How People use Presentation to Search for a Link: Expanding the Understanding of Accessibility on the Web

CAROLINE JAY

caroline.jay@manchester.ac.uk

ROBERT STEVENS

robert.stevens@manchester.ac.uk

MASHHUDA GLENCROSS

mashhuda@manchester.ac.uk

University of Manchester, Oxford Road, Manchester M13 9PL

ALAN CHALMERS

alan@compsci.bristol.ac.uk

CATHY YANG

yang@compsci.bristol.ac.uk

University of Bristol, Tyndall Avenue, Bristol BS8 1TH

Abstract

It is well known that many Web pages are difficult for visually disabled people to use. Without access to a rich, visual display, the intended structure and organisation of the page is obscured. To fully understand what is missing from the experience of visually disabled users, it is pertinent to ask how the presentation of Web pages on a standard display makes them easier for sighted people to use. Here, we report on an exploratory eye-tracking study that addresses this issue by investigating how sighted readers use the presentation of the BBC News Web page to search for a link. We compare the standard page presentation with a "text-only" version and observe both qualitatively and quantitatively that the removal of the intended presentation alters "reading" behaviours. The demonstration that the presentation of information assists task completion suggests that it should be re-introduced to non-visual presentations if the Web is to become more accessible. We also explore the extent to which algorithms that generate maps of what is perceptually salient on a page match the gaze data recorded in the eye tracking study. The correspondence between a page's presentation, knowledge of what is visually salient and how people use these features to complete a task might offer an opportunity to re-model a Web page to maximise access to the most important parts of that page.

1 Introduction

In this paper we explore the problems encountered by people with profound visual disabilities when browsing Web pages. We ask two questions: firstly, what is it about an audio screenreader presentation of a Web page that really makes it difficult to use (assuming the page has been authored for screenreader use);

secondly, what is it about the standard presentation that is missing and needs to be replaced? The obvious and unattainable facility that is lacking in a screen reader presentation is the external memory provided by paper or a screen, together with the speed and accurate control of information flow afforded by the human visual system [17]. All assistive technology for visually disabled people attempts, in some way, to provide some kind of replacement for this, but the gulf between sighted and visually disabled users' experience of a web page remains enormous.

The difficulties visually disabled people have accessing Web content are well documented [6, 5]. However, what is less well documented is the exact nature of the problems encountered by those users. Profoundly visually disabled people usually use a screenreader (such as Jaws [1, 8]), which speaks screen content under the direction of the user, in order to "read" what is on the screen. When a Web page is loaded into a browser, by default the page contents are typically spoken from top-left to bottom-right. Naturally, such a spoken presentation is difficult to use—memory cannot handle such quantity and review is not really possible. The screenreader allows finer control over what is spoken than this cumbersome mechanism. A user can move around a page using cursor keys at the level of lines, characters, words, paragraphs, *etc*. It would seem that all parts of a page can be accessed; so, what is the problem?

Unfortunately, providing access to the individual parts of a page is not the whole solution. Much has been achieved in terms of access to content – not just to text [19], but also to graphics, tables and figures [29, 30, 31, 32]. However, cues indicating the overall structure of the information, provided in visual presentations by the layout of the page, are still missing from non-visual presentations.

If we are to support visually disabled users' access to Web pages it is important to characterize the situation beyond "difficult". A fundamental part of understanding this problem is gaining insight into exactly how sighted users use Web presentation to access the content. Information about this process will arm us with the knowledge to reintroduce not the presentational qualities themselves, but the facilities that the formatting and layout provide to users. It will not tell us *how* to achieve our goal, but it will provide the pre-requisite of knowledge of the features an adequate presentation should afford. Eye tracking provides a valuable means of obtaining such information. In this paper, we examine eye-tracking data during the exercise of basic Web based tasks on standard and text-only versions of a page—which features are looked at and for how long in each presentational style?

In the current study, we investigate this by comparing eye movements on a 'standard' Web page with those on a 'text-only' page that has exactly the same content, but has the majority of the formatting removed. The pages are taken from the BBC News website, which has a filter program called *betsie* that is designed to make pages more screen reader friendly[1]. Betsie removes images and other formatting unnecessary for screenreader use, and presents the content at the top of the page and the main menu at the bottom. The user is given what amounts to a traditional linear document, but with little differentiation between text serving different purposes, such as headings, plain text and emphasized text that usually supports the processing of the printed information [20].

---

[1] http://www.bbc.co.uk/education/betsie/

Figure 1: The standard version of the Web page used in the study.

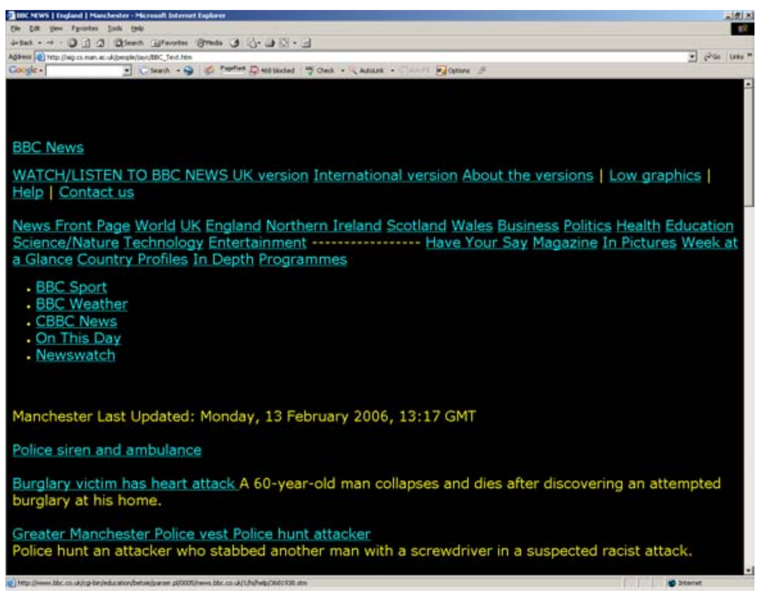

Figure 2: The text-only version of the Web page used in the study.

A standard Web page is shown in Figure 1. Figure 2 illustrates its text-only presentation, where the majority of the formatting and layout is removed. The presentation of the standard page generally facilitates a sighted user in achieving their goal, be it searching, browsing, following links *etc*. It is well known from the principles of graphic design and information presentation that how content is laid out can influence how a task is performed and what interpretation is made of data [20]. By comparing gaze patterns on each type of page, we hope to increase our understanding of exactly *how* presentation helps users to accomplish their task.

Once we understand the benefits that visually disabled users are denied through a lack of visual presentation, we can start the difficult task of re-introducing the requisite facilities into audio presentations. In order to accomplish this, it is necessary to find a means of automating the process. A starting point for this may be to use the notion of *saliency* to identify the areas of a page that users are likely to look at.

Sighted humans have the ability to glance at and automatically focus their attention on the most interesting parts of the environment, using a combination of top-down and bottom-up visual processing. Bottom-up models of visual attention suggest that people spend the greatest amount of time focusing on the most salient parts of an image, on the grounds that they will be most interesting. For example, a red apple (a source of nutrition) is more visually salient, and therefore attractive, than the green leaves surrounding it. Top-down models, on the other hand, explain visual search driven by semantics, or knowledge about the environment: when asked to describe the emotion of a person in a picture, for instance, people will automatically look to the person's face, rather than their clothes.

Both forms of visual attention will inevitably play a part in helping people to navigate and understand Web pages. Bottom-up processing would allow people to quickly detect items such as bold text and images, which help to explain how the content is organized. It also helps people to group the information into 'sections', such as blocks of text, headings and menus. Top-down processing would enable people to interpret the information using prior knowledge and heuristics. For example, people may look for menus at the top and sides of the page, and main content in the middle.

It is possible to algorithmically generate models of what is salient in an image [23]. As a follow-up to our eye-tracking studies, we investigate whether these techniques could be used to generate maps of visual salience for Web pages. The correspondence between these saliency maps and gaze data might give insight into how salient features are used in performing tasks on Web pages. If we can then find salient features computationally, and identify how they influence visual attention on Web pages, we have a potential mechanism to computationally interpret a page and guide how it might be 'transcoded', or re-presented in an alternative modality and order. The second part of this paper thus explores the extent to which 'saliency maps', greyscale images indicating the relative attractiveness of different parts of the page, correspond to the eye tracking data, and can therefore provide the basis for modelling how visual attention is allocated on Web pages.

[2] http://www.bbc.co.uk/education/betsie/

In Section 2 we describe the context in which this work has been performed. We then describe our experiment: its design, procedure and equipment (see Section 3). Our results are presented in Section 4.1, and the correspondence between the saliency map and the eye tracking data is discussed in 4.2. A discussion of the results is given in Section 5.

2 Related Work

Profoundly visually impaired users usually access the Web using screen readers [8] or specialized browsers [1, 26, 27, 28]. There are many sources that state the difficulty that such users have using the Web *via* tools such as screenreaders. The Disability Rights Commission (DRC) report of 2004 [6] concludes that most UK Web sites (81%) fail to satisfy even basic accessibility requirements. A similar situation exists in the USA [5]. Coyne and Nielsen [5] concluded that "the Web is about three times more difficult to use for people who are visually impaired than it is for sighted users". The question still remains, however, as to exactly what makes it so difficult.

One part of the answer is in simply extracting information encoded within the HTML of a Web page. Thatcher *et al.* [19] give practical advice on constructing accessible Web sites by elucidating guidelines [4] and using HTML to best advantage. Text-only alternatives to standard Web pages are offered as "accessible" options. However, Thatcher et al [19, Chapter 1] state "the issue of text-only versions crosses into the idea of separate versus inclusive design". Why should visually disabled people use different pages to sighted users? As well as problems of maintenance, there is a resistance to the development of Ghettos. Regan [15] observes that, "designers look at sites that are meant to serve as models of accessibility and are appalled by the aesthetics. For most designers, accessibility equates to boring, uninteresting designs. The state of accessibility on the Web today represents a failure of the imagination".

Even when pages are built according to guidelines that are meant to increase accessibility, there still seem to be "disabilities". Takagi *et al.* [18] classified the problems into three categories:

1. Adherence to guidelines, not usability issues;
2. Over-reliance on syntactic checking of Web pages;
3. No attention on the time component in the operations provided to aid access.

These observations usefully describe some of the sources of difficulties encountered by visually disabled users, without really describing the nature of those difficulties. The problem is summed up by Hanson [7], who says, "specifications for accessibility of Web pages do not necessarily guarantee a usable or satisfying Web experience for persons with disabilities. It is not uncommon to have pages that meet standards but are still difficult to use by persons who have difficulties."

Goble *et al.* [3] and Yesilada *et al.* [22] have used the metaphor of travel to raise the notion of using a Web page above that of dealing with mere 'sensory translation'. Just as people use *travel objects* in the environment (signs, landmarks and other cues) to help them orientate and navigate, so a Web user can use *travel objects* on a Web page to aid mobility—the ease of travel in a Web page. The layout and presentation of a Web page provides these travel objects. A well designed page eases travel and a badly designed page hinders travel and decreases mobility. A text-only page will tend to lack more of these travel objects and thus the reader is again hindered in their task.

Screenreaders, unlike sighted users, cannot see the implicit structural and mobility knowledge encoded within the visual presentation of Web pages [21]. "It is impossible for blind users to distinguish visually fragmented groupings only from the sequence of tags read to them" [2]. So, we begin to see that basic sensory translation of what is encoded within HTML still lacks what is necessary to support usable access to the Web. We can identify the Web correlates of what in printed material enables and supports effective information processing [20], as something lacking in current non-visual renderings of Web based material.

We have used eye-tracking in this study because we wish to find out how sighted users exploit the visual presentation of a Web page. Whilst eye tracking has been used to investigate cognitive processes for over 100 years [14], monitoring people's gaze during Web use is a relatively new discipline. Recording the pattern of fixations on a Web page is a powerful tool, enabling us to determine those areas that are most salient (attract the most fixations), and those that receive little attention.

Eye tracking's most obvious applications are in improving the standard design and layout of Web pages, and evaluating their usability [16]. Studies have also examined the saliency of items on a page under varying conditions (*e.g.*, [9, 12, 10]), how eye movements vary according to information scent [13] and how looking for a menu is influenced by page complexity and prior expectations [11]. To our knowledge, no one has investigated how gaze patterns differ when searching on either the standard or text-only versions of the same page. We address this issue by tracking participants' eye movements as they search for specific links on the standard and text-only versions of the BBC News Web site.

An interesting question concerns the extent to which the saliency of parts of a page influences how much participants look at them. Saliency maps are used to identify the most visually attractive portions of an image. Models such as that of Itti and Koch [23] have been successfully employed in computer graphics to segment images into regions the user is most likely to focus upon. These regions are then selectively rendered in higher quality than less salient portions, at a reduced overall computational cost, but without the viewer being aware of this quality difference [25]. Here, we investigate qualitatively the correspondance between saliency maps of a Web page and the gaze data obtained from eye tracking.

3 Method

Participants' eye movements were tracked while they were searching for a link on either the standard or text-only versions of the BBC News Website (see Figures 1 and 2). The time to locate the link provided a measurement of task difficulty. However, it is important to state that the goal is not to elucidate whether it is 'easier' to find a link on a standard page per se, but to understand the differences in the search techniques that the participants use on each page. Eye tracking data provided both a qualitative and quantitative means of monitoring these differences. The number and position of fixations indicated the amount of attention allocated to different areas of the page in order to find the link, and the average duration of fixations indicated the relative complexity of the information presented on each page.

3.1 Equipment

The experiment was run on a SONY VAIO VGN-FS315S laptop. Stimuli were displayed on a SONY VAIO SDM-HS75P monitor positioned on a desk. Participants sat in front of the monitor and browsed the web with a mouse. 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 analyze eye movement and event data.

3.2 Task

The study used a between-subjects design in which participants searched for links on the BBC News Manchester Web page. This site was chosen as it is widely known in the UK, but the actual content on any one day will change. BBC pages are automatically provided with a text-only page, which has identical text content to the graphical page, but information is provided in the order of the html (rather than a table) and graphics, variations in font etc. are removed. Sighted users will experience much of the effect of an audio screenreader rendering when using a text-only version. Of course, the correspondence between sighted users reading a text-only page and the reality of profoundly visually disabled people's use of an audio screenreader is only partial at best. It does, however, provide some flavour of the interaction.

Half of the participants searched for a link to the BBC Manchester Website (hereafter referred to as the 'Manchester' link) on the standard version of the page, and a link to a story about the Chinese community (the 'China' link) on the text-only version of the page; the other half searched for the Manchester link on the text-only page, and the China link on the standard page. The links were chosen because they were next to each other on the page, and positioned on the right in the standard version and approximately two-thirds of the way down in the text-only version, so participants were unlikely to see them immediately on entering the page. The presentation order of the two site versions alternated between users. At the end of the experiment, users were asked to indicate whether they used the BBC Website more than once a week, and whether they found it easier to perform the search on the standard or text-only version of the page.

3.3 Participants and Procedure

Eighteen participants between the ages of 17 and 50 with normal or corrected vision took part. The study was carried out during visit days by prospective students to the School of Computer Science. Both prospective students and their parents were asked to volunteer to take part in the study. The procedure was explained and the study performed, which took about five minutes. Participants were shown the recording of their eye movements after the experiment.

Each participant sat 50 cm from the monitor and went through a calibration process. The participant was initially presented with a page containing links to either the standard or text-only page, which allowed him or her to locate the mouse pointer on the screen visually, and was asked to look for either the BBC Manchester or Chinese community link. The participant entered the appropriate version of the page and started searching, indicating that he or she had located the link by hovering over it with the mouse. The participant then searched for the other link on the alternative version of the page.

3.4 Saliency maps

We used an implementation based on the Itti and Koch saliency model to generate the maps. In order to find the visually important areas, this algorithm combines information from three channels. First, an edge map is used to determine perceptually important edge features, then a metric for intensity and colour contrast is combined with this edge map to obtain the salient features [24]. The algorithm was originally designed for use with computer generated images. Here, we test its efficacy on Web pages.

Our method takes as input image-space information normally captured as a screen-shot in TIFF format. This is processed and the saliency map is supplied as a greyscale TIFF image. The algorithm for calculating the saliency of an area is described in more detail in [24] and [25]. The lighter an area, the more salient the algorithm predicts it to be. The technique was designed for use with computer graphics, so it is not clear how successfully it will work with Web pages. We hypothesize there will be a broad correspondence between the areas on the standard Web page predicted to be the most salient, and the areas on the Web page that receive the most fixations. However, the match is unlikely to be exact, as the saliency map fails to take into account any top-down knowledge (relating to the task and/or prior experience with the BBC Website) that may influence eye movements.

4 Results

4.1 Eye tracking data

The eye tracker recorded the position of the participant's gaze on the Web page throughout the experiment[3]. We used the ClearView analysis software to calculate the number of fixations that occurred, their position, their order, and their average duration. Figures 3 and 4 show the areas of the page that received the most fixations in the standard page when participants were looking for the 'Manchester' and 'China' links respectively (red = 9 or more fixations, green = 4 or more, grey = 0). Participants tended to fixate on the salient areas when searching the standard page: headlines, some images, and prominent words in the text. The areas with the most fixations also appeared to vary according to the link participants were searching for – when looking for the Manchester link, they looked more at the menus, as this may seem a likely place from which to navigate to a BBC regional Website.

[3] The full size webpage stimuli, gaze plots and heat maps are available at

http://aig.cs.man.ac.uk/research/attention/attention.php.

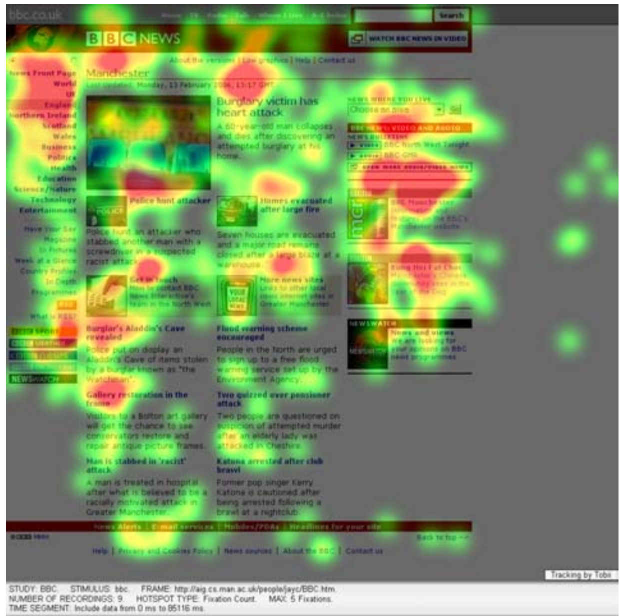

Figure 3: Gaze 'hotspots' when searching for the 'Manchester' link on the standard page.

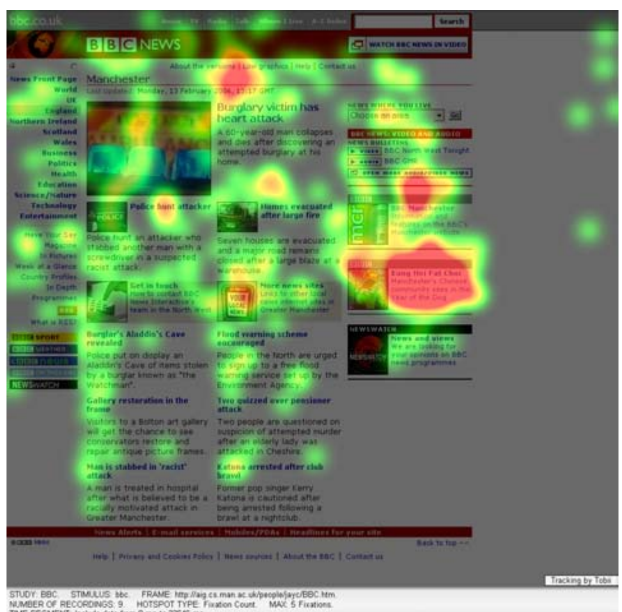

Figure 4: Gaze 'hotspots' when searching for the 'China' link on the standard page.

On the text-only page, participants fixated in a uniform manner on nearly all of the text on the left-hand side of the page, indicating that it was allocated equal importance as they read down from the top until they found the link (see Figures 5 and 6). Due to the length of the page, participants had to scroll down several times, but this pattern continues on every screen they encounter.

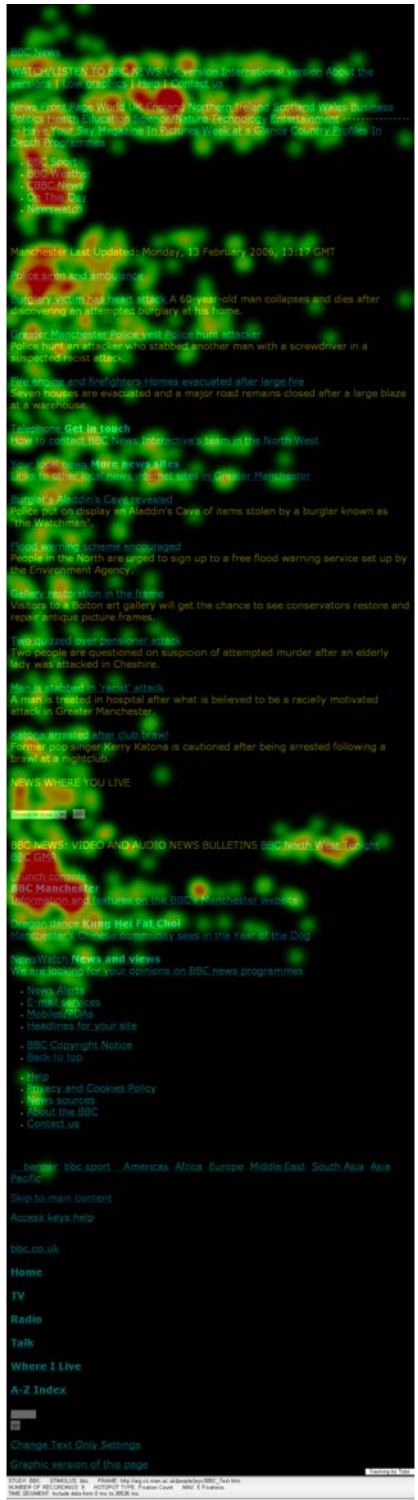

Figure 5: Gaze 'hotspots' when searching for the 'Manchester' link on the text-only page.

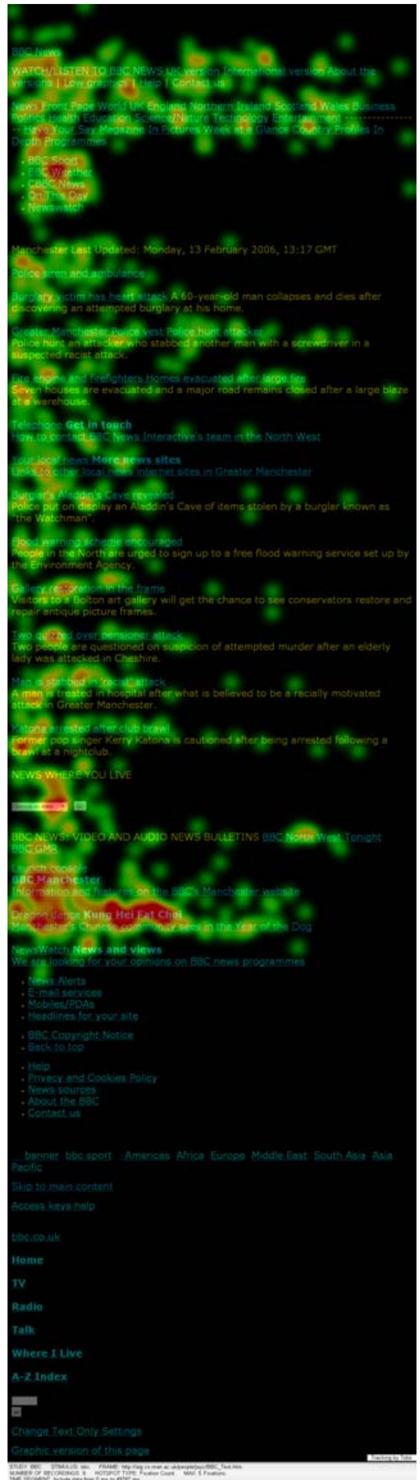

Figure 6: Gaze ‘hotspots’ when searching for the ‘China’ link on the text-only page.

Qualitative analysis of the gaze replay and fixation order data confirms that participants simply read down the page in a serial fashion when looking for the link on the text-only page. On the standard page, however, participants' eyes dart around as they attempt to locate what appears visually to be the next most likely location for the link. The gaze plots in Figures 7 and 8 illustrate this process for participants searching for each link on the standard page and those in Figures 9 and 10 show the same two participants searching for the other link on the text-only page. When looking at the standard page, participants make large saccades to move from one salient area to the next; on the text page however, saccades are much smaller, as the participants read through the text from top to bottom.

Figure 7: Gaze plot for a participant searching for the 'Manchester' link on the standard page.

Figure 8: Gaze plot for a participant searching for the 'China' link on the standard page.

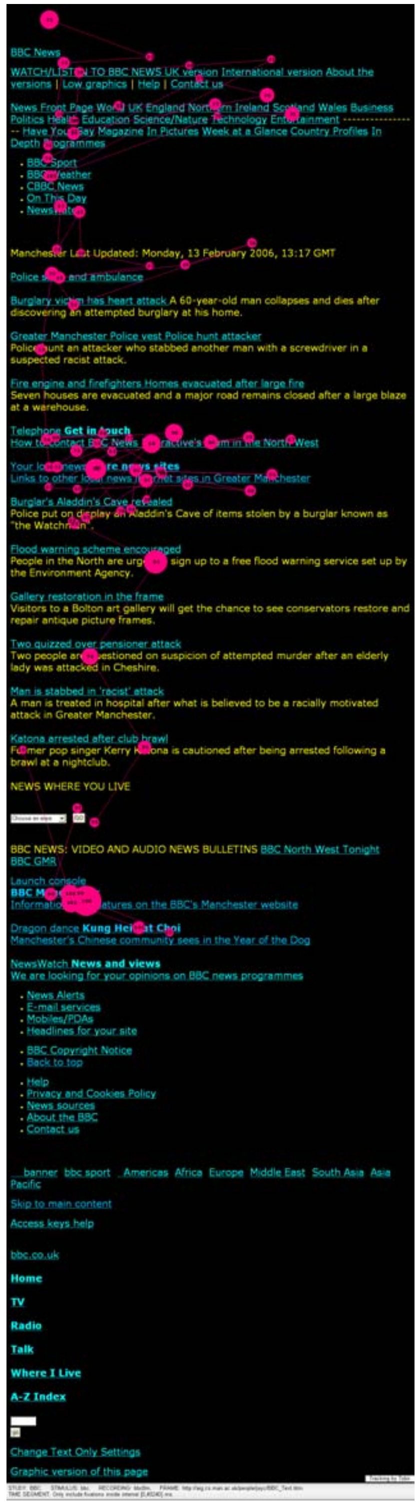

Figure 9: Gaze plot for a participant searching for the ‘Manchester’ link on the text-only page.

Figure 10 Gaze plot for a participant searching for the 'China' link on the text-only page.

A two-way ANOVA (page × link) showed that participants made significantly more fixations on the text-only site than they did for the standard site, regardless of the link they were searching for ($F_{1,32} = 2.183$, $p<0.005$; see Figure 11). The fixations also appeared to be spread over a greater proportion of the text-only page, in a more uniform manner. The duration of a fixation, however, lasted significantly longer on the standard page ($F_{1,32} = 0.208$, $p<0.005$), indicating that a greater amount of cognitive processing occurred in a fixation on the standard page (see Figure 12).

It may be that the information obtained in a single fixation on the standard page was more complex, or that some of the extra processing time was due to participants orientating themselves and planning their navigation to the next part of the page (fixations during visual search are known to be longer than fixations during silent reading [14]).

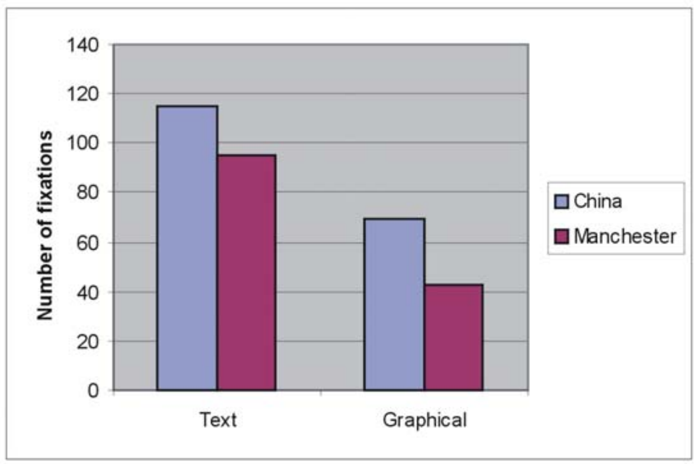

Figure 11: Mean number of fixations.

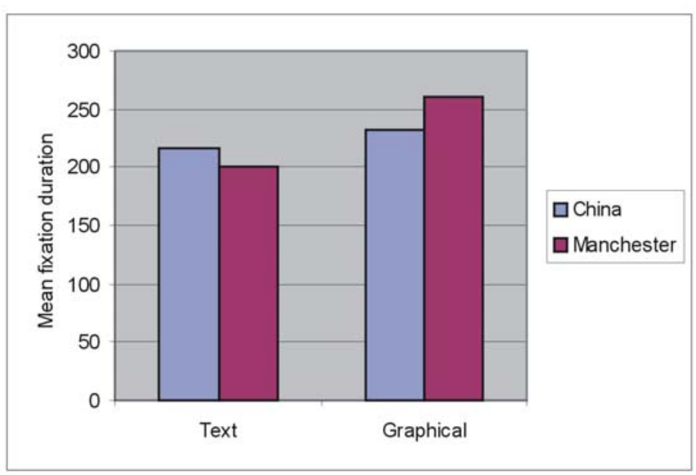

Figure 12: Mean fixation duration in msec.

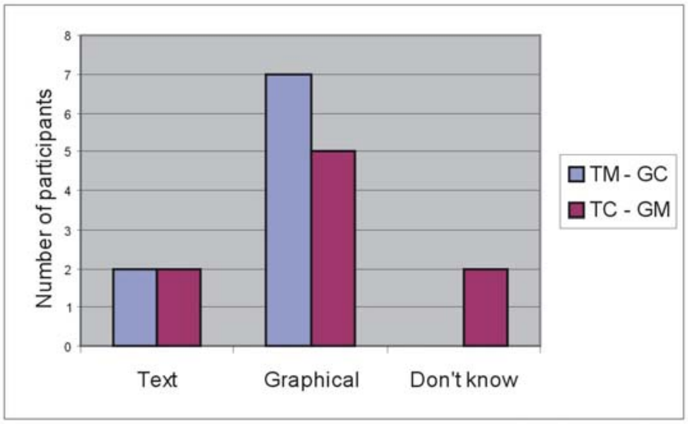

Figure 13: Number of participants who found it easier to locate each link on the standard and text-only pages.

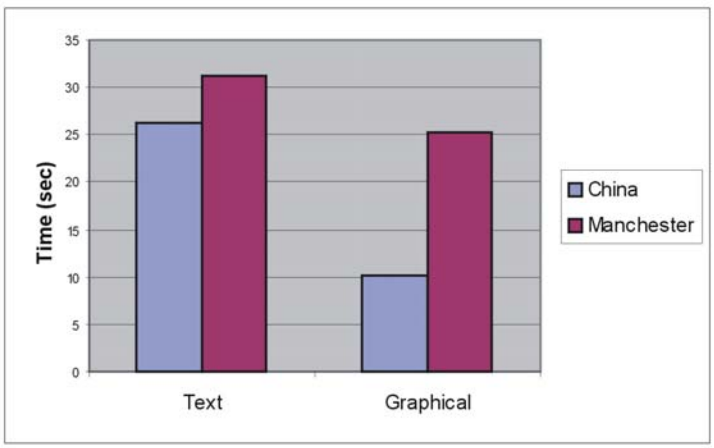

Figure 14: Mean time to locate a link in msec.

Significantly more participants, $\chi^2(2, N = 18) = 9.33$, $p < 0.01$, found it easier to search for the link on the standard version of the page (see Figure 13). It is important to acknowledge that familiarity with the page may have had a strong influence – all but two of the participants used the BBC Website more than once a week, and of those who did not, one rated the text-only page as easier and the other did not express a preference. However, the majority of participants found it easier to search for the information in the standard page.

Completion times (shown in Figure 14) did not vary significantly as a result of page type. Closer analysis shows that this result occurred due to the serious difficulty two participants had locating the Manchester link (taking more than 50 seconds to find it, in contrast to the 13 seconds it took the other participants). This may have arisen due to the prior expectations of the participants. Both spent a long time looking at menus, convinced that the link should be located on one, rather than positioned on the right of the page. A t-test considering the China link searches alone shows that participants locate it significantly more quickly on the standard page ($t_{16} = 3.696$, $p<0.005$). Although this difference can be ascribed to layout, it is also worth considering the fact that the two pages varied in colour. Sighted users, used to seeing white text on a black background, may have had more difficulty processing yellow text on a black background.

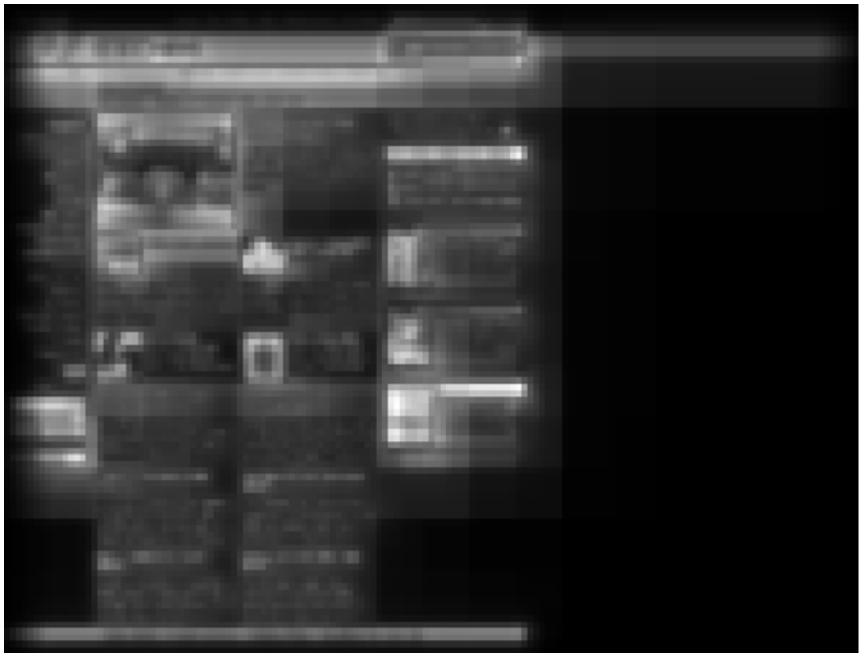

Figure 15: Saliency map for the standard page.

.

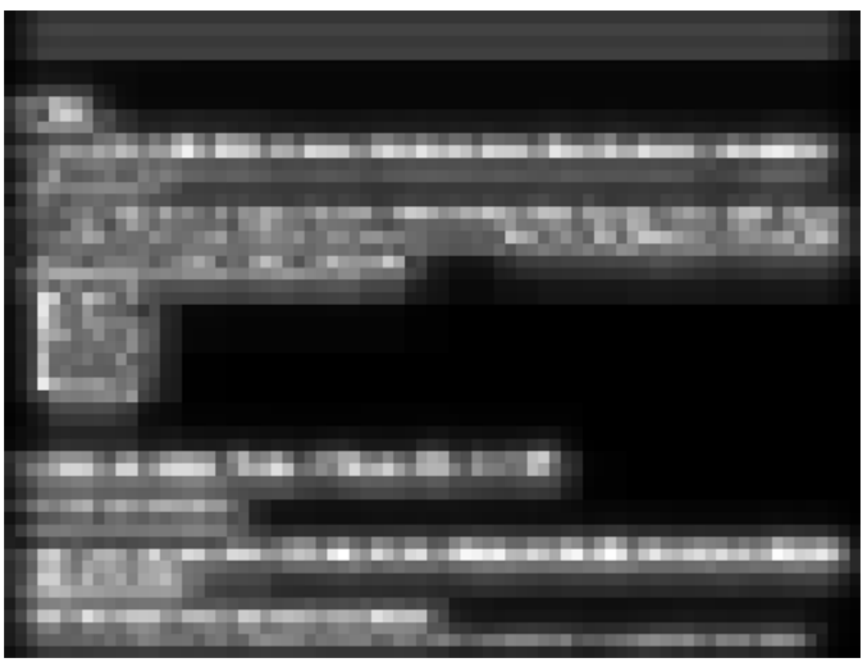

Figure 16: Saliency map for the text-only page

4.2 Saliency Map Results

Figure 15 shows a saliency map for the standard version of the page and Figure 16 shows a saliency map for the top screen of the text-only page. All of the content appears highly salient: although underlined or bulleted text is slightly lighter, it is difficult to differentiate between different sections of content. It is much easier to see how content is divided up on the standard page, as the space between different sections acts as a border between them, and is picked up by the saliency algorithm. There is also far more variation in the amount of saliency attributed to different parts of the page. Images and headings, for example, are shown as much more salient than body text.

Figure 17 shows the map for the standard page superimposed on the original stimulus with the cumulative gaze data. The gaze data, to a large extent, falls into sections as identified by the saliency algorithm. There is also a correspondence between gaze hotspots and some bold text, headings and images rated as highly salient. However, the match is far from perfect. The menu on the left hand side, for example, is not rated as particularly salient by the algorithm, but receives a large number of fixations. In contrast, images, which are highly salient, receive fewer fixations than the text next to them. Tuning the algorithm to take knowledge about the task and Website into account will improve the correspondence between the saliency map and gaze data. In the current situation, a specific example of task-related knowledge is that users are more likely to fixate on menus (albeit incorrectly in this task) when searching for an item they consider to be a category heading, like the BBC Manchester Website. A more general example of an adjustment to the algorithm concerns the relative salience of text and images. Illustrated text tends to receive more fixations than the 'salient' image next to it, but it also tends to receive more fixations than text that is not illustrated.

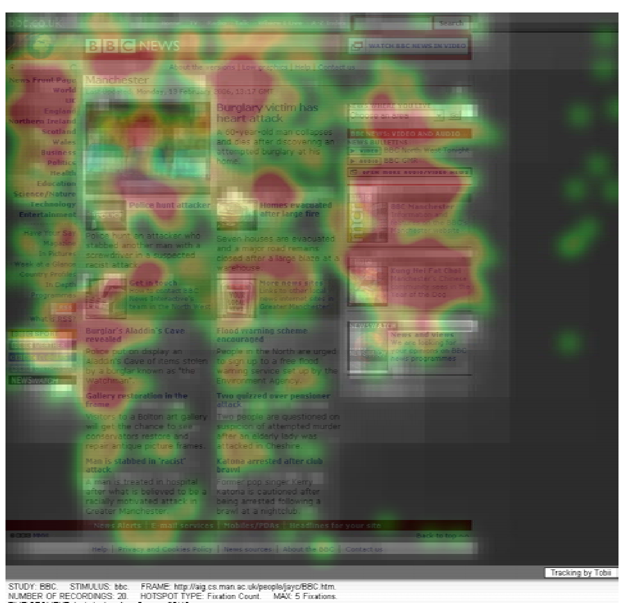

Figure 17: Saliency map for the standard page superimposed on the original stimulus with cumulative gaze hotspots.

5 Discussion

An Audio rendering (and test-only versions of Web pages) lacks many of the very facilities that make the visual system the predominant human information gathering system, that is, the ability to control information flow from the outside with speed and accuracy, because there is an external memory holding the information [17]. In the case of Web pages, there is the additional factor of even being able to access what is being presented—the focus of much of the work reviewed in Section 2. Current audio renderings essentially lack the formatting and presentation available in the visual mode: layout, spacing; typefaces; font-size; colour; *etc*. Non-speech and audio cues can potentially add some of this information, but this has not so far been greatly exploited in the rendering of Web pages. A key question is what features need to be added via other modalities. Once we know this, we can tackle the question of how they should be added.

In this exploratory experiment we have attempted to further characterize the problem of Web accessibility, by asking how formatting features are used during a particular activity. Although only one of the two search tasks was completed significantly faster on the standard page, we qualitatively show a stark difference in eye-movement behaviour on standard and text-only Web pages. In the standard, formatted presentation (preferred by the majority of participants), saccades are greater and fixations longer. The link menu is being used, along with the distinct presentational fragmentation of the page into areas. In contrast, the text-only page provoked eye-movements more akin to those seen in the reading of ordinary printed text. The page is being "read", rather than used to navigate towards the link goal. The observation is that a sighted user uses the formatting of the page to achieve the task and that the longer gaze duration is a consequence of decision making on orientation and navigation. The text-only page is in effect similar to that which a blind user encounters in an audio presentation. The eye-movements seen in this condition are similar to those seen in cursor movements during reading. The opportunity to "dart" around the page to orientate and navigate are removed. Text-only versions of a Web page have been advocated as accessible options for visually disabled users. Whilst they make the content available, the observations of this investigatory study suggest they lack the facilities that a formatted presentation affords a sighted Web user. Further studies, using different tasks, will help to determine more precisely the nature of how formatting aids search (enables a user to locate an item more quickly), and illuminate what exactly it is about the page that helps the user.

The ultimate goal of this research is to provide information with which to restructure or remodel web pages to improve their non-visual presentation. The question of how to replace the facilities provided by layout in non-visual presentations is open to debate, and it is our intention that it should remain so. However, we believe that one obvious use of the proposed models is in driving transcoding. Layout provides borders for portions of information. Ideally, the objects on a page guide a reader through the information such that he or she can accomplish a task [20]. This is the basis of the use of the travel metaphor [3] in increasing Web accessibility. Here, the *travel objects* that facilitate movement through a Web page are identified and semantically marked up [22]. This markup is then used to help transcode a page into either an order or a fragmentation that facilitates use.

Obviously, mark-up and transcoding cannot occur by hand alone. The correspondence between gaze data and saliency maps seen here shows that such maps could provide the basis for a model used to automate this process. We can see that even a saliency map originally designed for computer generated images gives a reasonable correspondence to the features on a Web page, providing a starting point for computationally detecting those areas of a page that will draw a reader's attention, and may thus constitute important travel objects.

The technique undoubtedly needs refinement. For example, whilst images are rated as highly salient, in reality they receive fewer fixations than the text or heading next to them – a result that would have to be incorporated in the algorithm. We can also see from the study that task requirements, heuristics and expectations have a strong influence on gaze patterns. Incorporating these factors into a model would be challenging, but could be achieved through further eye tracking and behaviour studies. Conducting studies examining how people's use of salient information varies according to such issues constitutes the next step in this work.

It is vital to point out that this approach is not trying to mimic what sighted people do when reading a page. We wish to reproduce instead the support such readers gain from a page's presentation. The vast majority of Web pages are, and will continue to be, designed for the sighted world. Our approach is to exploit an understanding of how that world works and transfer what it enables to alternative modalities. Our aim is to see whether we can create models of how presentation is used for navigation and orientation about a page such that we can look at a page's formatting and layout and infer how it will be used. Such models should be able to drive the transcoding of a page and begin to re-introduce into an audio rendering the same support for tasks that presentational cues provide to sighted users.

Acknowledgements: Caroline Jay is funded by EPSRC grant **EP/D036518/1**.

[1] C. Asakawa and T. Itoh. User interface of a home page reader. In *ASSETS'98*, pages 149–156. ACM Press, 1998.

[2] C. Asakawa and H. Takagi. Annotation-based transcoding for nonvisual Web access. In *ASSETS'00*, pages 172–179. ACM Press, 2000.

[3] C. A. Goble, S. Harper, and R.D. Stevens. The Travails of Visually Impaired Web Travellers. In *ACM 2000 Hypertext (Proceedings of the Eleventh ACM Conference on Hypertext and Hypermedia)*, pages 1–10. ACM Press, 2000.

[4] W. Chisholm, G. Vanderheiden, and I. Jacobs. Web content accessibility guidelines 1.0. World Wide Web Consortium (W3C), 1999.

[5] K. P. Coyne and J. Nielsen. Beyond ALT text making the Web easy to use for users with disabilities. Nielson Norman Group, 2001.

[6] Disability Rights Commission. The Web access and inclusion for disabled people. Technical report, Disability Rights Commission (DRC), UK, 2004.

[7] V.L. Hanson. The user experience designs and adaptations. In S. Harper, Y. Yesilada, and C. Goble, editors, *Proceedings of the international cross-disciplinary workshop on Web Accessibility*, pages 1–11, 2004.

[8] Henter-Joyce, Inc. *Jaws*.

[9] T. Joachims L. Granka and G. Gay. Eye-tracking analysis of user behaviour in www search. In *SIGR'04*, pages 478–479, Sheffield, UK, 2004. ACM Press.

[10] A. Tholke M. Schiessl, S. Duda and R. Fischer. Eye tracking and its application in usability and media research. In *MMI-Interactiv '03*, pages 41–50, 2003.

[11] J. McCarthy, A. Sasse, and J. Riegelsberger. Could I have a menu please? an eye tracking study of design conventions. In *Proceedings of HCI2003*, pages 401–414, 2003.

[12] B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, and J. K. Newman. The determinants of Web page viewing behavior: An eye-tracking study. In *Eye Tracking Research & Applications Symposium on Eye-Tracking Research & Applications*, pages 147–154, San Antonio, Texas, 2004. ACM Press.

[13] P. L. Pirolli, S. K. Card, and M. Van Der Wege. Visual information foraging in a focus+context visualization. In *Proceedings of ACM Conference on Human Factors and Computing Systems (CHI 2001)*, pages 506–513, Seattle, 2001. ACM Press.

[14] K. Rayner. Eye movements in reading and information processing: 20 years of research. In *Psychological Bulletin*, pages 372–422, 1998.

[15] B. Regan. Accessibility and design a failure of the imagination. In S. Harper, Y. Yesilada, and C. Goble, editors, *Proceedings of the international cross-disciplinary workshop on Web accessibility*, pages 29–37, 2004.

[16] M. Russell. Using eye-tracking data to understand first impressions of a Website. In *Usability News 7.2*, 2005.

[17] R. D. Stevens, A. D. N. Edwards, and P. A. Harling. Access to mathematics for visually disabled students through multi-modal interaction. *Human-Computer Interaction (Special issue on Multimodal Interfaces)*, pages 47–92, 1997.

[18] H. Takagi, C. Asakawa, K. Fukuda, and J. Maeda. Accessibility designer visualizing usability for the blind. *ASSETS'04*, pages 177–184, 2004.

[19] J. Thatcher, P. Bohman, M. Burks, S.L. Henry, B. Regan, S. Swierenga, M.D. Urban, and C.D. Waddell. *Constructing Accessible Web Sites*. Glasshaus, 2002.

[20] P. Wright. The psychology of layout: Consequences of the visual structure of documents. In R. Power and D. Scott, editors, *American Association for Artificial Intelligence Technical Report FS-99-04*, pages 1–9. AAAI Press, 1999.

[21] Y. Yesilada, S. Harper, C. Goble, and R. Stevens. Screen readers cannot see (ontology based semantic annotation for visually impaired Web travellers). In N. Koch, P. Fraternali, and M. Wirsing, editors, *Web Engineering - 4th International Conference, ICWE 2004 Proceedings (LNCS 3140)*, pages 445–458. Springer, 2004.

[22] Yeliz Yesilada, Robert Stevens, and Carole Goble. A foundation for tool based mobility support for visually impaired Web users. In *Proceedings of the Twelfth International World Wide Web Conference*, 2003.

[23] L. Itti, C. Koch and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. Pattern Analysis and Machine Intelligence (20), pages 1254-1259, 1998.

[24] P. Longhurst, K. Debattista and A. Chalmers. Snapshot: A rapid technique for driving a selective global illumination renderer. *Winter Conference on Computer Graphics (WSCG) 2005*, 2005.

[25] V. Sundstedt, K. Debattista, P. Longhurst, A. Chalmers and T. Troscianko. Visual attention for efficient high-fidelity graphics. Spring Conference on Computer Graphics SCCG 2005, pages = 162--168, 2005.

[26] T. V. Raman. Emacspeak - direct speech access. In *ASSETS'96*, pages 32–36. ACM Press, 1996.

[27] M. Zajicek, C. Powell, and C. Reeves. Evaluation of a World Wide Web scanning interface for blind and visually impaired users. In *HCI '99*, volume 2, pages 980–984, 1999.

[28] pwWebSpeak. http://www.soundlinks.com/pwgen.htm.

[29] E. Pontelli, D. Gillan, W. Xiong, E. Saad, G. Gupta and A. I. Karshmer. Navigation of HTML Tables, Frames, and XML Fragments. In *ASSETS'02*, pages 25-32. ACM Press, 2002.

[30] Y. Yesilada, R. Stevens, C. Goble and S. Hussein. Rendering Tables in Audio: The Interaction of Structure and Reading Styles. In *ASSETs'04*. Pages 16-23. ACM Press, 2004.

[31] A. Brown, S. Pettifer and R. Stevens. Evaluation of a Non-Visual Molecule Browser. In *ASSETS'04*. Pages 40-47. ACM Press, 2004.

[32] Z. B. Fredj and D. A. Duce. GraSSML: accessible smart schematic diagrams for all. *W4A: Proceedings of the 2006 International Cross- Disciplinary Workshop on Web Accessibility*. Pages 57-60, 2006.