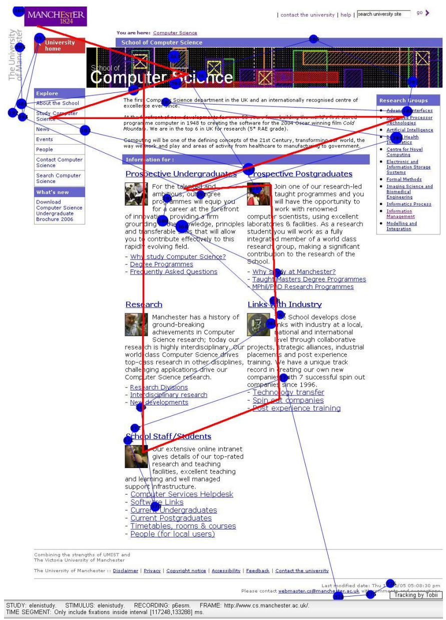
Glancing through the BBC web page (Figure B.2(a)), almost 90% of the subjects fixated on the big picture of an upcoming show. With more than half of the participants, the logo of the BBC's page had the next most fixation duration along with the headings of the news. About half of the participants did use the scroll bar to glance through the whole page but only a few fixated on random places on the second half of the page, with one quarter of the participants fixating on the footnote. It is important to note the difference between the menu and the main content of the page. Even if both areas have menu structure, people gazed through the keywords on the left menu and on the headings of the main content of today's news.

As Figure B.2(b) shows, along with time, subjects usually looked first on the big picture and then on the logo above. The left hand side menu was next, followed by today's news.

B.1.2 Computer Science Department



(a) Hotspot



(b) Gaze Plot

Figure B.3: Eye Tracking - Computer Science

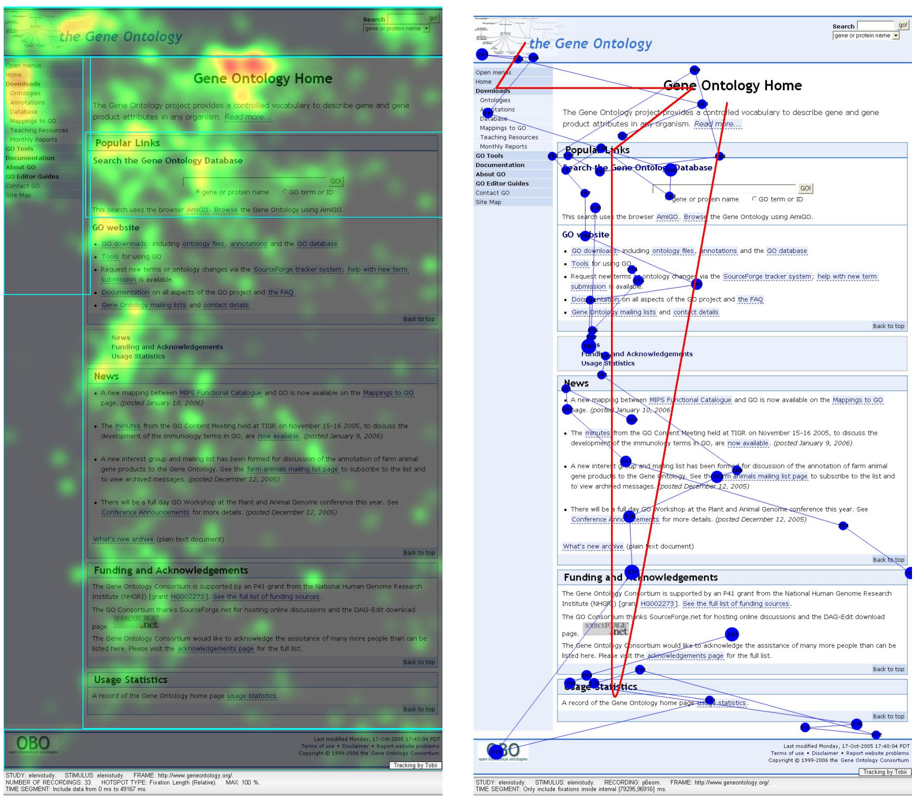
Hotspot analysis on the Computer Science Department web page (Figure B.3(a)) reveals that all the participants had the most fixation (100%) on the Computer Science title/picture and then on the right and left menus. However, the fixations on both of the menus were about 70% on the first two links and about 30% on the rest. The subjects also glanced through the rest of the page fixating more on the headings of each category with few short fixations on the links provided for every category. While scrolling and reading through the main content, the participants looked at the first two horizontal news items, then on the right column and when scrolling was done with no more right column content, moved to the left column (in the main content).

Analyzing the hotspots with respect to time and observing the gaze plots, as Figure B.3(b) shows, the participants first gazed on the page's logo/picture, then glanced through the department's description, followed by the left and right menus and finally the main content.

B.1.3 Gene Ontology

The title "Gene Ontology Home" attracted all subjects while reading the page with 100% fixation length (Figure B.4(a)). The first three links of the menu had the next highest fixation with about 80%, followed by the page's graphic on the top left corner. The rest of the page was randomly gazed through it with the longest and most fixations (60%) on the headings of the first two sections of the main content.

The subjects first gazed at the Gene Ontology's page title, then on the small graphic on the top left corner and then on the search box. The menu was then gazed followed by short fixations all over the page (Figure B.4(b)).



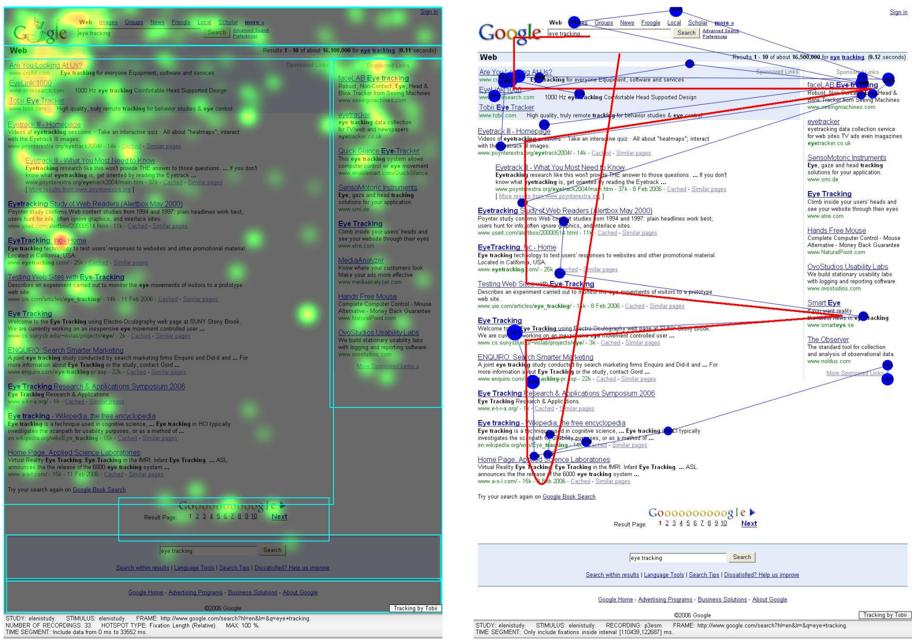
(a) Hotspot (b) Gaze Plot

Figure B.4: Eye Tracking - Gene Ontology

B.1.4 Google Search Results

The longer fixations on the google search results' page (Figure B.5(a)) were the google logo, the keyword on the search box and the first two headings of the results. Even though these elements had the longer fixations, they did not have the maximum (100%) fixation. The sponsored links column earned about 50% of the fixation length and as the hotspot shows, people tended to glance through the rest of the results with at least one gaze, either at the heading or at the description. The number of pages with results also earned about 40% of the fixation length.

It is interesting to note that when reaching the google results page (see Figure B.5(b)), the participants first looked on the first two results with fixations around their description text. Then they looked on the google's logo followed by the keyword on the search box. The sponsored links column was next followed by random fixations around the whole result page. Within miliseconds the participants gazed vertically through the whole list of the results.



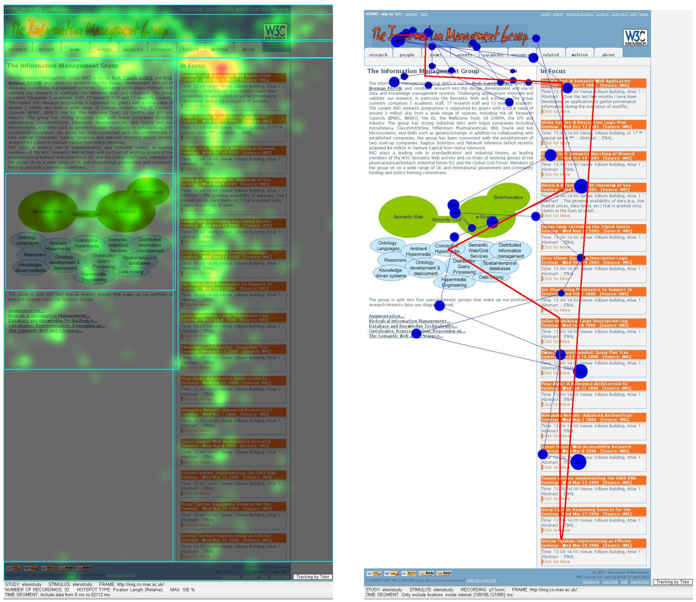
(a) Hotspot *(b) Gazeplot*

Figure B.5: Eye Tracking - Google Search Results

B.1.5 Information Management Group

Browsing the Information Management Group (IMG) web page (Figure B.6(a)), people fixated most on the right hand menu which consists of small boxes. The longer fixations were at the first title of the seminar column, specifically on the first box. The next longer fixations were the IMG's logo with 60% of the relative fixation length along with parts of the right hand side column.

The participants' gaze plots showed (Figure B.6(b)) that they tended to look first at the horizontal menu, specifically the events link, and the IMG logo. After a few fixations on the IMG's group description, the participants fixated on the "In Focus" heading of the right hand column and then on its first box. After glancing on the first three seminar descriptions, the subjects fixated longer on the group's picture and skipped a couple of seminars on the right hand side. The participants, then, tended to glance through the rest of the seminars following a straight vertical line.



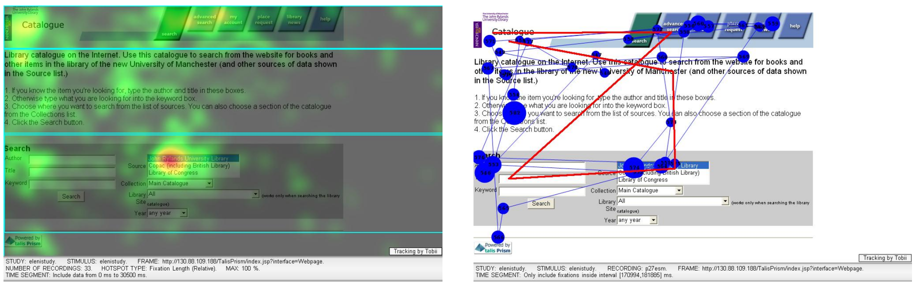
(a) Hotspot (b) Gaze Plot

Figure B.6: Eye tracking - IMG

B.1.6 John Rylands Library Catalogue

The John Rylands Library Catalogue web page (Figure B.7(a)) had only one place that attracted more than 50% of the fixation duration: the source of the catalogue search. That form element had around 98% of the fixation length which shows that the participants spent a longer time fixating on the description of the sources origin. The next longest fixations were the first three menu elements along with the “Catalogue” title. The rest of the page was just gazed through without any signifigant fixations.

As Figure B.7(b) shows, the subjects tended to first look on the horizontal menu, more specifically on the first link, and then move their eyes towards the search form, on the three form descriptions. Then they fixated for a long time on the source form description and next they glanced through the text of the page, which is the description of the web page. For the rest of the browsing, they gazed between the menu and the source form fixating longer on the John Rylands University Library source.



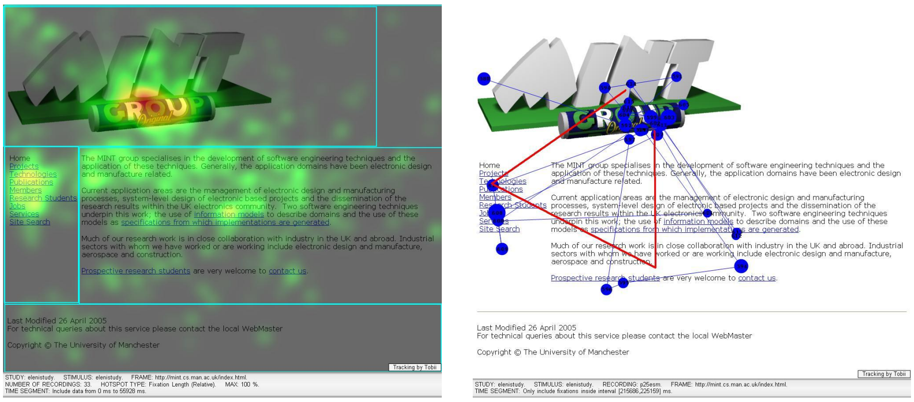
(a) *Hotspot* (b) *Gazeplot*

Figure B.7: Eye Tracking - John Rylands Library

B.1.7 Mint Group

By first glancing on the Mint Group’s web page, one can see that the group’s logo takes about half of the page. Hence, it is not surprising that the hotspot analysis (Figure B.8(a)) reveals that 98% of the fixation length is on the logo and mostly on the “Group Original” text. About 50% of the gaze time is on the left menu, while the main content has fixations over the text without any specifically long gaze. It is also important to note that within the first five seconds of the gaze time, the participants went through the whole page.

The analysis also shows that the first fixations of the participants are on the logo, then on the menu followed by fixations on the main content where participants fixated more on the links that are within the text (Figure B.8(b)).



(a) Hotspot (b) Gaze plot

Figure B.8: Eye Tracking - Mint Group

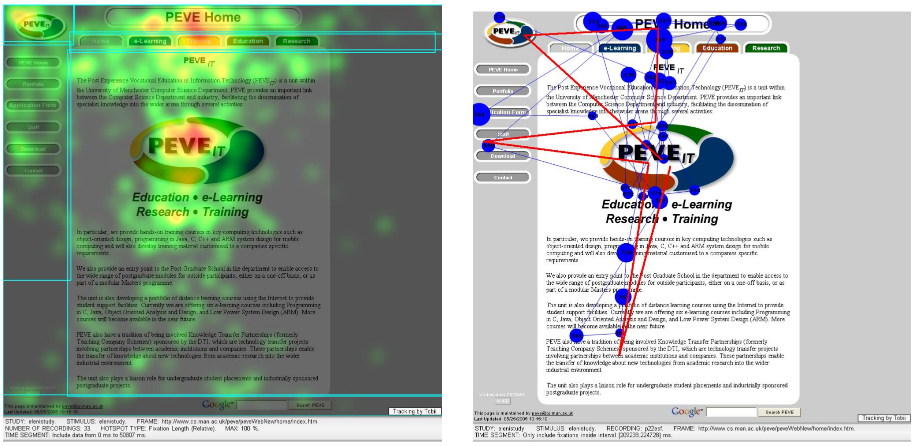
B.1.8 Peve IT Group

The hotspot analysis on the Peve Home page (Figure B.9(a)) reveals that the subjects fixated most and longest on the Peve graphic in the middle of the page (97%) and on the title of the page, "PEVE Home" (90%). The logo on the top left corner also attracted long fixations (57%). The horizontal and vertical menus were gazed through with small fixations, along with the rest of the main content of the page without fixations on specific parts.

Within the first two seconds (see also Figure B.9(b)), the participants first fixated on the graphic in the centre of the page followed by the title of the page and then on the top left corner logo. Next they gazed on the left menu fixating only on a couple of the menu links. Then they fixated on the centre graphic and then on the horizontal menu. Next, some participants gazed through the main content of the page and others on the big graphic's text. The final seconds of the browsing session were spent on the top half of the context and on the horizontal menu. It is important to note that not everybody gazed through the whole page, with around 10% of the subjects fixated around the bottom half of the page.

B.1.9 Vodafone

The longer fixations on the Vodafone web page were at the centre of the page, where a flashing image attracted almost 90% of the participants. 75% of the relative fixation length was on the link and picture exactly below the animation (Figure B.10(a)). The horizontal menu, the site's logo and the first box from each

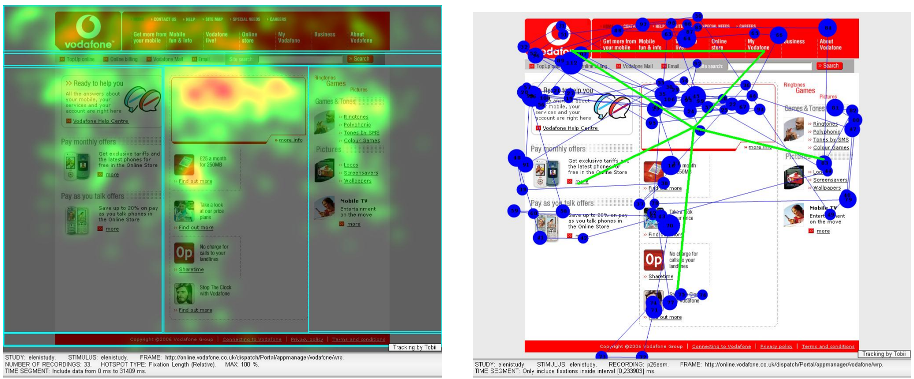


(a) Hotspot (b) Gaze plot

Figure B.9: Eye Tracking - PEVE Group

right and left column are the rest of the fixations while the subjects only glanced through the rest of the page, without fixating on any particular place.

The gazing session (see also Figure B.10(b)) started from the page's logo and went diagonally to the center animation while fixating a bit longer on the first box of the left column. Next the participants gazed back to the animation to fixate more and then back to glancing though the rest of the page without any common order. It is important to notice that the participants did not fixate on the smaller pictures of each box but on the links that are listed within each box/area.



(a) Hotspot (b) Gazeplot

Figure B.10: Eye Tracking - Vodafone Online

Appendix C

Webpages Used During the Pilot Evaluations

The screenshots of the ten Web pages used for the evaluations described in Chapter 4. The screenshots were taken on the 1st December 2006

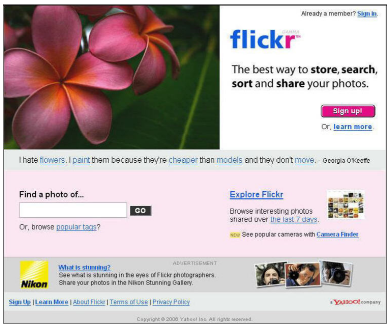
C.1.1: Amazon

C.1.2: Annotea

Figure C.1: Screenshots used for the Pilot Evaluations

C.1.3: BBC UK

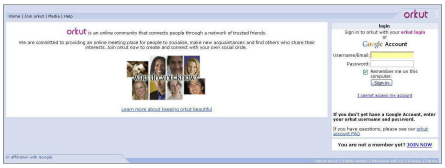
C.1.4: Firefox



C.1.5: Flickr

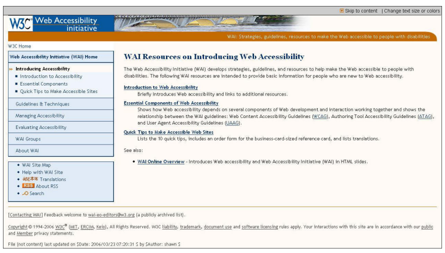
C.1.6: GoogleSearch

Figure C.1: Screenshots used for the Pilot Evaluations



C.1.7: Orkut

C.1.8: MSN



C.1.9: WAI

C.1.10: Yahoo

Figure C.1: Screenshots used for the Pilot Evaluations

Appendix D

Pilot Evaluation I

D.1 Materials

D.1.1 Observer's Script

The observer followed this script while introducing the participants to the experiment

Hello everyone,

First I would like to thank you for participating in this experiment!

What we are trying to do is to understand how people perceive images and their visuall complexity, that is their appearance.

During this experiment, you will you will be asked to do image comparison and choose which image is more visually complex.

Now, please read the Information Sheet and when you are ready sign the Consent form. Then have the Questionnaire ready to start the experiment.

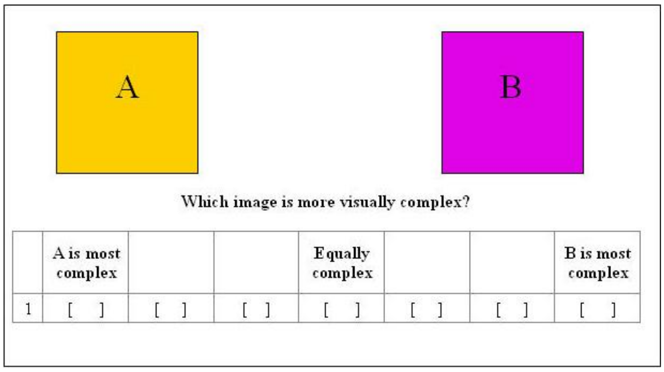
(when they are ready to start): Please note that every slide/comparison will appear for 20seconds. So within this time you have to see the images and mark on the sheet whatever you believe is the appropriate answer.

D.1.2 Information Sheet

The Information Sheet given to participants to read after the observer's introduction:

During this experiment you will be asked to do image comparison with respect to their visual complexity. You will look at a projector screen showing two images at a time and asked to answer which image is more visually complex on a scale by marking the appropriate box.

An example of what you will see is below:



There are no right or wrong answers; the answers reflect your opinions only. It is important for you to know that any information that you provide will be confidential. You do not need to identify yourself by name on any materials. All of the data will be summarized and no individual can be identified from these summarized results.

D.1.3 Consent Form

The Consent Form given to participants and the observer:

Participant:

The study's observer has explained to me what is asked of me as a volunteer, and I have read the information sheet.

I consent to take part as a volunteer and be recorded. I understand that I am free to withdraw at any time without giving any reason and without detriment to myself.

Signed: .... Date: ....

NAME: ....

Observer:

I confirm that I have fully explained the purpose and nature of the investigation and the risks involved

Signed: .... Date: ....

NAME: ....

D.1.4 Demographic Data Questionnaire

The Demographic Data Questionnaire that each participant had to complete:

The University of Manchester

ViCRAM - Visual Complexity Study

1. Please select the age range you belong:

16--25	26--35	36--45	46--65
[ ]	[ ]	[ ]	[ ]

2. Please select your gender:

Male	Female
[ ]	[ ]

3. Is English your native language?

Yes	No
[ ]	[ ]

4. How often do you use the Internet?

Every Day	A few times a week	Once a week to once a month	Never
[ ]	[ ]	[ ]	[ ]

5. Purpose for using the Internet? (mark all that applies)

Business /Work	Email /Chat	Special Interests	Online Purchases
[ ]	[ ]	[ ]	[ ]

6. Are you colour blind?

Yes	No
[ ]	[ ]

If you answered Yes, please specify what type:.......

7. Do you have any other problems with your vision?

Yes	No
[ ]	[ ]

If you answered Yes, please specify what problems:........

8. Can you see the screen ahead of you?

Yes	No
[ ]	[ ]

If you answered No, please let the observer know.

If you wear glasses please use them for this experiment.

Participant #

0	[ ]	[ ]
1	[ ]	[ ]
2	[ ]	[ ]
3	[ ]	[ ]
4	[ ]	[ ]
5	[ ]	[ ]
6	[ ]	[ ]
7	[ ]	[ ]
8	[ ]	[ ]
9	[ ]	[ ]

Group #

0	[ ]	[ ]
1	[ ]	[ ]
2	[ ]	[ ]
3	[ ]	[ ]
4	[ ]	[ ]
5	[ ]	[ ]
6	[ ]	[ ]
7	[ ]	[ ]
8	[ ]	[ ]
9	[ ]	[ ]

Figure D.1: Demographic Data Questionnaire

D.1.5 Answer Sheet Questionnaire

The Answer Sheet that participants used to record their answers:

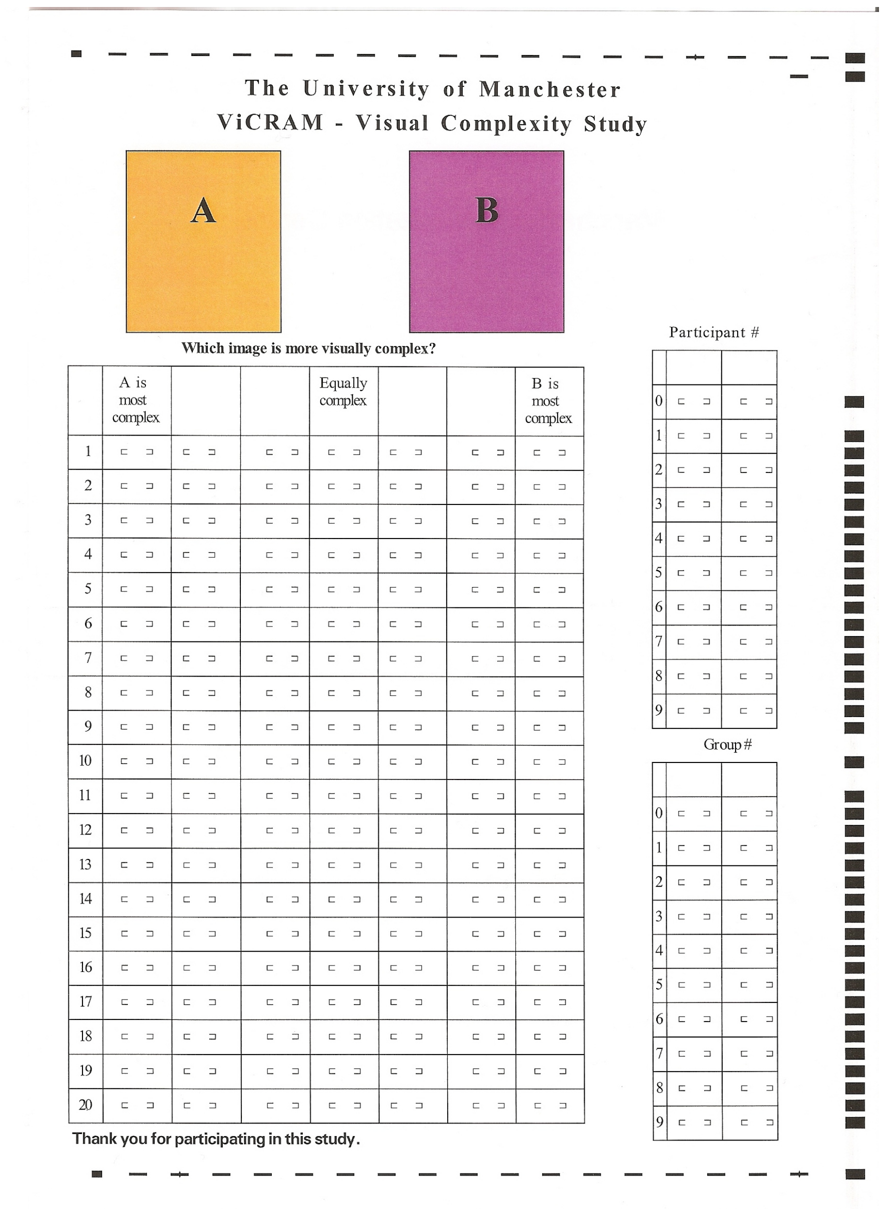


Figure D.2: Answer Sheet Questionnaire

D.2 Data

Tables D.1 and D.2 show the value given for each comparison by the subjects (participants). pXY is the paired comparison between Web page X versus Y. A negative value means that X is considered as more visually complex and a positive value that the Y Web page is ranked as more complex by the participants.

Tables D.3 and D.4 list the mean values calculated for each comparisons and the results of the t-test.

Table D.1: Participants' Image Comparison Data (Part A)

Subjects	p12	p13	p14	p15	p16	p17	p18	p19	p110	p23	p24	p25	p26	p27	p28	p29	p210	p34	p35	p36	p37	p38
1						-1				1	1		-1			0					-0.5	
2						0				1.5	-1		-1			-1					1	
3						-1				-2	-1		-1			-2					2	
4	-2	1				0				2	-2		0			0					0	
5	1	0		0	-1		-1.5	-1					0		1		2	-2	-2	-2		
6	0	-1		-2	-1		-1	-2					-2		0		2	-1	2	-3		
7	-2	1		-1	-1		-2.5	-1					-1		-2		-1	-2	-2	-1		
8	-2	1		-1	-2		-2.5	-2					-2		-2		2.5	-3	-2	-2		
9	2	-1		-1	-1		-1.5	0					2		2		-1.5	0	-3	-1		
10				-1	-2		-1.5	-2					-1		1		1.5	-2	-1	-1		
11	0	0		-3		0			0						-3	-3			-1.5			-2
12	-2	-1		-3		0			0.5						1	2			-3			-3
13	-1	-3		-3		-1			-1.5						1	-1			-1.5			-2
14	-1	1		-3		0			0.5						-1	0			-2			3
15	-2	-1		-3		1			-1						0	0			-2			-3
16	-1	0		-2		0			0.5						-1	-1			-2			-3
17			-2		2						1	2.5								-1		-0.5
18			-3		-3						3	0								3		-3
19			-3		0						3	1.5								0		-1.5
20			-2		-1						1	-0.5								0		1.5
21			-2		-3						3	-2.5								2		1.5
22			-2		-1						1	-2								2		-0.5
23			-2		-1						1	-1								1		-1.5
24			-1		-2						2	-1								1		0
25		2	-2	-3	-2	-1						-1		1.5					-1		1	
26		1	-1	-2	-1	0						-1		-1.5					0		-3	
27		2	-1	-2	-2	0						1		2.5					-1		-2	
28		2	-3	-3	-2	0						-1		0					-1		1	
29		2	-2	-2	-2	0						-1		-1					-2		1	
30		2	-1	-2		0						1		2.5				2			0	
31		1	-1.5	-1		1						1		2.5				2			0	
32		2	-2	-2		0						-3		-2.5				3			2	
33		3	-0.5	-2		1						-3		-1.5				2			0	
34		2	-1	-1		0						-3		-0.5				1			1	
35		0	-1	1		0						3		0.5				1			-1	
36		3	0	0		1						1		2.5				0			1	
37		0	-0.5	0		1						1		2				2			1	
38		1	-1	-1		-1						-1		0.5				1			-2	
39			-3			1				3	0		2			1					-1	
40			-3			3				3	3		3			3					3	
41			0			0				2	1		0			1					0	
42	-2		2			-2				3	3		3			2					-3	
43	-2		1	0	0		0	-1					1		2		3	-1		-1		
44	-2		3	1	-2		-1.5	-2					1		1		1.5	-2		-3		
45	-2		0	-2	-2		-1	-1					1		-2		2.5	-2		-1		
46	-2		1	0	-2		-0.5	-2					1		1		1.5	-2		-2		
47			0	-1	-1		-2.5	-1.5					1		-1		1.5	-2		-1		
48	0								0				0		1	0		0	-1			1
49	-2								-2				0		-1	0		-2	-2			-2
50	2								0				0		-3	0		-3	-2			-3
51	2								-1.5				0		-2	-2		-2	1			-2
52	3							-3	0				-3	-3	-3			-3	-3			0
MEAN	-0.682	0.769	-1.081	-1.452	-1.375	0.071	-1.455	-1.542	-0.409	1.688	1.188	-0.409	0.125	0.300	-0.455	-0.056	1.409	-0.600	-1.455	-0.526	0.068	-1.053

Table D.2: Participants' Image Comparison Data (Part B)

Subjects	p39	p310	p45	p46	p47	p48	p49	p410	p56	p57	p58	p59	p510	p67	p68	p69	p610	p78	p79	p710	p89	p810	p910
1			1	1	2	-1	-1		1		-1	1	1		0							2	
2			0	1	2	0	0		1		0	1	2		-2							2	
3			0	2.75	1	1	2		2		1	2	2		2							2	
4			2	2	3	2	2		1		-1	0	2		0							2	
5																	2			0	1	2	
6																	1			-1	-2	2	
7																	1.5			0	1	2	
8																	2.5			0	2	3	
9																	2.5			0	1	2	
10																	1			-1	1	2	
11		0							3	2			2			0				-1	1		0.5
12		0							-1	2			1			-3				0	1		2.5
13		1							2	2			3			-2				-1	-2		2.5
14		0							1	2			1			1				-1	0		1.5
15		0							0	3			2			0				0	1		2
16		-1							1	2			1			-2				-1	1		1.5
17	-1	0			2		0			1	-2	-2			1	1		0	-1.5				
18	3	0			-3		-3			3	-3	3			0	3		-3	3				
19	3	0			3		-3			-3	3	-3			0	0		-3	1.5				
20	0	-1			1		0			1	0	-1			0	0		1	0				
21	1	-1			-3		0			3	-3	3			2	2		-3	-1				
22	0	-1			-1		1			3	-2	2			-2	-1		-1	0				
23	-1	0			-1		1			2	-1	1			-1	1		-1	1.5				
24	1	0			-2		-1			3	-2	1			0	0		-2	0				
25	-1		0			0	0	2						-0.5						1	-1		
26	-2		2			-1	1	2						2.5						-1	2		
27	-1		0			-1	1	2.5						1						0	-1		
28	0		0			1	1	2.5						0						1	0		
29	-1		-1			1	1	2.5						0.5						-1	0		
30	-2		1			1	0	2						2	0			1					
31	-1		1			1	-1	2						1	-1			1					
32	0		-1			-2	2	3						3	-3			3					
33	1		0			-2	1	2.5						0	-2			2					
34	1		-1			1	1	2.5						2	0			2					
35	-1		3			-2	2	1						2	1			2					
36	1		1			1	0	2.5						3	0			2					
37	-1		1			-1	0	1						2	1			1					
38	0		1			1	1	1.5						1.5	0			1					
39			1	-1	1	-2			-1		2	0	2		1							3	2
40			3	-3	3	0			-3		-3	-3	-3		3							3	3
41			0	-1.5	1	-1			-1		-1	1	1		0							1	1
42			3	-2.5	2	-3			-2		-2	-1	2		-2							3	2
43			1													-1	1				-1	2	
44			-2													0	1				-1	1	
45			1													2	2.5				1	2	
46			2													-1	2				-1	1	
47			2													0	1.5				1	2	
48		0		0					1	1	1	0	1			1				0			0.5
49		-2		-1					-3	2	0	1	1			3				-1			1.5
50		0		3					3	3	0	2	3			0				-3			3
51		-2		1					-2	-2	-2	-2	0			0				-2			1.5
52		3		0					0	3	0	3	3			3				0			3
MEAN	-0.045	-0.211	0.778	0.135	0.688	-0.273	0.308	2.107	0.158	1.737	-0.762	0.429	1.421	1.429	-0.080	0.292	1.682	0.176	0.438	-0.545	0.227	2.053	1.867

Table D.3: One-Sample Descriptive Statistics

Comparisons	N	Mean	Std. Deviation	Std. Error Mean
P12	22	-0.682	1.673	0.357
P13	26	0.769	1.423	0.279
P14	31	-1.081	1.478	0.265
P15	31	-1.452	1.207	0.217
P16	24	-1.375	1.056	0.215
P17	28	0.071	0.940	0.178
P18	11	-1.455	0.820	0.247
P19	12	-1.542	0.782	0.226
P110	11	-0.409	0.917	0.276
P23	8	1.688	1.668	0.590
P24	16	1.188	1.601	0.400
P25	22	-0.409	1.736	0.370
P26	24	0.125	1.513	0.309
P27	15	0.300	1.916	0.495
P28	22	-0.455	1.654	0.353
P29	18	-0.056	1.514	0.357
P210	11	1.409	1.411	0.425
P34	25	-0.600	1.871	0.374
P35	22	-1.455	1.214	0.259
P36	19	-0.526	1.679	0.385
P37	22	0.068	1.576	0.336
P38	19	-1.053	1.817	0.417
P39	22	-0.045	1.362	0.290
P310	19	-0.211	1.084	0.249
P45	27	0.778	1.281	0.247
P46	13	0.135	1.889	0.524
P47	16	0.688	2.056	0.514
P48	22	-0.273	1.386	0.296
P49	26	0.308	1.320	0.259
P410	14	2.107	0.594	0.159
P56	19	0.158	1.834	0.421
P57	19	1.737	1.661	0.381
P58	21	-0.762	1.640	0.358
P59	21	0.429	1.859	0.406
P510	19	1.421	1.346	0.309
P67	14	1.429	1.124	0.300
P68	25	-0.080	1.412	0.282
P69	24	0.292	1.574	0.321
P610	11	1.682	0.643	0.194
P78	17	0.176	1.976	0.479
P79	8	0.438	1.474	0.521
P710	22	-0.545	0.912	0.194
P89	22	0.227	1.193	0.254
P810	19	2.053	0.621	0.143
P910	15	1.867	0.834	0.215

Table D.4: One-Sample t-test Results, Test Value = 0, Significant at ≤ 0.05

Comparisons	t	df	Sig. (2-tailed)	Mean Difference
P12	-1.912	21.000	0.070	-0.682
P13	2.757	25.000	0.011	0.769
P14	-4.070	30.000	0.000	-1.081
P15	-6.698	30.000	0.000	-1.452
P16	-6.382	23.000	0.000	-1.375
P17	0.402	27.000	0.691	0.071
P18	-5.882	10.000	0.000	-1.455
P19	-6.828	11.000	0.000	-1.542
P110	-1.480	10.000	0.170	-0.409
P23	2.862	7.000	0.024	1.688
P24	2.967	15.000	0.010	1.188
P25	-1.105	21.000	0.282	-0.409
P26	0.405	23.000	0.689	0.125
P27	0.606	14.000	0.554	0.300
P28	-1.289	21.000	0.211	-0.455
P29	-0.156	17.000	0.878	-0.056
P210	3.312	10.000	0.008	1.409
P34	-1.604	24.000	0.122	-0.600
P35	-5.619	21.000	0.000	-1.455
P36	-1.366	18.000	0.189	-0.526
P37	0.203	21.000	0.841	0.068
P38	-2.525	18.000	0.021	-1.053
P39	-0.157	21.000	0.877	-0.045
P310	-0.846	18.000	0.408	-0.211
P45	3.155	26.000	0.004	0.778
P46	0.257	12.000	0.802	0.135
P47	1.337	15.000	0.201	0.688
P48	-0.923	21.000	0.367	-0.273
P49	1.189	25.000	0.246	0.308
P410	13.270	13.000	0.000	2.107
P56	0.375	18.000	0.712	0.158
P57	4.557	18.000	0.000	1.737
P58	-2.129	20.000	0.046	-0.762
P59	1.056	20.000	0.303	0.429
P510	4.600	18.000	0.000	1.421
P67	4.755	13.000	0.000	1.429
P68	-0.283	24.000	0.779	-0.080
P69	0.908	23.000	0.373	0.292
P610	8.673	10.000	0.000	1.682
P78	0.368	16.000	0.718	0.176
P79	0.839	7.000	0.429	0.438
P710	-2.806	21.000	0.011	-0.545
P89	0.894	21.000	0.381	0.227
P810	14.402	18.000	0.000	2.053
P910	8.671	14.000	0.000	1.867

Appendix E

Pilot Evaluation II

E.1 Materials used during the study

E.1.1 Observer's Script

The observer followed this script while introducing the participants to the experiment

Hello everyone,

First I would like to thank you for participating in this experiment!

What we are trying to do is to understand how people perceive images and their visuall complexity, that is their appearance.

During this experiment, you will you will be asked to do image comparison and choose which image is more visually complex. You will have 10 seconds for each comparison and within this time you will have to decide if they look different and which one is more complex.

Then, you will look on ten images and asked to give a ranking score from 1 to 10. You will be able to see each image twice so if you change your mind for a ranking you already had you can put the second score next to the existing one.

Now, please read the Information Sheet and when you are ready sign the

Consent form and complete the Demographic Data Questionnaire. Then have the Questionnaire 1 ready to start the experiment.

(when they are ready to start): Please note that every slide/comparison will appear for 10 seconds. So within this time you have to see the images and mark on the sheet whatever you believe is the appropriate answer.

E.1.2 Information Sheet

See Section D.1.2 on page 215.

E.1.3 Consent Form

The Consent Form given to participants and the observer: See Section D.1.3 on page 216.

E.1.4 Demographic Data Questionnaire

The Demographic Data Questionnaire that each participant had to complete: See Section D.1.4 on page 217.

E.1.5 Answer Sheet Questionnaire - Part I

The Answer Sheet that participants used to record their answers: See Section D.1.5 on page 217.

E.1.6 Answer Sheet Questionnaire - Part II

The Answer Sheet that participants used to record their answers for the second part of this study (Pilot IIb):

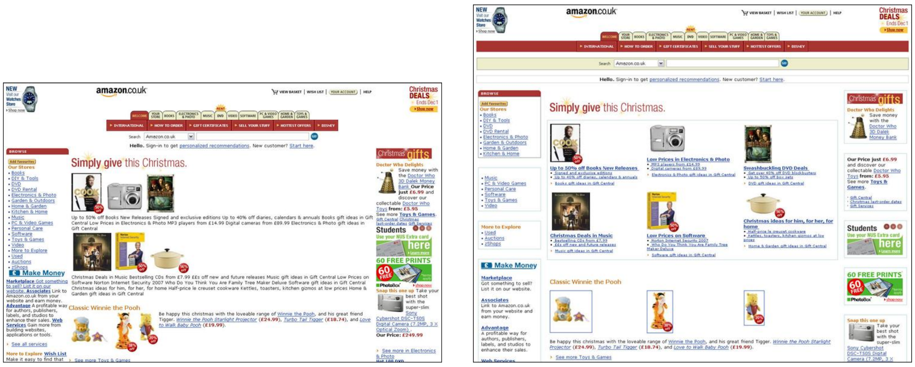
Participant #……… Croup #………..

Page Number	Visual Complexity Ranking (1=simplest---10=most complex)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Figure E.1: Answer Sheet Questionnaire - Part II

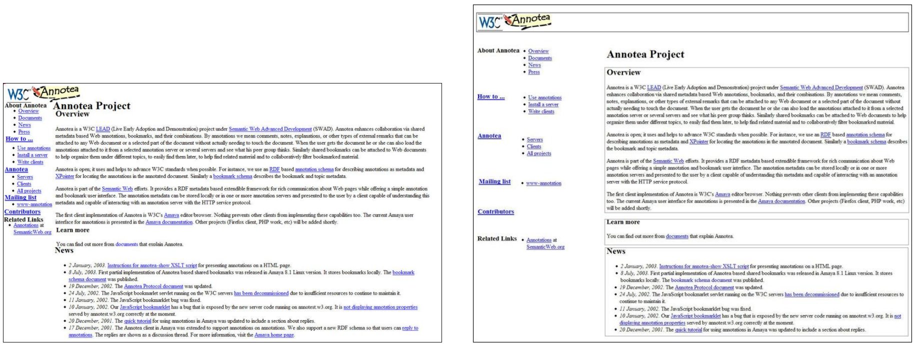
E.2 Webpages used in this experiment

The screenshots of the original versions of the ten Web pages were taken in February 2007 (images are shown in Appendix C on page 211). The source code was saved the same time and manipulated to create the modified versions shown in pairs below.



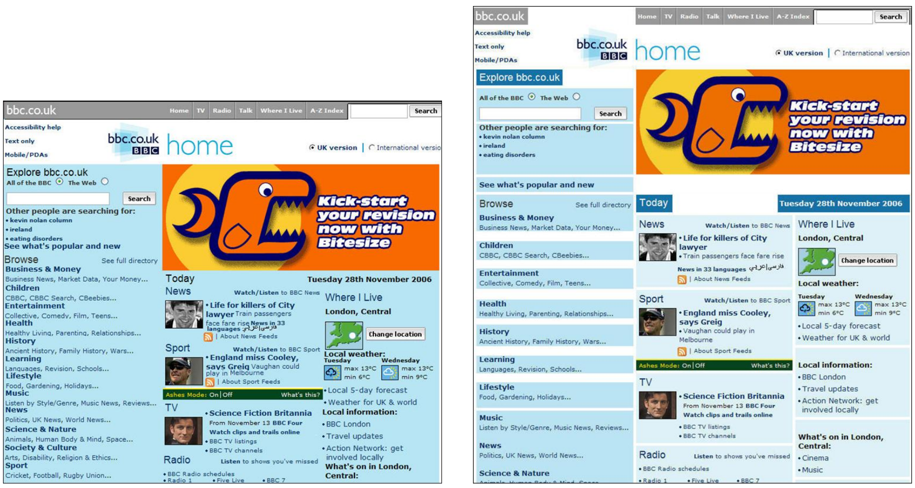
E.2.1: Less Edges E.2.2: More Edges

Figure E.2: Pilot IIb - Amazon Modified Page



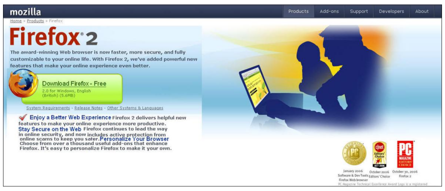
E.3.1: Less Edges E.3.2: More Edges

Figure E.3: Pilot IIb - Annotea Modified Page



E.4.1: Less Edges E.4.2: More Edges

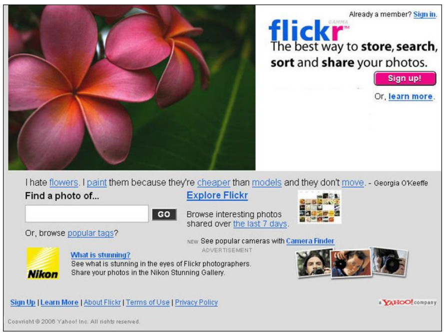
Figure E.4: Pilot IIb - BBC-UK Modified Page



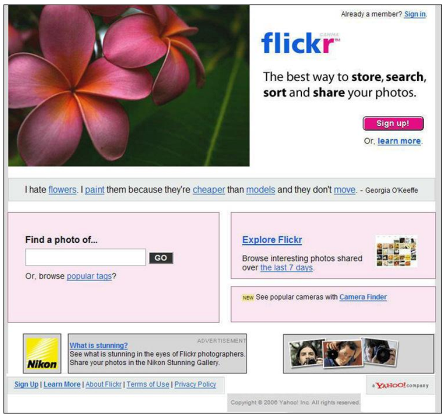
E.5.1: Less Edges

E.5.2: More Edges

Figure E.5: Pilot IIb - Firefox Modified Page

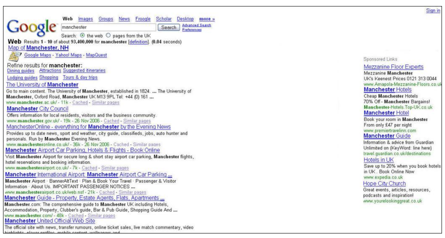


E.6.1: Less Edges

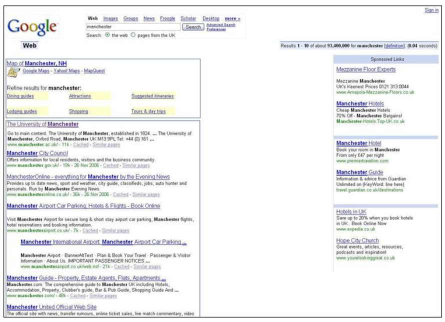


E.6.2: More Edges

Figure E.6: Pilot IIb - Flickr Modified Page



E.7.1: Less Edges



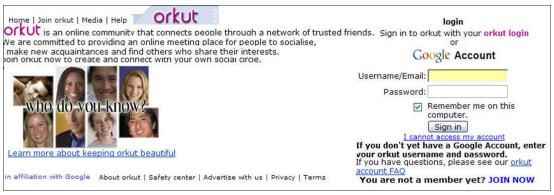
E.7.2: More Edges

Figure E.7: Pilot IIb - GoogleResults Modified Page

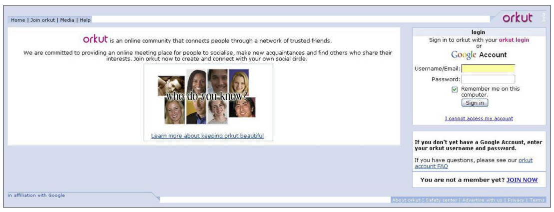
E.8.1: Less Edges

E.8.2: More Edges

Figure E.8: Pilot IIb - MSN Modified Page

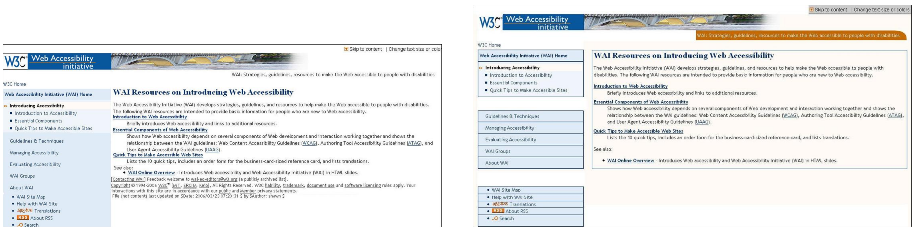


E.9.1: Less Edges



E.9.2: More Edges

Figure E.9: Pilot IIb - Orkut Modified Page



E.10.1: Less Edges E.10.2: More Edges

Figure E.10: Pilot IIb - WAI Modified Page

E.11.1: Less Edges E.11.2: More Edges

Figure E.11: Pilot IIb - Yahoo Modified Page

E.3 Data

Table E.1 shows the value given for each comparison by the subjects (participants). pXY is the paired comparison between Web page X versus its version Y, where Y can be a (less edges) or b (more edges). A negative value means that X is considered as more visually complex and a positive value that the Y Web page is ranked as more complex by the participants.

Table E.1: Pilot IIa: Participants' Image Comparison Data (Part A)

P	p1a	p1b	p2a	p2b	p3a	p3b	p4a	p4b	p5a	p5b	p6a	p6b	p7a	p7b	p8a	p8b	p9a	p9b	p10a	p10b
1	1	0	1	1	-1	1	1	-1	2	0	1	-1	0	1	-1	1	-1	-1	0	-1
2	1	0	1	0	2	1	2	-2	1	0	1	-2	1	-1	1	1	1	1	0	-1
3	-3	3	0	0	-3	3	0	0	3	-3	3	3	-3	-3	3	0	-3	-3	3	0
4	-1	0	2	-1	0	0	-2	-1	-2	0	1	1	1	0	-1	-2	0	0	0	-1
5	1	0	-3	-1	1	-1	3	-1	-1	-2	2	-3	2	0	-3	-1	2	0	0	-1
6	-3	-3	-1	2	-1	-1	1	-1	-2	3	0	0	0	0	0	-3	3	3	0	-1
7	1	0	1	1	1	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0
8	3	0	2	2	2	1	3	1	0	1	3	-1	-2	1	0	0	2	-1	0	1
9	0	0	-2	-1	-1	-1	-1	-1	0	0	-1	0	0	0	-1	-1	-2	-1	0	0
10	2	-1	2	1	2	0	1	1	1	-1	1	1	2	1	3	1	2	0	2	-1
11	2	0	2	1	1	-1	2	1	1	1	2	1	1	-1	0	-1	1	0	0	-1
12	-2	0	1	-2	1	-1	1	-1	1	-1	2	1	1	1	0	0	-1	0	1	0
13	1	0	-1	1	1	-1	-1	-1	-1	0	1	1	1	-1	1	0	-1	-1	0	0
14	0	1	1	1	2	-1	2	-1	-1	-1	0	0	-1	1	2	1	1	1	1	1
15	0	1	1	-1	0	1	0	-1	-1	-1	0	-1	2	0	-1	1	0	1	1	1
16	1	1	-1	-2	2	1	-2	-1	-2	-1	-1	0	1	1	-2	-1	-2	-1	1	-1
17	0	0	2	1	1	-2	2	3	1	-2	2	1	2	0	-3	-1	1	1	1	0
18	1	0	-1	-2	2	1	-2	-3	-3	-2	2	1	2	-1	3	-3	3	1	2	0
19	2	0	3	-2	3	-1	3	0	-3	-3	3	2	3	2	3	0	3	1	2	1
20	2	1	2	1	-2	-1	-3	-2	-2	1	2	2	2	0	-2	-2	-2	-2	-1	-2
21	3	-1	3	1	1	2	2	3	2	0	2	1	2	3	0	1	-1	2	3	2
22	2	0	2	1	-2	2	2	2	1	2	2	2	3	2	2	0	1	1	2	0
23	1	0	1	1	1	1	-1	1	-1	1	1	-1	2	0	1	-1	0	2	2	-2
24	2	2	-1	-1	2	3	3	-2	2	1	-2	2	1	1	3	-1	2	1	3	-2
25	3	0	-2	2	3	0	3	-2	2	-2	-2	3	2	0	2	-2	1	2	3	0
26	0	0	-3	-3	3	3	-3	-3	3	3	-3	-3	-3	-3	-1	3	3	3	-3	0
27	0	1	1	1	-2	0	-1	3	-2	1	2	3	1	0	3	0	-1	2	-1	-2
28	-2	-2	-2	-2	0	2	3	0	0	-1	0	2	2	0	-2	1	0	0	0	0
29	0	0	0	-2	2	-2	-1	-1	0	1	1	1	-1	0	-2	0	1	-1	-1	-2
30	3	3	2	-3	3	3	3	3	2	3	3	3	3	0	1	0	-1	2	0	-3
31	-3	0	-2	-3	0	-1	3	0	-1	0	0	1	0	0	-2	1	3	-1	3	1
32	2	1	2	-1	2	0	2	1	1	0	2	1	2	-1	1	0	2	0	2	-1
33	0	0	2	2	1	1	1	1	0	-2	1	1	3	-1	-1	-1	2	1	0	0
34	3	0	1	0	2	3	3	1	3	0	2	3	2	0	2	0	0	0	2	1
35	1	-1	2	-2	1	0	2	-1	2	2	2	1	-1	1	3	1	2	-2	2	0
36	3	-1	2	-2	2	1	-1	-1	1	-2	-1	-1	1	1	1	0	1	-1	2	0
37	1	1	1	1	0	-1	1	1	-1	0	0	1	1	1	1	-1	1	-1	1	1
38	2	0	2	3	3	3	3	3	-2	1	3	1	3	2	1	0	0	-2	1	2
39	3	-1	2	2	3	3	2	2	-1	-2	2	2	3	-2	2	2	1	3	3	3
40	1	-1	3	3	3	3	2	3	3	-2	3	3	3	3	3	2	3	3	1	0
41	3	1	0	2	3	2	3	3	2	2	3	3	2	2	2	2	0	0	1	1

Table Continues on the next page

Participants' Image Comparison Data (Part B)

P	p1a	p1b	p2a	p2b	p3a	p3b	p4a	p4b	p5a	p5b	p6a	p6b	p7a	p7b	p8a	p8b	p9a	p9b	p10a	p10b
42	-1	-2	2	1	1	1	1	1	-2	2	1	-2	-1	-1	2	1	-2	-2	-1	-2
43	1	0	-1	-1	-1	-1	2	-1	1	-1	0	-2	1	0	-2	-2	0	-1	0	0
44	1	0	0	0	1	1	2	1	1	-1	2	1	2	0	-2	1	-1	-2	0	-1
45	2	0	-3	0	-1	-2	2	2	0	-1	-2	-2	2	-1	-3	-3	-3	-2	-2	0
46	2	0	0	0	0	0	1	1	2	0	-1	0	2	0	-1	0	-1	0	1	0
47	-3	-1	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	3	-3	-3	-3	-3	-3	-3	0
48	-2	1	0	-1	-1	-1	2	-1	0	0	-2	0	-1	0	2	0	1	1	-1	1
49	2	0	-2	-3	2	3	1	-3	-2	-1	-1	2	3	-2	-2	-1	0	2	-1	0
50	1	-1	2	-1	1	1	1	1	-1	1	2	2	2	-2	-1	-1	3	1	-2	1
51	3	-2	3	-3	3	0	2	0	3	2	-3	3	3	-3	-3	2	1	-3	3	-1
52	1	0	1	2	2	-1	-2	-2	0	-1	1	1	2	0	-3	-1	1	-1	-3	0
53	-1	1	1	2	-1	-1	1	1	0	1	-1	2	0	0	-2	0	-1	0	-1	0
54	2	0	3	2	3	1	2	2	3	-2	1	2	2	1	2	1	2	-1	1	0
55	1	0	0	-2	2	-3	3	-2	2	-3	2	2	1	0	2	2	2	0	2	1
56	3	0	3	-3	3	2	3	2	-2	-3	3	2	3	-3	0	2	3	-3	0	0
57	-2	1	1	1	-2	1	-1	2	-1	0	-2	0	-1	-1	-2	0	0	0	-2	-1
58	0	0	0	2	0	-2	-3	-2	1	0	-2	-1	-3	-2	-2	0	0	0	0	0
59	1	-1	2	-1	-1	1	2	2	1	2	2	2	3	1	2	0	2	-1	2	0
60	1	1	2	0	1	3	-2	1	1	0	2	2	2	0	1	1	0	-1	1	1
61	2	0	-3	-2	0	-2	-2	-2	-2	-2	-2	2	2	0	-2	-2	1	0	0	0
62	0	1	-1	0	2	1	-2	1	-2	-2	2	1	2	2	2	-2	-1	-1	1	2
63	2	0	0	1	3	-1	2	2	-1	-1	3	1	2	1	2	-2	1	-1	1	0
64	1	-1	0	0	2	-1	-2	-2	-2	0	0	-1	2	0	-1	-1	-1	0	-1	0
65	-3	-3	3	3	3	0	0	0	3	-3	0	0	3	3	-3	-3	-3	0	0	0
66	2	0	1	-1	1	-1	2	-2	1	-1	1	1	2	0	-3	0	-2	0	-1	0
67	2	-2	2	-2	-2	-2	2	-2	-2	-2	2	0	2	-2	-2	-2	-2	-2	-2	-2
68	1	-1	-2	-1	-2	0	-1	-1	1	-1	-1	0	1	0	-1	-1	-1	-1	-1	0
69	-1	0	-3	-3	-2	0	-1	-2	0	-1	0	-2	2	-1	-2	-2	-1	-2	-1	-2
70	-3	3	3	3	3	3	-3	3	3	-3	0	3	-3	3	3	3	3	3	3	3
71	1	1	-1	-1	2	1	2	1	1	-2	1	-2	2	-2	-2	-3	1	1	1	1
72	2	1	3	3	-1	0	2	2	2	-1	2	2	1	0	2	1	2	1	0	0
73	1	1	0	-2	-1	1	3	-2	-1	1	-1	2	3	-1	-2	1	0	-1	2	1
74	3	0	0	0	2	1	2	1	2	2	2	1	3	-2	1	-3	1	2	0	0
75	3	0	0	2	2	0	2	3	3	-3	-2	3	2	0	2	3	1	3	2	3
76	1	0	0	-1	1	1	-1	-1	1	1	1	1	2	1	1	0	1	1	1	0
77	3	2	2	-1	2	2	2	2	2	-1	1	2	2	2	1	1	1	-1	-3	-1
78	-1	0	0	1	-1	1	0	2	-1	0	0	0	-1	1	1	0	1	1	0	1
79	3	3	0	3	-2	0	-2	2	-3	2	-3	-2	-3	0	-3	2	0	2	-3	0
80	2	0	-1	-2	2	1	2	-1	0	0	2	-1	3	0	-2	-1	-1	0	1	0
81	0	1	-1	1	-1	0	2	2	2	0	1	1	0	0	3	1	2	1	2	0
Mean	0.864	0.099	0.506	-0.086	0.815	0.383	0.802	0.173	0.235	-0.370	0.667	0.728	1.235	-0.012	0.037	-0.173	0.432	0.037	0.444	-0.025

Table E.2: Pilot IIa: One-Sample Descriptive Statistics

Comparisons	N	Mean	Std. Deviation	Std. Error Mean
P1A	81	0.864	1.716	0.191
P1B	81	0.099	1.147	0.127
P2A	81	0.506	1.740	0.193
P2B	81	-0.086	1.783	0.198
P3A	81	0.815	1.711	0.190
P3B	81	0.383	1.554	0.173
P4A	81	0.802	1.887	0.210
P4B	81	0.173	1.780	0.198
P5A	81	0.235	1.756	0.195
P5B	81	-0.370	1.585	0.176
P6A	81	0.667	1.696	0.188
P6B	81	0.728	1.605	0.178
P7A	81	1.235	1.638	0.182
P7B	81	-0.012	1.410	0.157
P8A	81	0.037	2.009	0.223
P8B	81	-0.173	1.515	0.168
P9A	81	0.432	1.620	0.180
P9B	81	0.037	1.537	0.171
P10A	81	0.444	1.589	0.177
P10B	81	-0.025	1.162	0.129

Table E.3: Pilot IIa: One-Sample t-test Results, Test Value = 0, Significant at $\leq .05$

Comparisons	t	df	Sig. (2-tailed)	Mean Difference
P1A	4.533	80	.000	0.864
P1B	0.775	80	.441	0.099
P2A	2.618	80	.011	0.506
P2B	-0.436	80	.664	-0.086
P3A	4.286	80	.000	0.815
P3B	2.217	80	.029	0.383
P4A	3.828	80	.000	0.802
P4B	0.874	8	.385	0.173
P5A	1.203	80	.233	0.235
P5B	-2.104	80	.039	-0.370
P6A	3.539	80	.001	0.667
P6B	4.085	80	.000	0.728
P7A	6.785	80	.000	1.235
P7B	-0.079	80	.937	-0.012
P8A	0.166	80	.869	0.037
P8B	-1.027	80	.308	-0.173
P9A	2.401	80	.019	0.432
P9B	0.217	80	.829	0.037
P10A	2.517	80	.014	0.444
P10B	-0.191	80	.849	-0.025

Table E.4: Pilot IIb: Participants' Image Ranking Data

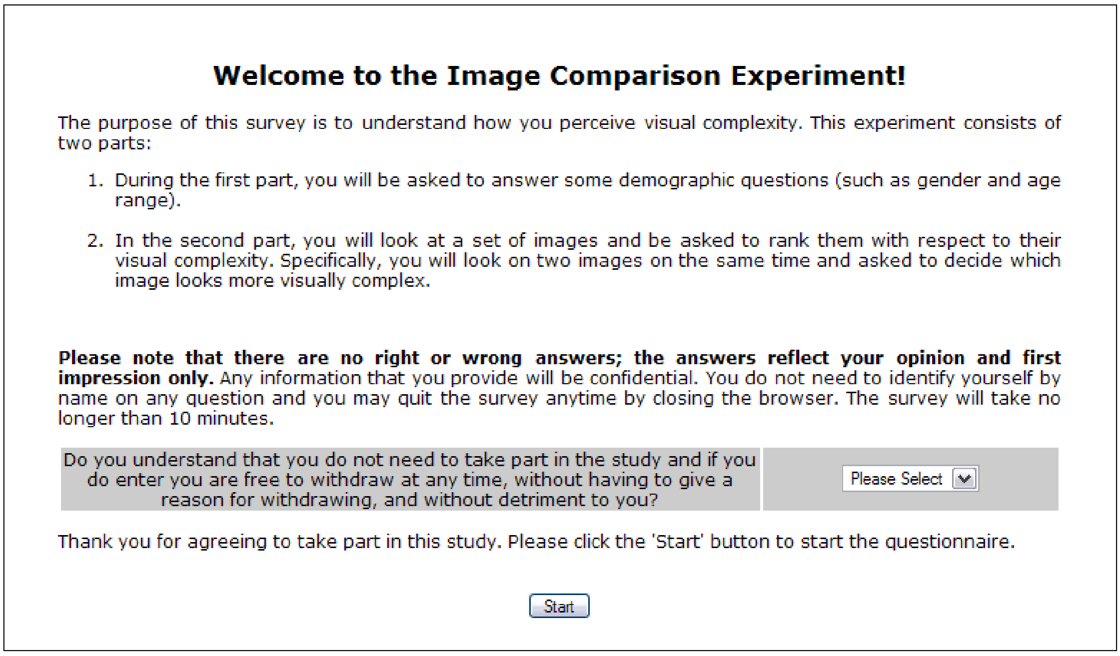
Participants	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
1	6	7	4	3	4	7	8	5	5	9
2	6	8	8	5	4	8	9	6	5	9
3	8	5	9	5	6	5	9	5	5	9
4	4	8	2	1	1	5	5	3	3	4
5	3	8	1	4	1	5	6	4	5	2
6	2	3	3	2	1	3	5	3	4	5
7	4	3	2	2	2	3	3	2	2	2
8	7	5	9	3	5	4	9	6	5	7
9	10	8	7	1	2	4	9	2	6	7
10	6	7	6	4	5	3	5	5	6	4
11	9	3	6	2	2	3	7	2	4	5
12	6	7	5	2	4	3	5	3	6	5
13	4	4	5	5	7	6	5	7	4	5
14	7	2	6	3	5	3	6	4	4	6
15	4	7	3	2	2	4	3	1	5	2
16	6	9	5	4	2	7	8	3	7	7
17	6	3	4	2	3	3	6	2	4	4
18	5	9	2	2	3	6	8	1	4	1
19	8	2	8	6	6	8	5	4	5	7
20	9	2	7	7	3	7	5	4	6	8
21	9	3	3	1	2	8	3	6	5	7
22	7	7	6	2	9	8	6	3	3	7
23	10	2	9	7	6	9	3	4	4	8
24	7	3	7	3	5	7	4	4	2	7
25	7	7	5	2	3	4	4	6	5	8
26	7	2	9	8	5	10	7	5	7	9
27	6	3	7	7	6	9	4	6	6	7
28	5	3	3	1	4	6	1	5	2	3
29	6	2	6	5	4	5	4	4	6	8
30	6	2	8	2	6	5	7	7	5	9
31	7	10	3	2	1	5	9	6	8	4
32	7	10	8	3	2	5	9	5	6	9
33	6	9	5	2	4	7	4	5	5	7
34	7	9	8	4	5	3	6	4	4	8
35	7	10	6	2	8	4	9	3	5	1
36	10	7	6	4	3	4	9	2	5	7
37	1	3	1	1	1	1	1	1	1	1
38	10	5	2	1	9	4	6	8	4	3
39	8	4	6	3	7	5	7	3	5	8
40	6	7	4	3	8	5	7	7	7	6
41	4	6	4	5	4	3	7	6	7	8
42	9	8	7	5	5	5	8	4	5	8
43	7	7	3	3	2	6	6	3	4	7
44	5	8	4	2	2	6	5	3	7	6
45	7	10	5	3	2	9	7	5	9	8
46	7	3	6	1	5	1	10	5	3	10
47	9	5	5	3	1	2	9	1	1	4
48	9	1	6	2	4	8	7	3	5	10
49	8	4	3	1	2	8	7	2	3	7
50	8	1	5	3	7	3	5	2	3	4
51	8	7	5	2	1	2	3	1	5	6
52	8	7	7	5	4	7	8	5	7	6
53	8	1	7	3	6	2	5	3	4	7
54	8	4	3	1	3	1	3	3	2	5
55	8	5	6	2	4	5	7	4	3	7
56	5	3	2	7	6	2	2	2	1	4
57	8	6	5	9	5	6	7	8	4	7
58	8	3	3	7	6	5	5	8	1	8
59	3	1	7	4	8	5	2	6	9	10
60	8	1	6	5	4	9	3	2	7	10
61	8	7	3	6	4	2	2	7	10	7
62	9	5	3	7	6	1	4	2	10	8
63	6	2	1	4	2	3	1	5	7	8
64	5	9	4	5	4	9	4	4	8	6
65	9	10	8	4	3	7	9	4	7	8
66	9	10	5	3	2	7	6	2	5	6
67	9	9	2	5	2	8	9	6	4	5
68	8	9	3	6	4	6	7	7	6	7
69	10	9	7	5	3	8	9	4	6	6
70	8	7	6	5	8	6	6	4	5	8
71	8	8	5	3	4	6	7	5	5	6
72	7	7	6	3	3	6	9	2	3	5
73	9	8	3	2	2	6	5	2	4	4
74	8	9	6	5	4	5	8	7	2	6
75	4	3	8	4	5	5	7	6	6	7
76	5	4	5	4	3	3	4	4	4	3
77	7	3	5	2	3	3	8	4	3	4
78	5	1	10	2	5	1	10	2	1	4
79	8	4	7	7	6	4	7	6	6	8
80	10	9	5	2	2	8	7	4	3	6
81	2	2	8	7	7	3	7	6	3	8

Appendix F

Web Page Chunking Evaluation

F.1 Online Evaluation Procedure

The Figures below show the stages in which the online experiment consisted.



Welcome to the Image Comparison Experiment!

The purpose of this survey is to understand how you perceive visual complexity. This experiment consists of two parts:

1. During the first part, you will be asked to answer some demographic questions (such as gender and age range).
2. In the second part, you will look at a set of images and be asked to rank them with respect to their visual complexity. Specifically, you will look on two images on the same time and asked to decide which image looks more visually complex.

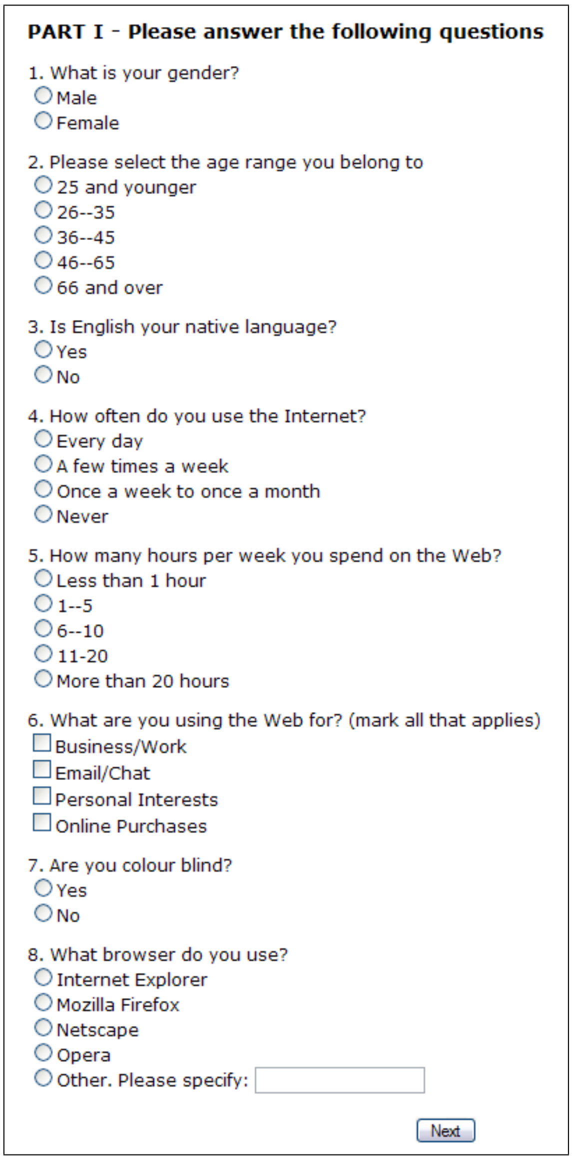
Please note that there are no right or wrong answers; the answers reflect your opinion and first impression only. Any information that you provide will be confidential. You do not need to identify yourself by name on any question and you may quit the survey anytime by closing the browser. The survey will take no longer than 10 minutes.

Do you understand that you do not need to take part in the study and if you do enter you are free to withdraw at any time, without having to give a reason for withdrawing, and without detriment to you?	Please Select

Thank you for agreeing to take part in this study. Please click the 'Start' button to start the questionnaire.

Start

Figure F.1: Introduction to the Online Evaluation - the participant could only proceed if agreed to take part by selecting the 'Yes' option



PART I - Please answer the following questions

1. What is your gender?
- Male
- Female

2. Please select the age range you belong to
- 25 and younger
- 26--35
- 36--45
- 46--65
- 66 and over

3. Is English your native language?
- Yes
- No

4. How often do you use the Internet?
- Every day
- A few times a week
- Once a week to once a month
- Never

5. How many hours per week you spend on the Web?
- Less than 1 hour
- 1--5
- 6--10
- 11-20
- More than 20 hours

6. What are you using the Web for? (mark all that applies)
- Business/Work
- Email/Chat
- Personal Interests
- Online Purchases

7. Are you colour blind?
- Yes
- No

8. What browser do you use?
- Internet Explorer
- Mozilla Firefox
- Netscape
- Opera
- Other. Please specify:

Next

Figure F.2: Demographic Questionnaire

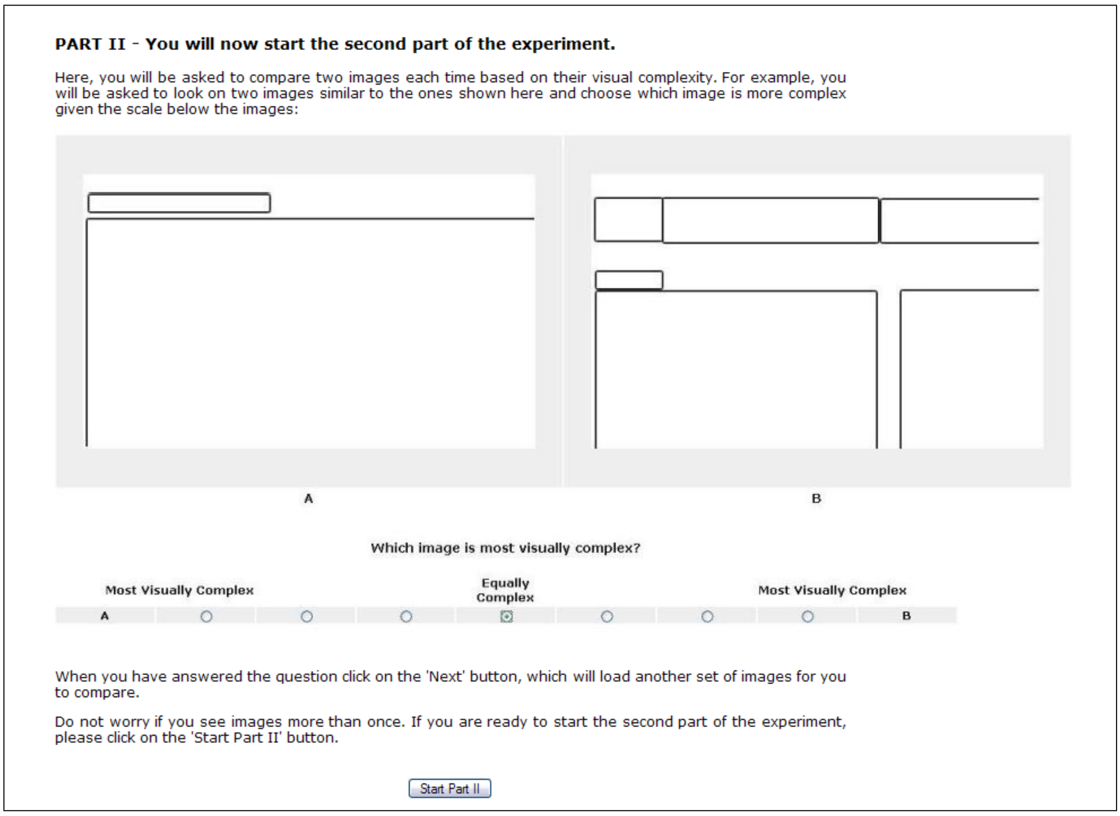


Figure F.3: Introduction to the Second Part of the Evaluation

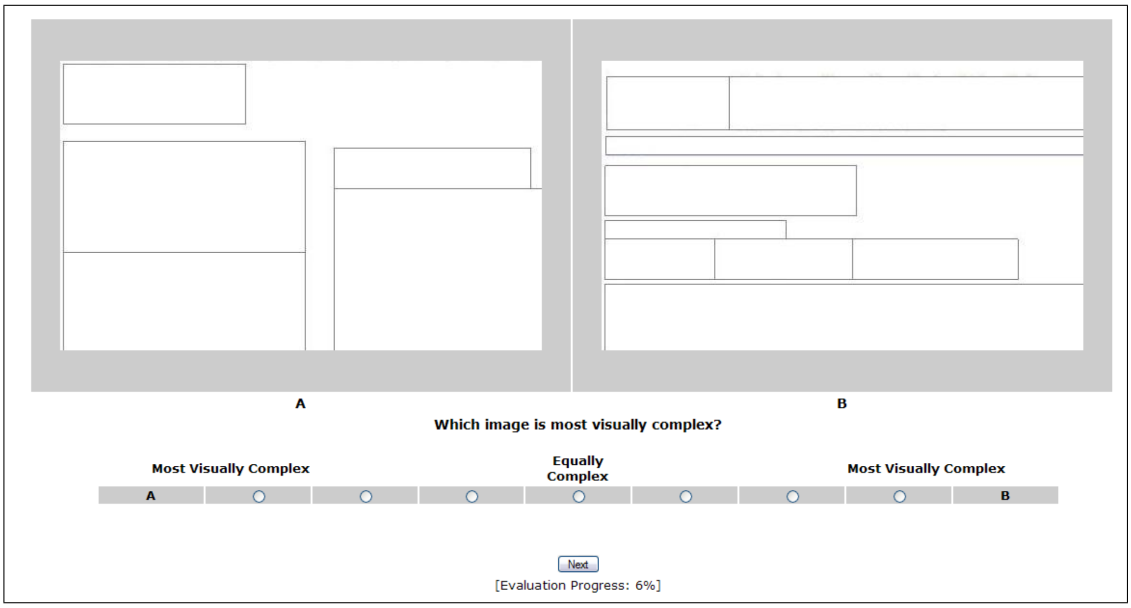
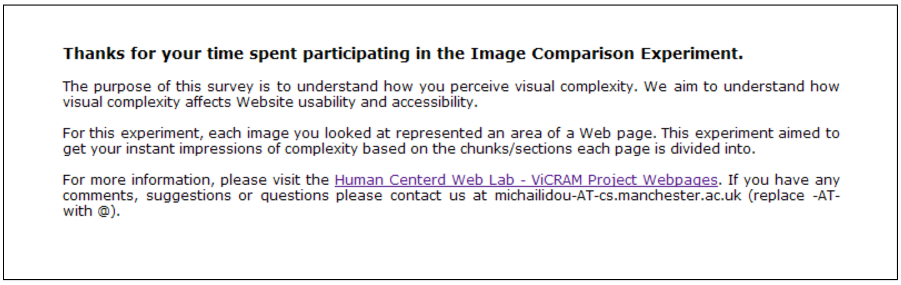


Figure F.4: A sample of a pair of images and the comparison question


Thanks for your time spent participating in the Image Comparison Experiment.

The purpose of this survey is to understand how you perceive visual complexity. We aim to understand how visual complexity affects Website usability and accessibility.

For this experiment, each image you looked at represented an area of a Web page. This experiment aimed to get your instant impressions of complexity based on the chunks/sections each page is divided into.

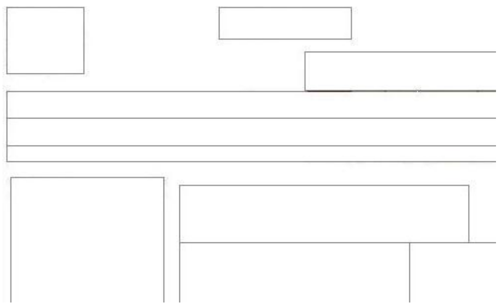
For more information, please visit the Human Centerd Web Lab - ViCRAM Project Webpages. If you have any comments, suggestions or questions please contact us at michailidou-AT-cs.manchester.ac.uk (replace -AT- with @).

Figure F.5: A 'Thank You' page which determined the end of the evaluation

F.2 Webpages and Experiment Stimuli

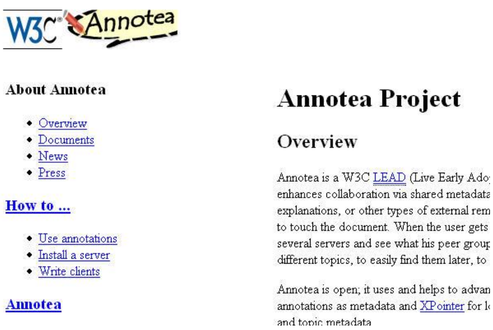
The screenshots shown below of the ten Web pages used as in their original look when the screenshot was taken and then as the modification took place to be used a the stimuli for the evaluation. The screenshots can also be found in the associated experimental material folder linked with this report.

F.6.1: Original

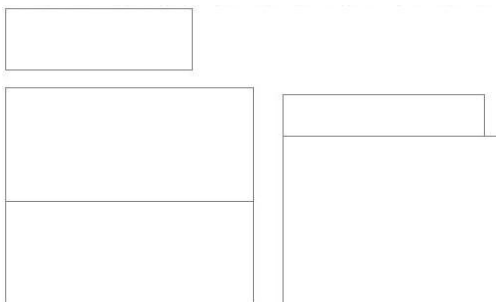


F.6.2: Chunk Rendering

Figure F.6: Chunk Ev.: ID1-Amazon Original and Chunk Rendering



F.7.1: Original



F.7.2: Chunk Rendering

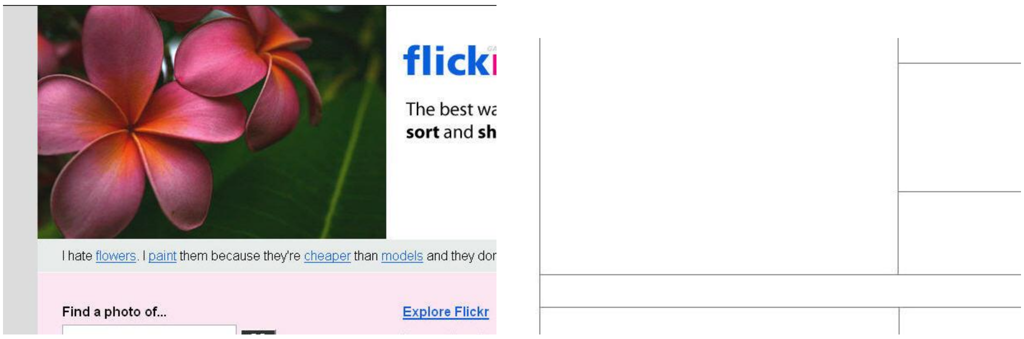
Figure F.7: Chunk Ev.: ID2-Annotea Original and Chunk Rendering

F.8.1: *Original* F.8.2: *Chunk Rendering*

Figure F.8: Chunk Ev.: ID3-BBC UK Original and Chunk Rendering

F.9.1: *Original* F.9.2: *Chunk Rendering*

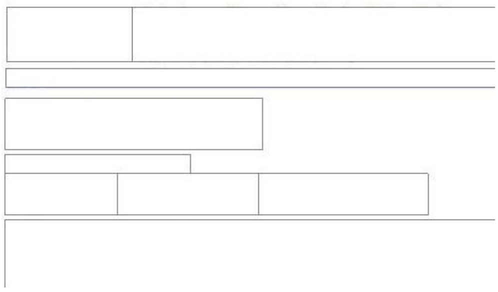
Figure F.9: Chunk Ev.: ID4-Firefox Original and Chunk Rendering



F.10.1: *Original* F.10.2: *Chunk Rendering*

Figure F.10: Chunk Ev.: ID5-Flickr Original and Chunk Rendering

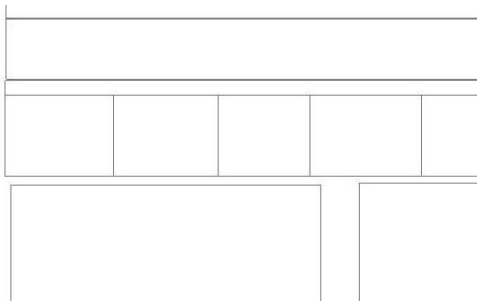
F.11.1: Original



F.11.2: Chunk Rendering

Figure F.11: Chunk Ev.: ID6-GoogleSearch Original and Chunk Rendering

F.12.1: Original



F.12.2: Chunk Rendering

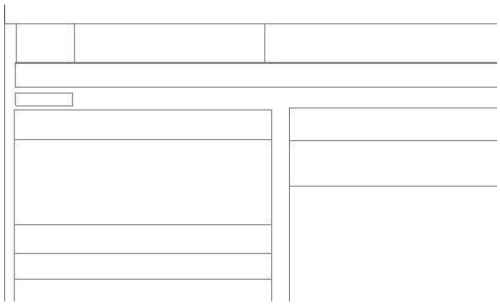
Figure F.12: Chunk Ev.: ID7-MSN Original and Chunk Rendering

F.13.1: Original

F.13.2: Chunk Rendering

Figure F.13: Chunk Ev.: ID8-Orkut Original and Chunk Rendering

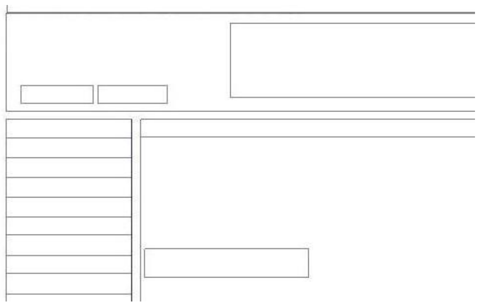
F.14.1: *Original*



F.14.2: *Chunk Rendering*

Figure F.14: Chunk Ev.: ID9-WAI Original and Chunk Rendering

F.15.1: *Original*



F.15.2: *Chunk Rendering*

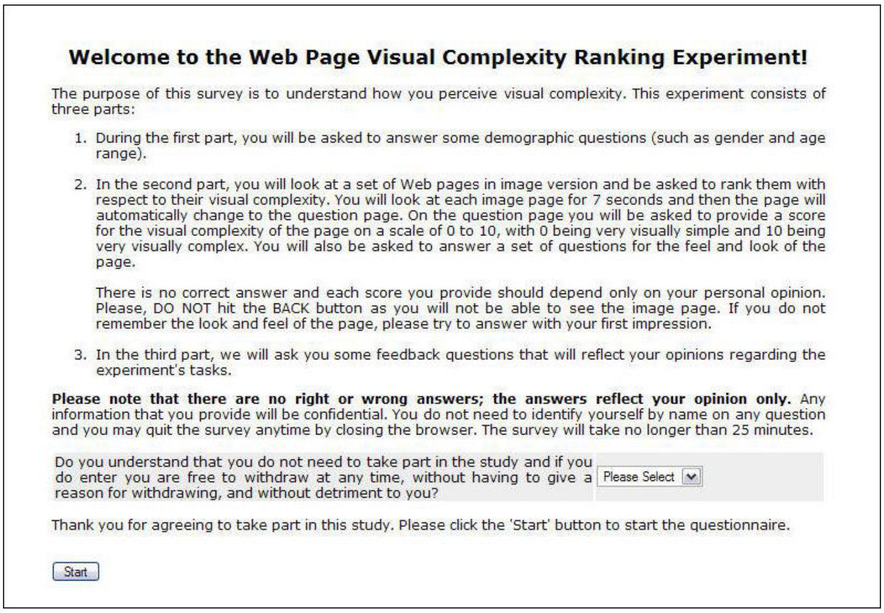
Figure F.15: Chunk Ev.: ID10-Yahoo Original and Chunk Rendering

Appendix G

Complexity and Aesthetics Evaluation

G.1 Online Evaluation Procedure

The Figures below show the stages in which the online experiment consisted. For the Demographic Questionnaire (stage 2) and Thank-You (final stage) screenshots of the experiment see pages 235 and 237 respectively.



Welcome to the Web Page Visual Complexity Ranking Experiment!

The purpose of this survey is to understand how you perceive visual complexity. This experiment consists of three parts:

1. During the first part, you will be asked to answer some demographic questions (such as gender and age range).
2. In the second part, you will look at a set of Web pages in image version and be asked to rank them with respect to their visual complexity. You will look at each image page for 7 seconds and then the page will automatically change to the question page. On the question page you will be asked to provide a score for the visual complexity of the page on a scale of 0 to 10, with 0 being very visually simple and 10 being very visually complex. You will also be asked to answer a set of questions for the feel and look of the page.

   There is no correct answer and each score you provide should depend only on your personal opinion. Please, DO NOT hit the BACK button as you will not be able to see the image page. If you do not remember the look and feel of the page, please try to answer with your first impression.
3. In the third part, we will ask you some feedback questions that will reflect your opinions regarding the experiment's tasks.

Please note that there are no right or wrong answers; the answers reflect your opinion only. Any information that you provide will be confidential. You do not need to identify yourself by name on any question and you may quit the survey anytime by closing the browser. The survey will take no longer than 25 minutes.

Do you understand that you do not need to take part in the study and if you do enter you are free to withdraw at any time, without having to give a reason for withdrawing, and without detriment to you? Please Select

Thank you for agreeing to take part in this study. Please click the 'Start' button to start the questionnaire.

Start

Figure G.1: Introduction to the Online Evaluation - the participant could only proceed if agreed to take part by selecting the 'Yes' option

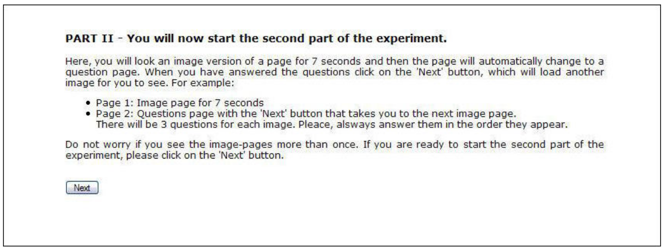


Figure G.2: Introduction to the Second Part of the Evaluation

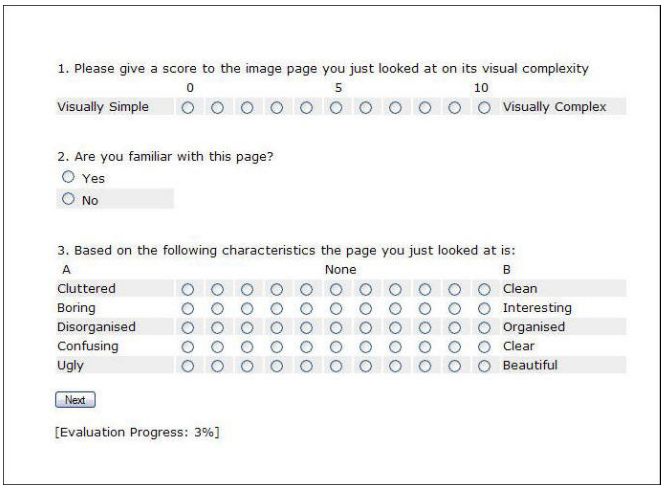


Figure G.3: The questions asked to each participant, which included rankings for visual complexity, familiarity with the page and questions for the page's aesthetic presentation

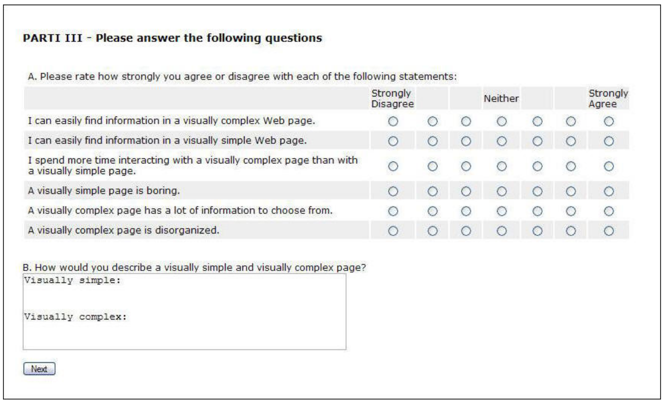


Figure G.4: The questions asked to each participant at the end of the evaluation for further feedback

G.2 Web pages used during the evaluation

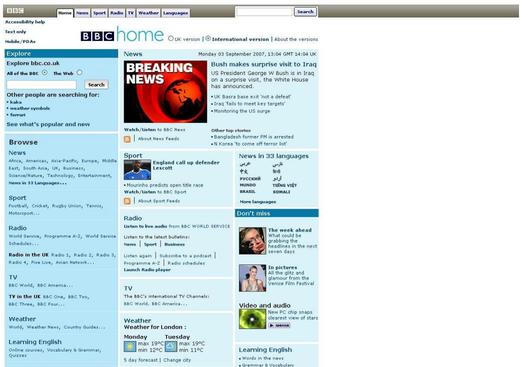
The screenshots shown below of the thirty Web pages used as stimuli for the evaluation. The screenshots can also be found as jpeg in the associated experimental material folder linked with the technical report (see page 265).

G.5.1: ID1 - Amazon

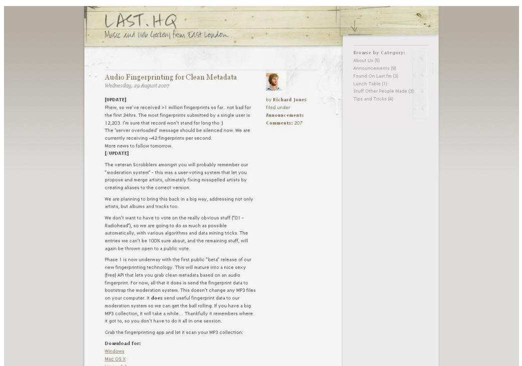
G.5.2: ID2 - Annotea

G.5.3: ID3 - AutoTrader

G.5.4: ID4 - BBC UK News

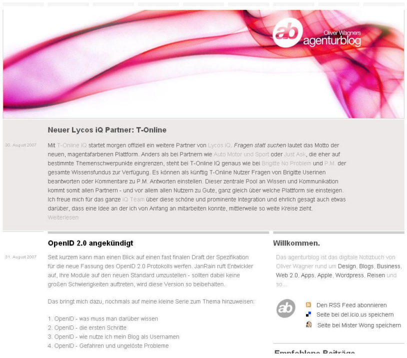


G.5.5: ID5 - BBC UK Home



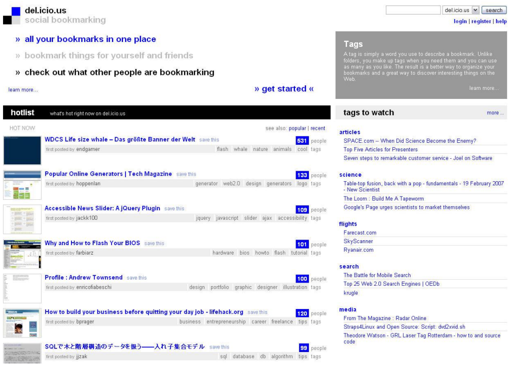
G.5.6: ID6 - Blogger HQ

Figure G.5: Screenshots used for the Complexity and Aesthetics Evaluation



G.5.7: ID7 - Blogger DE

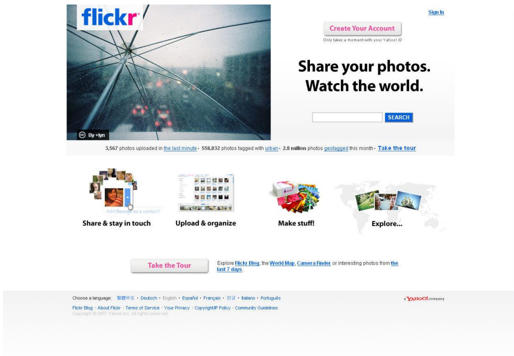
G.5.8: ID8 - Blogger Dashboard



G.5.9: ID9 - Delicious

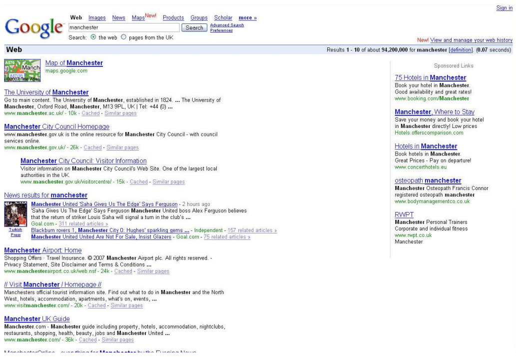
G.5.10: ID10 - Ebay UK

G.5.11: ID11 - Firefox

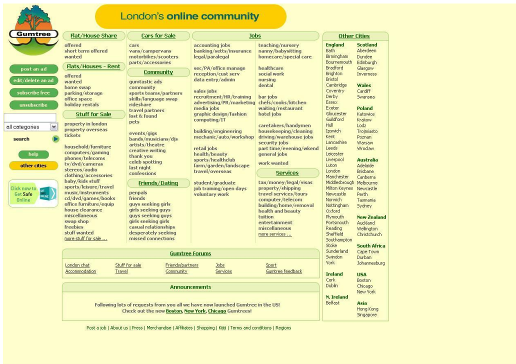


G.5.12: ID12 - Flickr

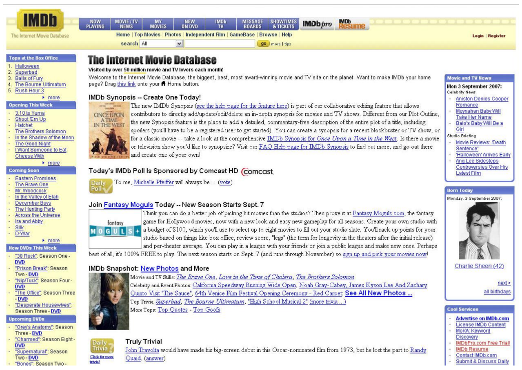
Figure G.5: Screenshots used for the Complexity and Aesthetics Evaluation



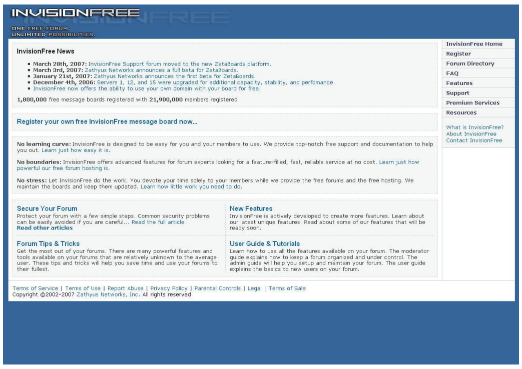
G.5.13: ID13 - Google Search Results



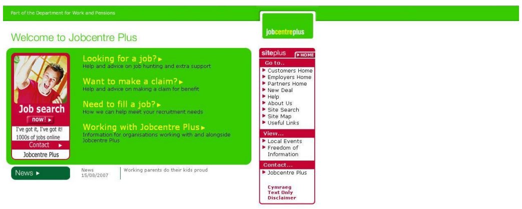
G.5.14: ID14 - GumTree



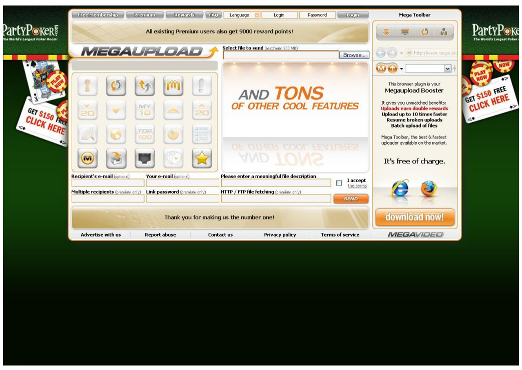
G.5.15: ID15 - Internet Movie Database



G.5.16: ID16 - Invision Free

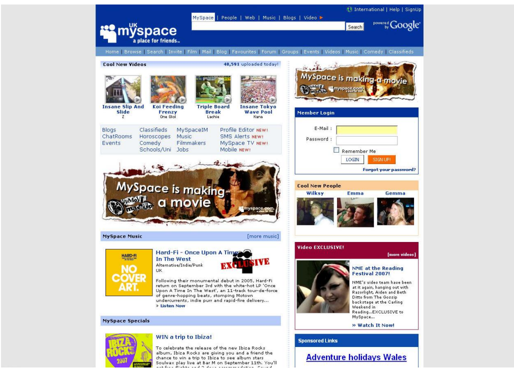


G.5.17: ID17 - Job Centre UK



G.5.18: ID18 - Megaupload

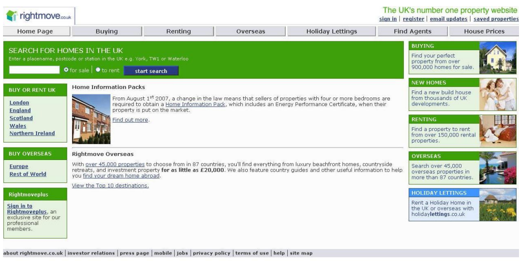
Figure G.5: Screenshots used for the Complexity and Aesthetics Evaluation



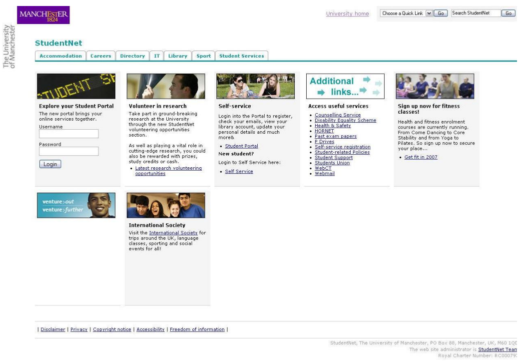
G.5.19: ID19 - MySpace

G.5.20: ID20 - Orkut

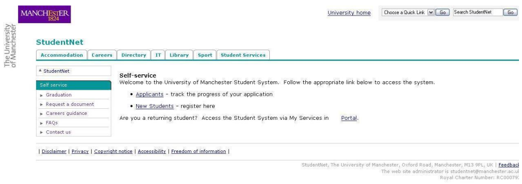
G.5.21: ID21 - Rapidshare



G.5.22: ID22 - Rapidmove UK

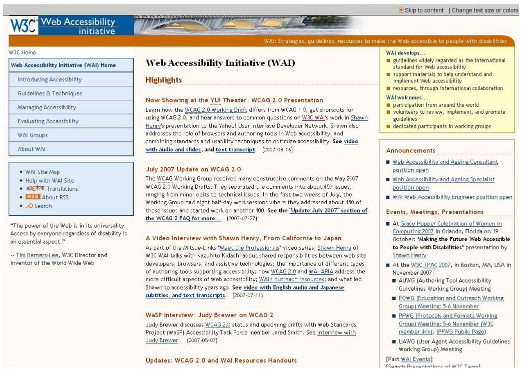


G.5.23: ID23 - University of Manchester Student Net

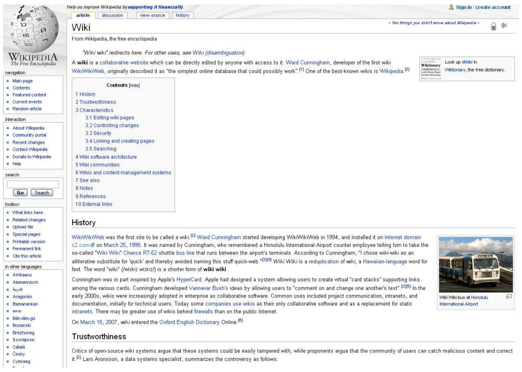


G.5.24: ID24 - University of MAnchester Self Service

Figure G.5: Screenshots used for the Complexity and Aesthetics Evaluation



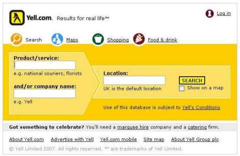
G.5.25: ID25 - Web Accessibility Initiative (WAI)



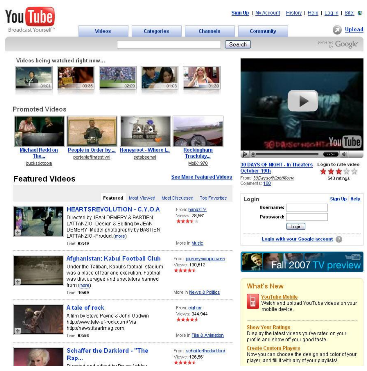
G.5.26: ID26 - Wikipedia Result Page

G.5.27: ID27 - Wikipedia Home

G.5.28: ID28 - Yahoo UK



G.5.29: ID29 - Yell UK



G.5.30: ID30 - YouTube Home

Figure G.5: Screenshots used for the Complexity and Aesthetics Evaluation

G.3 Data

Table G.1: Stepwise Regression - Clean Score Model Summary

Dep.Variable	**Model**	**R**	R^2	**Adj.** R^2	**SE Est.**	ANOVA **F-ratio***	**Durbin-Watson**
Clean B	1	0.702	0.493	0.475	1.049	27.262	
	2	0.761	0.580	0.549	0.973	18.631	
	3	0.821	0.674	0.636	0.874	17.901	
	4	0.857	0.734	0.691	0.805	17.220	2.135
Av. Clean	1	0.704	0.495	0.477	1.031	27.471	
	2	0.757	0.573	0.541	0.966	18.086	
	3	0.813	0.661	0.622	0.877	16.894	
	4	0.850	0.723	0.679	0.808	16.319	2.213

All Significant at $p < 0.01$;
Model 1 Predictors: (Constant), TLC;
Model 2 Predictors: (Constant), TLC, words;
Model 3 Predictors: (Constant), TLC, words, familiarity;
Model 4 Predictors: (Constant), TLC, words, familiarity, images;

Table G.2: Stepwise Regression - Clean Score Model Coefficients

	Clean B					Average Clean				
	B	SE B	β	t	sig.	B	SE B	β	t	sig.
Model 1										
(Constant)	7.910	0.448		17.652	0.000	7.875	0.440		17.882	0.000
TLC	-0.150	0.029	-0.702	-5.221	0.000	-0.148	0.028	-0.704	-5.241	0.000
Model 2										
(Constant)	8.374	0.460		18.213	0.000	8.307	0.457		18.193	0.000
TLC	-0.115	0.030	-0.539	-3.777	0.001	-0.115	0.030	-0.549	-3.816	0.001
Words	-0.003	0.001	-0.336	-2.358	0.026	-0.003	0.001	-0.318	-2.211	0.036
Model 3										
(Constant)	8.326	0.413		20.149	0.000	8.261	0.415		19.917	0.000
TLC	-0.138	0.029	-0.647	-4.824	0.000	0.898	0.345	0.319	2.603	0.015
Words	-0.003	0.001	-0.360	-2.804	0.009	-0.137	0.029	-0.654	-4.782	0.000
Familiarity	0.940	0.344	0.329	2.736	0.011	-0.003	0.001	-0.341	-2.605	0.015
Model 4										
(Constant)	8.407	0.382		21.992	0.000	8.342	0.384		21.735	0.000
TLC	-0.095	0.032	-0.446	-2.982	0.006	1.069	0.326	0.380	3.279	0.003
Words	-0.004	0.001	-0.476	-3.718	0.001	-0.094	0.032	-0.450	-2.946	0.007
Familiarity	1.111	0.325	0.389	3.421	0.002	-0.004	0.001	-0.459	-3.517	0.002
Images	-0.059	0.025	-0.316	-2.372	0.026	-0.060	0.025	-0.322	-2.368	0.026

Table G.3: Enter Method Regression - Interesting Score Model Summary

Dep.Variable	**Model**	**R**	R^2	**Adj.** R^2	**SE Est.**	ANOVA **F-ratio***	**Durbin Watson**
Interesting B	1	0.736	0.542	0.525	0.73371	33.107	1.603
	2	0.495	0.245	0.219	0.942	9.111	
	3	0.608	0.369	0.323	0.876	7.910	1.703
Av. Interesting	1	0.733	0.537	0.521	0.748	32.487	
	2	0.519	0.270	0.244	0.940	10.337	
	3	0.608	0.369	0.322	0.889	7.897	1.675

All Significant at $p < 0.05$;
Model 1 Predictors: (Constant), familiarity;
Model 2 Predictors: (Constant), links;
Model 3 Predictors: (Constant), links, words;

Table G.4: Stepwise Regression - Interesting Score Model Coefficients

	Interesting B					Average Interesting				
	B	SE B	β	t	sig.	B	SE B	β	t	sig.
Model 1										
(Constant)	5.060	0.183		27.587	0.000	4.955	0.187		26.492	0.000
Familiarity	1.545	0.269	0.736	5.754	0.000	1.560	0.274	0.733	5.700	0.000
Model 2										
(Constant)	5.014	0.307		16.336	0.000	4.867	0.306		15.888	0.000
Links	0.017	0.006	0.495	3.018	0.005	0.019	0.006	0.519	3.215	0.003
Model 3										
(Constant)	5.536	0.365		15.175	0.000	5.342	0.370		14.429	0.000
Links	0.026	0.006	0.725	3.974	0.000	0.026	0.007	0.725	3.972	0.000
Words	-0.003	0.001	-0.420	-2.304	0.029	-0.003	0.001	-0.377	-2.063	0.049

Table G.5: Stepwise Method Regression - Organised Score Model Summary

Dep.Variable	Model	R	R^2	Adj. R^2	SE Est.	ANOVA F-ratio*	Durbin-Watson
Organised B	1	0.401	0.161	0.131	0.978	5.372	
	2	0.634	0.402	0.358	0.841	9.078	
	3	0.785	0.616	0.572	0.687	13.896	1.863
Av. Organised	1	0.384	0.147	0.117	0.966	4.835	
	2	0.617	0.381	0.335	0.838	8.318	
	3	0.765	0.585	0.537	0.700	12.204	1.709

Sig. at $p < 0.05$;
Model 1 Predictors: (Constant), words
Model 2 Predictors: (Constant), words, familiarity
Model 3 Predictors: (Constant), words, familiarity, images

Table G.6: Stepwise Regression - Organised Score Model Coefficients

	Interesting B					Average Interesting				
	B	SE B	β	t	sig.	B	SE B	β	t	sig.
Model 1										
(Constant)	7.423	0.399		18.622	0.000	7.297	0.394		18.528	0.000
Words	-0.003	0.001	-0.401	-2.318	0.028	-0.003	0.001	-0.384	-2.199	0.036
Model 2										
(Constant)	7.175	0.351		20.453	0.000	7.057	0.350		20.180	0.000
Words	-0.003	0.001	-0.517	-3.383	0.002	-0.003	0.001	-0.498	-3.202	0.003
Familiarity	1.043	0.316	0.505	3.299	0.003	1.007	0.315	0.497	3.195	0.004
Model 3										
(Constant)	7.595	0.307		24.734	0.000	7.459	0.313		23.838	0.000
Words	-0.004	0.001	-0.566	-4.505	0.000	-0.004	0.001	-0.545	-4.177	0.000
Familiarity	1.406	0.275	0.680	5.108	0.000	1.354	0.281	0.668	4.827	0.000
Images	-0.067	0.018	-0.493	-3.804	0.001	-0.064	0.018	-0.481	-3.569	0.001

Table G.7: Stepwise Method Regression - Clear Score Model Summary

Dep.Variable	Model	R	R^2	Adj. R^2	SE Est.	ANOVA F-ratio*	Durbin-Watson
Clear B	1	0.520	0.270	0.244	1.064	10.358	
	2	0.658	0.433	0.391	0.955	10.313	
	6	0.803	0.645	0.604	0.770	15.732	1.939
Av. Clear	3	0.462	0.213	0.185	1.084	7.590	
	4	0.670	0.448	0.408	0.924	10.973	
	5	0.731	0.534	0.480	0.866	9.931	
	6	0.780	0.608	0.563	0.794	13.431	1.742

Sig. at $p < 0.01$;
Model 1 Predictors: (Constant), words
Model 2 Predictors: (Constant), words, familiar
Model 3 Predictors: (Constant), TLC
Model 4 Predictors: (Constant), TLC, familiarity
Model 5 Predictors: (Constant), TLC, familiarity, words
Model 6 Predictors: (Constant), familiarity, words, images

Table G.8: Stepwise Regression - Clear Score Model Coefficients

	Clear B					Average Clear				
	B	SE B	β	t	sig.	B	SE B	β	t	sig.
Model 1										
(Constant)	7.423	0.399		18.622	0.000					
Words	-0.003	0.001	-0.401	-2.318	0.028					
Model 2										
(Constant)	7.175	0.351		20.453	0.000					
Words	-0.003	0.001	-0.517	-3.383	0.002					
Familiarity	1.043	0.316	0.505	3.299	0.003					
Model 3										
(Constant)						7.172	0.463		15.488	0.000
TLC						-0.082	0.030	-0.462	-2.755	0.010
Model 4										
(Constant)						7.067	0.396		17.842	0.000
Familiarity						1.231	0.363	0.520	3.392	0.002
TLC						-0.115	0.027	-0.650	-4.238	0.000
Model 5										
(Constant)						7.446	0.410		18.180	0.000
Familiarity						1.281	0.341	0.541	3.760	0.001
TLC						-0.087	0.028	-0.494	-3.086	0.005
Words						-0.003	0.001	-0.335	-2.186	0.038
Model 6										
(Constant)	7.595	0.307		24.734	0.000	7.365	0.355		20.735	0.000
Familiarity	-0.004	0.001	-0.566	-4.505	0.000	1.417	0.318	0.599	4.449	0.000
Words	1.406	0.275	0.680	5.108	0.000	-0.005	0.001	-0.597	-4.708	0.000
Images	-0.067	0.018	-0.493	-3.804	0.001	-0.082	0.020	-0.527	-4.026	0.000

Table G.9: Stepwise Method Regression - Beautiful Score Model Summary

Model	**R**	R^2	**Adj.** R^2	**SE Est.**	ANOVA **F-ratio**	**Durbin-Watson**
1	0.487	0.237	0.210	0.888	8.709*	
2	0.632	0.400	0.356	0.802	9.001**	1.492

Sig. at $p < 0.01$; ** Sig. at $p < 0.001$;
Model 1 Predictors: (Constant), familiarity;
Model 2 Predictors: (Constant), familiarity, TLC;

Table G.10: Stepwise Regression - Beautiful Score Model Coefficients

	B	SE B	β	t	sig.
Model 1					
(Constant)	4.603	0.222		20.737	0.000
Familiarity	0.959	0.325	0.487	2.951	0.006
Model 2					
(Constant)	5.358	0.344		15.595	0.000
Familiarity	1.267	0.315	0.643	4.025	0.000
TLC	-0.064	0.023	-0.433	-2.707	0.012

Table G.11: Correlation between Independent Variable (structural elements) and Aesthetic Scores (mean values)

Variables	**Clean**	**Interesting**	**Organised**	**Clear**	**Beautiful**
Menus	-0.302	0.265	-0.016	-0.095	-0.048
Images	-0.423^b	0.294	-0.249	-0.318	0.047
Words	-0.585^a	0.020	-0.384^b	-0.451^b	-0.275
TLC	-0.704^a	0.277	-0.345	-0.462^c	-0.200
Links	-0.342	0.519^c	0.004	-0.095	0.174
Familiarity	0.005	0.733^a	0.382^b	0.285	0.487^c

a. $p < .01$; b. $p < .05$; c. $p < .001$

Table G.12: Aesthetic Characteristics - Mean Values per Score Version per Web PageID

PageID	Complexity	Clean			Interesting			Organised			Clear			Beautiful		
		A	B	Mean	A	B	Mean	A	B	Mean	A	B	Mean	A	B	Mean
1	C	5.09	5.15	5.12	6.40	6.60	6.50	6.87	6.71	6.79	6.25	5.76	6.01	5.47	5.27	5.37
2	S	6.45	6.13	6.29	4.02	4.15	4.08	6.53	6.31	6.42	6.24	5.89	6.06	3.71	3.67	3.69
3	C	3.42	3.62	3.52	4.96	5.29	5.13	4.87	5.00	4.94	4.20	4.29	4.25	3.89	3.53	3.71
4	C	5.07	5.33	5.20	7.33	7.55	7.44	6.96	7.09	7.03	6.18	6.35	6.26	5.71	5.80	5.75
5	M	5.22	5.09	5.15	6.80	6.95	6.87	6.58	6.93	6.75	5.98	6.15	6.06	5.73	5.64	5.68
6	S	6.96	7.05	7.01	5.47	5.51	5.49	6.53	6.75	6.64	6.29	6.44	6.36	5.93	5.75	5.84
7	S	5.82	6.69	6.25	4.84	5.09	4.96	5.78	6.00	5.89	5.35	5.93	5.64	5.67	6.05	5.86
8	M	5.51	5.24	5.37	4.85	4.89	4.87	5.53	5.85	5.69	5.27	5.29	5.28	4.33	4.40	4.36
9	M	4.27	4.53	4.40	3.96	4.69	4.33	5.09	5.22	5.15	4.51	4.55	4.53	4.00	3.95	3.97
10	C	5.45	5.67	5.56	6.53	6.91	6.72	6.84	6.76	6.80	6.15	6.47	6.31	5.84	5.73	5.78
11	S	7.60	7.58	7.59	6.53	6.80	6.66	7.80	8.00	7.90	7.65	7.73	7.69	6.85	6.84	6.85
12	S	7.45	7.51	7.48	6.80	7.07	6.94	7.49	7.64	7.56	7.20	7.38	7.29	6.60	6.76	6.68
13	M	6.58	6.67	6.63	6.13	6.75	6.44	7.56	7.89	7.73	7.49	7.47	7.48	5.09	5.75	5.42
14	M	4.18	4.11	4.15	4.87	5.16	5.02	6.00	6.02	6.01	4.98	5.11	5.05	4.42	4.33	4.37
15	C	4.13	3.75	3.94	5.85	5.95	5.90	5.33	5.73	5.53	4.82	4.44	4.63	4.62	4.38	4.50
16	M	4.38	4.18	4.28	3.33	3.36	3.35	4.85	4.69	4.77	4.27	3.91	4.09	3.25	3.04	3.15
17	M	6.49	6.91	6.70	5.31	5.47	5.39	6.55	7.20	6.87	6.20	6.93	6.56	3.98	4.25	4.12
18	C	4.04	3.85	3.95	4.11	4.36	4.24	4.42	4.53	4.47	3.33	4.04	3.68	3.56	3.84	3.70
19	C	3.89	4.00	3.95	5.58	6.05	5.82	5.11	5.67	5.39	4.51	5.05	4.78	4.27	4.51	4.39
20	S	8.62	8.27	8.45	4.36	5.00	4.68	7.60	7.78	7.69	7.25	7.67	7.46	5.15	5.53	5.34
21	S	5.98	6.07	6.03	4.91	5.13	5.02	5.93	6.35	6.14	5.78	5.82	5.80	4.42	4.53	4.47
22	M	6.55	6.25	6.40	5.71	5.93	5.82	6.93	7.09	7.01	6.73	6.73	6.73	5.20	4.89	5.05
23	M	6.42	6.38	6.40	6.07	5.98	6.03	6.84	7.04	6.94	6.51	6.65	6.58	5.35	5.38	5.36
24	S	7.91	8.24	8.07	4.65	5.44	5.05	7.85	8.18	8.02	7.65	7.89	7.77	4.75	5.18	4.96
25	M	4.55	4.22	4.38	4.58	4.13	4.35	5.60	5.22	5.41	4.95	4.42	4.68	4.00	3.60	3.80
26	M	6.13	5.96	6.05	7.15	6.96	7.05	7.44	7.51	7.47	6.82	6.89	6.85	6.02	5.67	5.85
27	S	7.78	7.87	7.83	7.15	7.13	7.14	7.89	7.96	7.93	7.73	7.69	7.71	7.09	7.02	7.05
28	C	4.53	4.42	4.47	6.45	6.51	6.48	6.20	6.13	6.16	5.22	5.38	5.30	5.36	5.31	5.34
29	S	7.91	7.80	7.85	5.49	5.73	5.61	7.85	7.96	7.91	7.78	7.80	7.79	5.47	5.60	5.54
30	C	5.07	5.29	5.18	7.33	6.91	7.12	6.62	6.71	6.66	5.75	5.96	5.85	5.51	5.60	5.55

Appendix H

ViCRAM Tool Installation

At the time this thesis was submitted, the ViCRAM tool was in the process of being committed in Eclipse ACTF. Based on the latest official installation instructions one can download and use the latest available prototype version of the tool as follows:

1. Download Eclipse Classic 3.5 or Eclipse for RCP/Plug-in Developers 3.5 from: http://www.eclipse.org/downloads/.

2. Unzip and launch Eclipse

3. Install ACTF Visualization SDK from Eclipse Update site. Instructions are available at: http://download.eclipse.org/technology/actf/0.7/milestones/

   One can also install it by using the following steps at this stage:

   (a) Open Help > Install New Software menu

   (b) Add 'Name: Galileo, Location: http://download.eclipse.org/releases/galileo' from 'Work with:' pull down menu and wait for a while.

   (c) Select 'ACTF Visualization SDK Feature (Incubation)' from 'General Purpose Tools' category, then click Next.

   (d) Follow instruction of installation wizard.

4. Check out `org.eclipse.actf.examples.vicram` and `org.eclipse.actf.examples.vicram-feature` projects

   One can also import the project by:

   (a) Import ViCRAM project (ViCRAM.psf) into Workspace by using 'Import > Team > Team Project Set' menu. The folder 'ViCRAM.psf' can be downloaded from the associated materials of the tool's technical report (see page 264).

5. Open `vicram.product` file in the `org.eclipse.actf.examples.vicram` project.

6. Select 'Synchronize' and 'Launch an Eclipse application' to launch the tool from Eclipse

7. Select 'Eclipse Product export wizard' to export the tool from Eclipse

Appendix I

Technical Evaluation

Table I.1: Tool Technical Evaluation Data

		Visual Complexity Score		Clean		Interesting		Organised		Clear	
Spearman's ρ - Coefficient		**.559[b]**		.524[b]		0.372		.453[b]		0.338	
PageID	PageName	Study	Tool	Study	Tool	Study	Tool	Study	Tool	Study	Tool
1	Amazon UK	5.09	55.59	5.12	4.07	6.5	7.51	6.79	4.32	6.01	6.27
2	AnnoteaProject	3.43	43.52	6.29	4.08	4.08	4.74	6.42	5.03	6.06	5.78
3	AutoTrader	6.86	62.39	3.52	2.8	5.13	5.87	4.94	3.98	4.25	5.61
4	BBC UK News	5.81	42.11	5.2	6.04	7.44	7.52	7.03	5.03	6.26	7.17
5	BBC UK	5.62	28.92	5.15	4.92	6.87	7.91	6.75	5.86	6.06	5.78
6	BloggerPostHQ	3.42	.	7.01	.	5.49	.	6.64	.	6.36	6.76
7	BloggerPostDE	3.96	67.21	6.25	2.09	4.96	6.63	5.89	3.72	5.64	5.29
8	Blogger Dashboard	4.75	37.43	5.37	5.63	4.87	4.92	5.69	5.31	5.28	6.68
9	Delicious	5.63	25.77	4.4	6.62	4.33	7.85	5.15	5.96	4.53	6.93
10	Ebay	5.41	27.54	5.56	6.86	6.72	6.7	6.8	5.85	6.31	7.17
11	Firefox	3.09	14.75	7.59	7.02	6.66	5.31	7.9	6.59	7.69	6.76
12	Flickr	3.06	7.08	7.48	8.01	6.94	5.9	7.56	7	7.29	7.17
13	GoogleSearch	3.78	32.82	6.63	6.55	6.44	5.53	7.73	5.54	7.48	7.17
14	GumTree	5.76	25	4.15	6.99	5.02	10.06	6.01	5.98	5.05	7.17
15	IMDB	6.09	56.82	3.94	3.88	5.9	7.05	5.53	4.25	4.63	6.19
16	InvisionFree	5.44	22.62	4.28	7.02	3.35	5.05	4.77	6.12	4.09	7.09
17	Job Centre	3.26	8.11	6.7	7.95	5.39	5.9	6.87	6.94	6.56	7.17
18	MegaUpload	6.44	.	3.95	.	4.24	.	4.47	.	3.68	.
19	MySpace	6.25	27.72	3.95	6.84	5.82	6.37	5.39	5.83	4.78	7.17
20	Orkut	1.61	5.78	8.45	8.08	4.68	5.3	7.69	7.07	7.46	7.17
21	Rapidshare	4	15.52	6.03	7.42	5.02	5.06	6.14	6.52	5.8	7.09
22	Rightmove	4.18	16.56	6.4	6.81	5.82	5.58	7.01	6.49	6.73	6.68
23	StudentNet	3.92	15.53	6.4	6.65	6.03	5.67	6.94	6.56	6.58	6.52
24	StudentNet SelfService	2.08	11.7	8.07	7.63	5.05	5.75	8.02	6.74	7.77	7.09
25	WAI	6	66.77	4.38	2.33	4.35	4.9	5.41	3.73	4.68	5.45
26	Wiki Result	4.46	.	6.05	.	7.05	5.89	7.47	.	6.85	5.86
27	Wikipedia	2.6	32.47	7.83	6.58	7.14	9.77	7.93	5.57	7.71	7.17
28	Yahoo UK	6.21	30.42	4.47	5.05	6.48	7.65	6.16	5.76	5.3	5.94
29	Yell	2.2	22.9	7.85	6.89	5.61	5.93	7.91	6.11	7.79	7.01
30	YouTube	5.47	105.07	5.18	2.37	7.12	4.36	6.66	1.47	5.85	7.09

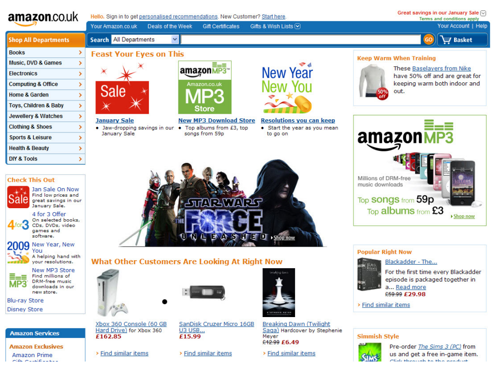
a. $p < .05$; b. $p < .01$;

Appendix J

Algorithm Evaluation

The screenshots shown below of the thirty Web pages used as stimuli for the evaluation and were taken in January 2009. The screenshots can also be found as jpeg in the associated experimental material folder linked with the respective technical report (see page 264).

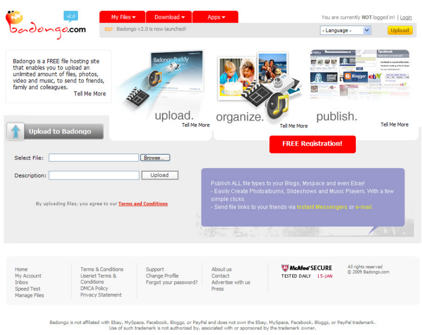
J.1.1: ID1 - 4Shared



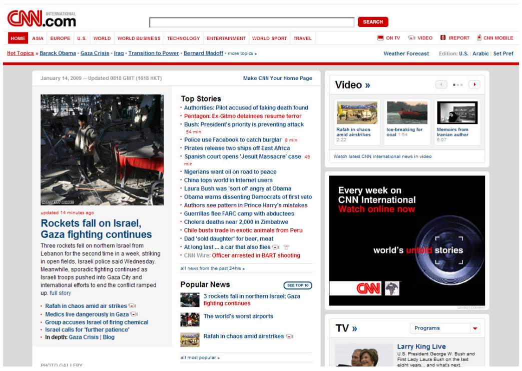
J.1.2: ID2 - Amazon UK

J.1.3: ID3 - AOL UK

Figure J.1: Screenshots used for the Algorithm User Evaluation

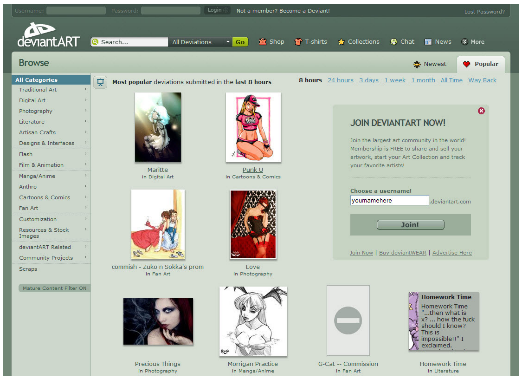


J.1.4: ID4 - Badongo



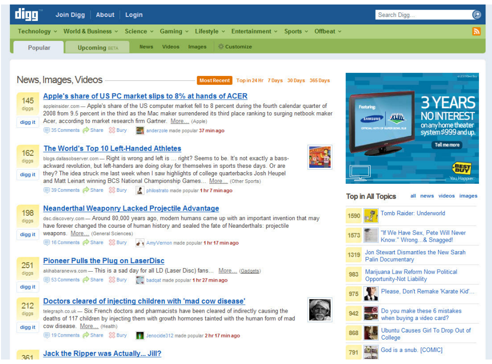
J.1.5: ID5 - CNN International

J.1.6: ID6 - Comcast



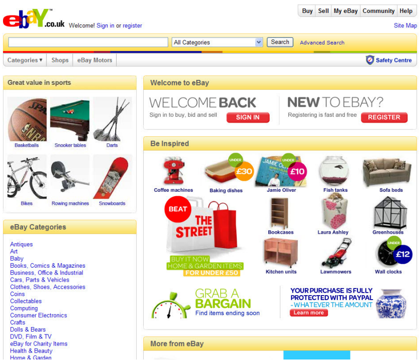
J.1.7: ID7 - DevianArt

Figure J.1: Screenshots used for the Algorithm User Evaluation

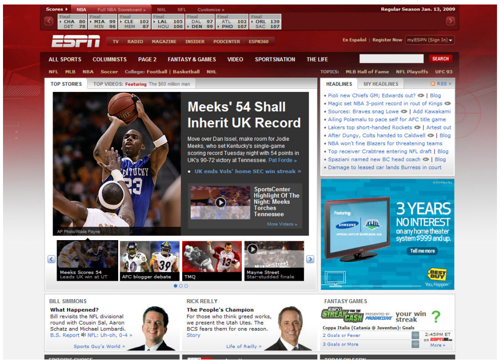


J.1.8: ID8 - Digg

J.1.9: ID9 - Download



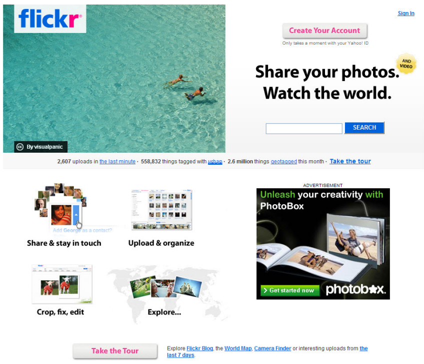
J.1.10: ID10 - Ebay UK



J.1.11: ID11 - ESPN

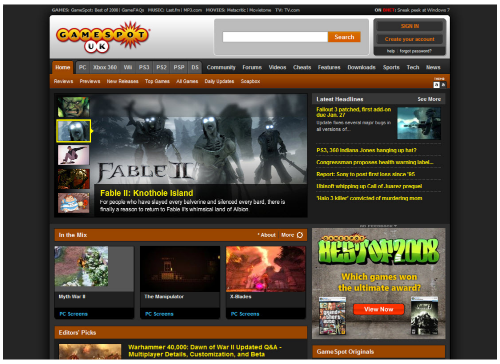
J.1.12: ID12 - Filefactory

Figure J.1: Screenshots used for the Algorithm User Evaluation



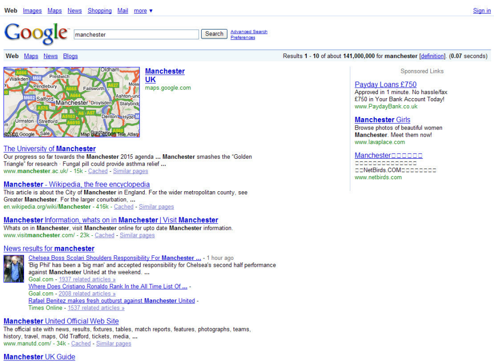
J.1.13: ID13 - Flickr

J.1.14: ID14 - Fotolog

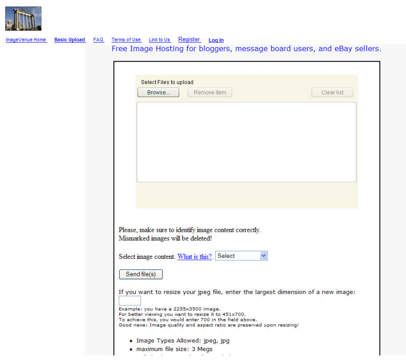


J.1.15: ID15 - GameSpot

J.1.16: ID16 - GO



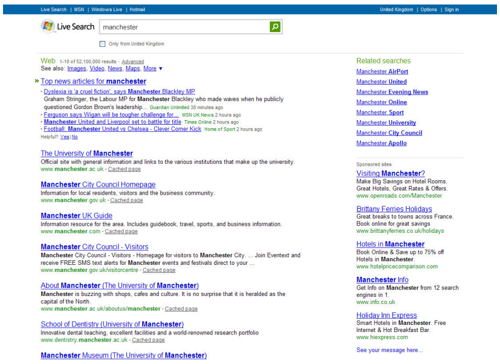
J.1.17: ID17 - Google Search Results



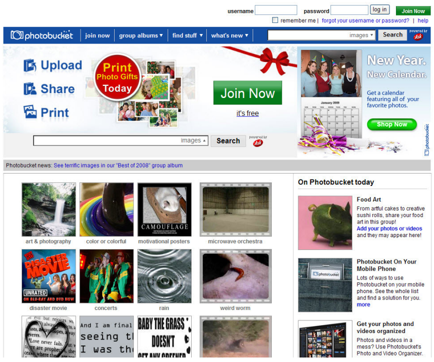
J.1.18: ID18 - Image Venue

Figure J.1: Screenshots used for the Algorithm User Evaluation

J.1.19: ID19 - Internet Movie Database

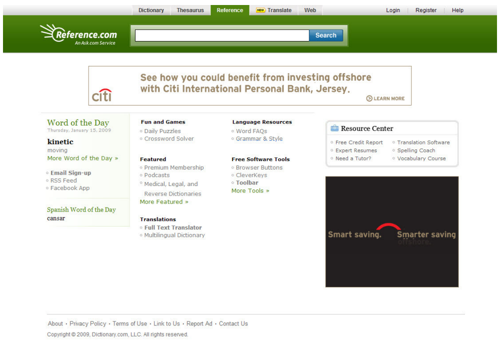


J.1.20: ID20 - Live Search Result (Manchester)



J.1.21: ID21 - Photobucket

J.1.22: ID22 - Rapidshare

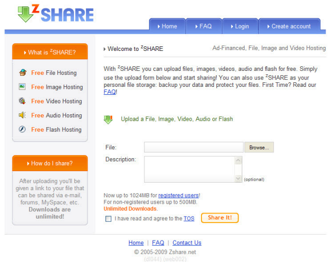


J.1.23: ID23 - Reference

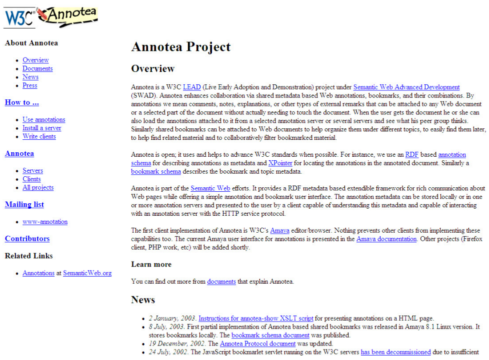
J.1.24: ID24 - WordPress

Figure J.1: Screenshots used for the Algorithm User Evaluation

J.1.25: ID25 - Yahoo UK

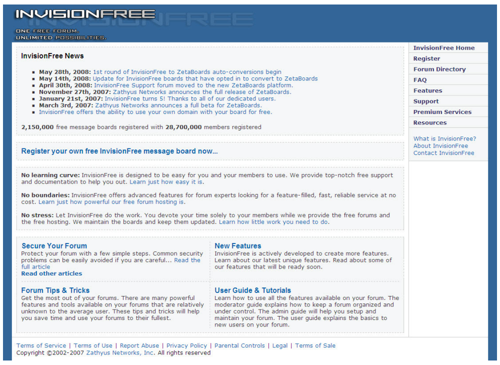


J.1.26: ID26 - zShare



J.1.27: ID27 - Annotea

J.1.28: ID28 - GumTree



J.1.29: ID29 - Invision Free

J.1.30: ID30 - Job Centre Plus

Figure J.1: Screenshots used for the Algorithm User Evaluation

Appendix K

Technical Reports

A list of technical reports written throughout the project:

1. Eleni Michailidou. The User Evaluation of the ViCRAM Visual Complexity Prediction Algorithm. Technical Report. Human Centred Web Lab, School of Computer Science. 2009.
   http://hcw-eprints.cs.manchester.ac.uk/112/

   This report describes the user evaluation of the visual complexity algorithm that was implemented into the ViCRAM tool. In order for the tool to be efficiently used by users and designers, the algorithm must provide realistic information. The user evaluation demonstrated that the models previously predicted, during the ViCRAM project, were appropriately implemented into the ViCRAM tool and therefore the tool provides information about the visual complexity of Web pages that is significantly correlated with users' perception. To achieve this, users were asked to rank Web pages based on their visual complexity. These rankings were then compared with the complexity information that the ViCRAM tool provides for the same pages.

2. Eleni Michailidou. The ViCRAM Tool Implementation - Visual Complexity Algorithm. Technical Report. Human Centred Web Lab, School of Computer Science. 2009.
   http://hcw-eprints.cs.manchester.ac.uk/105/

   Part of the ViCRAM project is to determine how Web page structural elements (such as font, tables, links, and images) and their characteristics (such as colour and size) can be used to determine the visual presentation and complexity level of a Web page. Through a sequence of user evaluations and eye tracking studies, statistical models, based on the structural elements of the page, were derived that siginificantly predict sighted users' perception on Web page visual complexity and aesthetic appearance. This report describes the implementation of these models into the ACTF Eclipse framework by extending the aDesigner accessibility tool to the ViCRAM tool. For each Web page, users and designers are provided with a complexity score that determines the page's level of visual complexity and an overlay heatmap that notes the areas that are most visually complex. A technical evaluation of the tool was executed that revealed outstanding issues that helped to improve the tool but also that the complexity score prediction is significantly correlated with the user scores.

3. Eleni Michailidou. Determining Users' Perception of Web Page Visual Complexity and Aesthetic Characteristics. Technical Report. Human Centred Web Lab, School of Computer Science. 2008.
   http://hcw-eprints.cs.manchester.ac.uk/104/

   The visual appearance of a Web page influences the way a user will interact with the page. Part of the ViCRAM project is to determine how Web page structural elements (such as font, tables, links, and images) and their characteristics (such as colour and size) can be used to determine the visual presentation and complexity level of a Web page. This report describes an investigation into how users perceive the visual complexity and aesthetic appearance of Web pages. Results show a strong correlation between users visual complexity perception of a Web page, structural elements and aesthetic appearance. Data analysis derived models that estimate the visual complexity and aesthetic level of a Web page using the structural elements of the page.

4. Eleni Michailidou. Investigating Users' Visual Perception based on Web page Chunks. Technical Report. Human Centred Web Lab, School of Computer Science. 2008. http://hcw-eprints.cs.man.ac.uk/103/

   The visual appearance of aWeb page influences the way a user will interact with the page. A pilot study, showed that the number of different sections a page is organized into affects user perception of the visual complexity level of the page. That is, the more chunks (sections) a Web page is arranged into, the more visually complex it becomes. This report describes an online pairwise experiment that extends the pilot study. A set of variables are defined and used to modify a Web page into its chunk rendering. Corner, blocks, boxes and top-left-corners are also defined and identified on a set of Web pages chunk renderings. Users had to compare this set of Web pages based on their visual complexity. Results show that the visual complexity score is significantly related with the organisation of the page determined by the number of boxes, blocks, corners and top-left-corners.

5. Eleni Michailidou. A Pilot Eye-Tracking Study: Understanding How Visually Complex Web Pages Influence Visual Attention. Technical Report. Human Centred Web Lab, School of Computer Science. 2008. http://hcw-eprints.cs.man.ac.uk/102/

   The World Wide Web has become the means of distribution and use of information by individuals, teams, organizations and communities. The way information is presented on the web is becoming more and more complex making access to information harder for web users with disabilities, especially visually impaired. Visually complex and simple web pages affect the way a user reads the page. This is an initial study using eye movement tracking methods. It provides supportive information on how sighted users perceive visual presentation of web pages, including where they glance first when they reach a page and for how long they fixate on specific areas of a web page. Qualitative analysis of this study's results gave feedback on the order of reading visually simple and complex web pages.

6. Darren Lunn, Eleni Michailidou. Observational Notes Acquired from Henshaws' Skillstep to Success Class: Observation Period 2. Technical Report. Human Centred Web Lab, School of Computer Science. 2008. http://hcw-eprints.cs.man.ac.uk/64/

   Henshaws' Society for Blind People is a charity for visually impaired people of all ages. One of the facilities the charity provides for its members is an IT course entitled "Skillstep to Success". The authors of this report spent four sessions at Henshaws' acting as observers and classroom assistants in the Skillstep class. The sessions lasted for two hours each, during which time the authors discussed accessibility issues with students and staff of the class in addition to observing students accessing the Web. This report presents the combined observations of the two authors made during the sessions in a cohesive manner but does not aim to provide any analysis or draw any conclusions from the visits to Henshaws'.

7. Eleni Michailidou, Huangmao Quan, Simon Harper. Web Page Visual Complexity: A Relationship Between Two Classification Methods. Technical Report. Human Centred Web Lab, School of Computer Science. 2008. http://hcw-eprints.cs.man.ac.uk/100/

   This is a short report on the comparison of two classification methods that are used to identify the visual complexity level of a set of Web pages. Ranking data are collected from two approaches: programmatic and perceptual. These ranks are analyzed to compare the correlation of the prediction using both approaches. It was determined that the ranking order assigned to the set of Web pages from the two approaches are significantly correlated.

8. Darren Lunn, Eleni Michailidou. Observational Notes Acquired from Henshaws' Skillstep to Success Class: Observation Period 1. Technical Report. Human Centred Web Lab, School of Computer Science. 2007. http://hcw-eprints.cs.man.ac.uk/61

   Henshaws' Society for Blind People is a charity for visually impaired people of all ages. One of the facilities the charity provides for its members is an IT course entitled "Skillstep to Success". The authors of this report spent seven sessions at Henshaws' acting as observers and, in some cases, classroom assistants in the Skillstep class. The sessions lasted for two hours each, during which time the authors discussed accessibility issues with students and staff of the class in addition to observing students accessing the Web. This report presents the combined observations of the two authors made during the sessions in a cohesive manner but does not aim to provide any analysis or draw any conclusions from the visits to Henshaws'.