

Wai Yu · Ravi Kuber · Emma Murphy · Philip Strain  
Graham McAllister

## A novel multimodal interface for improving visually impaired people's web accessibility

Received: 29 July 2005 / Accepted: 7 October 2005 / Published online: 13 December 2005  
© Springer-Verlag London Limited 2005

**Abstract** This paper introduces a novel interface designed to help blind and visually impaired people to explore and navigate on the Web. In contrast to traditionally used assistive tools, such as screen readers and magnifiers, the new interface employs a combination of both audio and haptic features to provide spatial and navigational information to users. The haptic features are presented via a low-cost force feedback mouse allowing blind people to interact with the Web, in a similar fashion to their sighted counterparts. The audio provides navigational and textual information through the use of non-speech sounds and synthesised speech. Interacting with the multimodal interface offers a novel experience to target users, especially to those with total blindness. A series of experiments have been conducted to ascertain the usability of the interface and compare its performance to that of a traditional screen reader. Results have shown the advantages that the new multimodal interface offers blind and visually impaired people. This includes the enhanced perception of the spatial layout of Web pages, and navigation towards elements on a page. Certain issues regarding the design of the haptic and audio features raised in the evaluation are discussed and presented in terms of recommendations for future work.

**Keywords** Multimodal interface · Haptics · Audio · Assistive technology · Web accessibility · Web navigation

### 1 Introduction

Limitations of current assistive technology and inaccessible Web design prevent blind and visually

impaired users from experiencing the full potential of the Internet, in comparison to their sighted counterparts. A recent survey conducted by Petrie et al. for the Disability Rights Commission reports that only 19% of 1,000 tested UK Websites' homepage pass the priority 1 check specified in the W3C's Web Content Accessibility Guidelines [1]. Barriers to accessibility on the Web, can be attributed to the predominantly visual nature of information presented to users via computer interfaces. The situation is further compounded by the limitations posed by assistive devices, such as screen readers and Braille displays. These assistive tools force blind users to browse the Web in a linear and time-consuming fashion, rendering graphs, animations and busy Web pages inaccessible. As key structural information is omitted, it is difficult to gain a full comprehension of the material presented. Developing an awareness of the spatial layout of objects on a Web page can also present a challenge. Thus, a need has been identified for a new approach to Web browsing for visually impaired users.

Traditional non-visual assistive tools for browsing have been designed with the aid of the auditory channel. WebSpeak [2] and BrookesTalk [3] have been developed to output text-to-speech, providing an aural overview of Web content for the user. A study by Donker et al. [4] has examined the development of auditory interaction realms to represent the layout of Web pages and to support navigation. It has been demonstrated that by embedding sounds in an environment, locational awareness of objects can be improved [4, 5]. Roth et al. [6] have investigated adding the haptic modality to an auditory environment. In this study, sounds represent the nature of the HTML tag touched, providing increased awareness of the position and meaning of the element. Audio and haptic techniques traditionally associated with expensive virtual reality technologies have recently become more feasible for the design of desktop solutions. In this study, virtual web objects are created through audio and haptic feedback to create a realistic non-visual spatial representation of a web page.

W. Yu (✉) · R. Kuber · E. Murphy · P. Strain · G. McAllister  
Queen's University of Belfast, University Road,  
BT7 1NN Belfast, Northern Ireland, UK  
E-mail: w.yu@qub.ac.uk  
Tel.: +44-28-90974781  
Fax: +44-28-90974332

The haptic modality has been exploited in order to improve access to interfaces. Force feedback devices have been developed in response to the lack of non-visual feedback on a graphical user interface (GUI) allowing icons, controls and menus on the screen to be tactually perceived [7, 8]. Previous studies have illustrated that advantage can be gained when the haptic modality is used in conjunction with the visual and auditory channels [9–11]. The IFeelPixel multimodal application [12] enables the user to mediate structures such as edges, lines and textures, depending on features of the pixels detected by the device. Both tactile and auditory feedback is experienced as a result. Multimodal solutions have the capacity to extend visual displays making objects more realistic, useful and engaging [9].

Multimodal interfaces also provide assistance in the mental mapping process, allowing the user to develop a greater awareness of objects contained within the environment. A clearer spatial representation created by multimodal feedback, enhances the user's ability to orientate and navigate within their environment [11, 13, 14]. As the majority of information required for mental mapping of an unknown space is gathered through the visual channel, it seems apparent that a multimodal assistive interface may provide a solution to reducing barriers that are currently faced by the visually impaired community.

Non-visual browsing methods have previously been examined in order to gain a complete picture of how multimodal feedback can be used to support the user in their browsing tasks. Jansson and Monaci [15] have found that by providing differentiated information within the contact area available with the haptic display, benefit will be derived as objects can be recognised more effectively. Similarly, by providing distinguishable icons in the auditory realm, benefit will also be brought to a non-visual environment. Auditory icons positioned strategically in the environment can contribute to the formation of a mental model, which would aid browsing [4, 5].

Findings from a user requirement survey conducted at Queen's University of Belfast with 30 blind and partially sighted people have revealed that the positions of images are particularly difficult to detect on a page due to the lack of feedback given. Images provide useful context to the corresponding text contained within a page. Alternative text descriptions are helpful, but the intention that the original image is trying to convey may not be immediately obvious to the user, thus rendering some pages difficult to interpret. The position of hyperlinks on a Web page also presents a challenge to locate. Links themselves may not provide meaningful cues to the user. The URL that the link would follow may not offer a description of the intended target. By removing structural and contextual information concerning objects on a Web page, additional time and attention must be spent on a page as the user tries to derive the meaning lost through the use of assistive technologies.

Our research aims to extend previous work by focusing specifically on improving accessibility for visually impaired users when interacting with Web pages, examining the presentation of information content, access to graphics, and navigation. It is hoped that by using the multimodal interface, additional structure can be brought to a page to give it more meaning, thus adding value to the perceptual experience described by Sharmin et al. [16].

---

## 2 Multimodal interface

A multimodal interface is currently being developed to improve blind and visually impaired people's Web accessibility. System design focuses on three main areas: (1) navigation on the Web, (2) representation of information, and (3) access to graphical content. To achieve the objectives, Web technology combined with haptic and audio representations are used to form the multimodal interface. The system structure is shown in Fig. 1.

In the first system prototype, a content-aware Web browser plug-in has been developed to assist Web navigation through the use of haptic and audio features. In this approach, users have an opportunity to explore a Web page's layout through active haptic interaction. A force feedback mouse is used and its cursor position is constantly monitored by the plug-in, which detects the surrounding objects. If the plug-in finds an object nearby then it will inform the user by enabling the haptic and audio features. Depending on the user's intention and the context of the task, appropriate prompts can be given, such as providing users with guidance to the desired destination or informing users about nearby objects. The content-aware Web browser plug-in, and the associated haptic and audio features are described in the following sections.

---

## 3 Browser plug-in

The plug-in is the crucial component of the multimodal interface because it monitors the cursor movements and activates the haptic and audio features accordingly. The two main browsers currently available for the development of plug-ins are Microsoft Internet Explorer 6.0 (IE) [17] and Mozilla Firefox 1.0 [18]. Both browsers offer distinct advantages and disadvantages. As Firefox is based on the open-source Mozilla project, extensions can be readily developed, as a result of its cross platform compatibility and accessibility of source code. Mozilla also fully implements the W3C standards [19].

### 3.1 Overview of extension architecture

Mozilla extensions use a range of programming languages and interfaces. Javascript is the primary scripting language, and is used in conjunction with cascading style

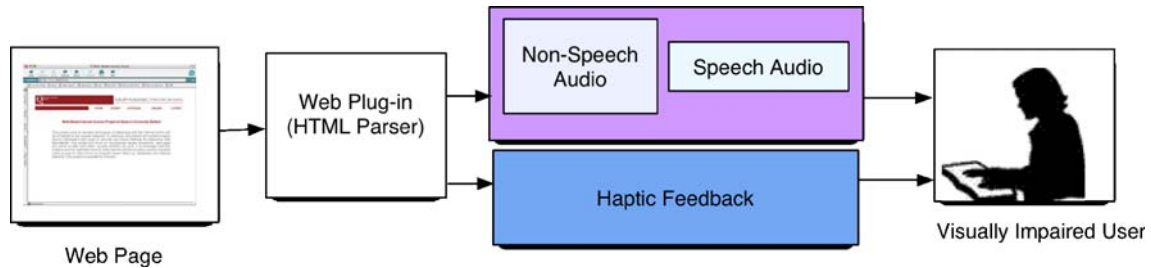


Fig. 1 Overview of the multimodal approach at QUB

sheets (CSS) and the Document Object Model (DOM) [20] to access and manipulate HTML elements in real time.

The Javascript API can be extended by writing a Cross Platform Object Module (XPCOM). XPCOM is a framework, which allows large software projects to be broken up into smaller, manageable components. To achieve this, XPCOM separates the implementation of a component from the interface, which is specified via the interface definition language (IDL). XPCOM is similar to Microsoft COM in structure, however it is designed to be used mainly at the application level [21]. XPConnect provides a bridge between the component and the Javascript external, allowing constructors and methods from the component to be accessed via Javascript. An overview of the architecture is shown in Fig. 2. This architecture enables the rapid development and prototyping of extensions for the Firefox browser.

There were four main requirements for the development of the plug-in:

1. The current position of the mouse cursor is captured by adding a mousemove listener to the browser window. Javascript is used to record values for movements made using the mouse.
2. The position of each HTML element on the screen is obtained by parsing the DOM via Javascript, and by

retrieving the co-ordinates via the stylesheet properties for each element.

3. The relative co-ordinates of the mouse pointer are calculated, if the mouse cursor is within a distance, DIST, of an HTML element. The element is divided into nine sections (Fig. 3), where each section has a particular co-ordinate range. The height and the width of the images and hyperlinks are taken into consideration for the calculation. Dimensions of images are calculated using the element's stylesheet properties. In the case of hyperlinks, the heights and widths are determined by the number of characters contained within the hyperlink, and the size of font used.
4. Finally, relative co-ordinates are passed to an external application; the real time audio simulation environment and a haptic device. An XPCOM component has been created in C++, to provide methods for sending control messages from the browser to the audio simulation environment via UDP, which is then parsed by the audio simulation application. A separate plug-in, provided by the haptic device manufacturer, is used to interface the browser with the device. Haptic effects are then accessed by the extension software discussed in this paper, allowing users to mediate elements on a Web page with force feedback.

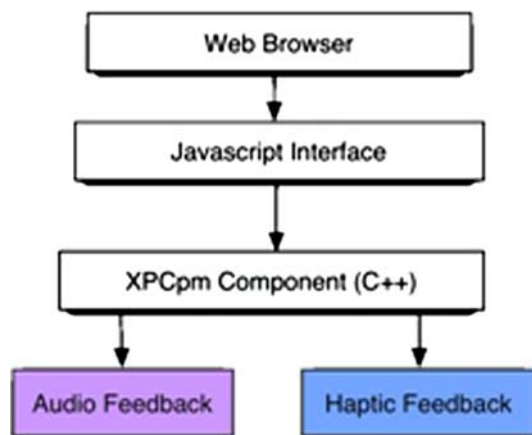


Fig. 2 Web browser plug-in architecture

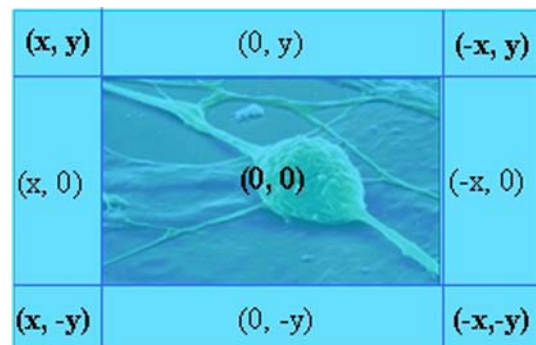


Fig. 3 Object co-ordinates

### 3.2 Haptics

The Logitech Wingman force feedback mouse (Fig. 4) has been selected to facilitate on-screen navigation, due to its capability to access haptic feedback and its compatibility with the Firefox browser. Moreover, the device is in the form of a computer mouse, which is a common tool used by sighted people for their day-to-day GUI-based activities. The immersion Web plug-in has been linked to the content-aware Web plug-in. Supporting software can model a small array of haptic effects including stiffness, damping, and various textures, which can then be called through the Javascript. This has facilitated the exploration of objects with additional force feedback.

The main objective of the haptic feedback here is to inform users about the presence and position of images and hyperlinks on a Web page. Haptic cues in the multimodal interface were developed, adhering to recommendations from [10, 22, 23]. General principles of developing distinctive sensations to aid object identification and providing constraints to facilitate navigation were taken into account for the 2D nature of a Web page. Appropriate design and mapping of haptic cues to suitable objects on a Web page would lead users to develop a clearer mental representation of spatial layout.

The following haptic primitives have been employed to develop a “roll-over” metaphor; the enclosure effect has been coupled with clipping effects bordering an image. This has given the illusion of a ridge, which needs to be mounted. Cursor clipping motion increases a user’s psychological perception of the wall’s stiffness. Upon rolling over the image, a buzz effect has been produced along with force feedback. The dual effect of audio coupled with force feedback, is intended to heighten the sense of positional awareness.

The periodic effect has been used to provide location awareness of the cursor when directly hovering over a hyperlink. This effect produces a wave that varies over time, promoting a locked sensation when directly hovering over the link. It is intended that this will promote a sense of orientation within a page for the visually impaired user.



Fig. 4 Logitech wingman force feedback mouse

### 3.3 Real-time audio

Audio feedback for the system consists of both speech and non-speech audio. Non-speech sounds complement haptic feedback to convey navigational information. Speech output conveys textual information through a text-to-speech synthesiser.

#### 3.3.1 Non-speech audio

The non-speech audio feedback for this system gives the user a sense of navigation in relation to an image or a link on the page. The audio has been designed in Max/MSP, a real-time audio programming environment. Audio is then played back using the same software. Netsend, an MSP external object is used to receive  $x$  and  $y$  location co-ordinates sent via UDP from the Web plug-in. Figure 5 show how the element is divided up, and the range of co-ordinates that are associated with each section.

As the user rolls over an image or a link with the force feedback mouse, an auditory icon is played to reinforce the haptic response. In this system, the sound icon that indicates an image is a short descriptive auditory clip of a camera shutter clicking, suggesting a photograph or graphic. The auditory icon used to depict a link is a short “metallic clinking” sound suggesting the sound of one link in a chain hitting off another.

Outside the image or link space the cursor location is mapped to panning and pitch-shift parameters of a continuous background sound. The  $x$ -value co-ordinates are mapped to a panning patch in Max/MSP so that as the user moves the cursor along the  $x$ -axis the audio is panned to that position. Similarly the pitch varies according to the position on the  $y$ -axis; as the user

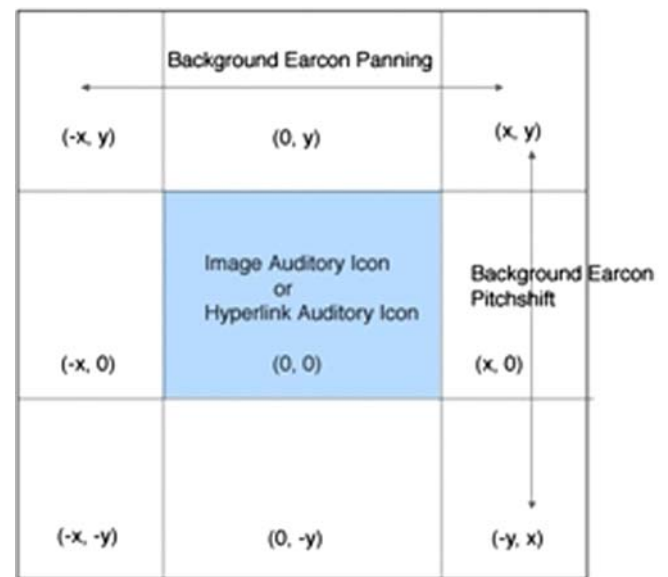


Fig. 5 Object co-ordinates and audio feedback

moves the cursor upwards, the background sound is pitch-shifted upwards to represent this movement.

### 3.3.2 Speech audio

The Microsoft Speech SDK [24] is utilised to provide speech synthesis via the Web plug-in. As the user rolls over non-link text on a page, the text is read out to the user by paragraph. The speech will stop when the user moves off the text on to another object. As the user rolls over an image, the corresponding alt text describing the significance of the image is read to the user while the auditory icon simultaneously informs the user that the object is an image. Similarly as the user rolls over a link, the speech synthesiser reads the text while the link auditory icon plays.

## 3.4 Evaluation

In order to assess the usability of the multimodal interface by visually impaired people, a series of experiments have been conducted. The experiments were divided into two parts: (1) assessment with sighted people and (2) assessment with blind and visually impaired people.

### 3.5 Assessment with sighted people

An experiment was designed to investigate the overall usability of the multimodal Internet browser, verifying strengths, and weaknesses of the system, over a commercial screen reader commonly accessed by visually impaired Internet users. The experiment intended to examine three main aspects: (1) spatial awareness of object layout on a Web page, (2) navigation towards these objects on a page, and (3) system usability. Therefore, the experiment was divided into three sections. The study was conducted on a group of fully sighted participants, blind-folded for all tasks. We did not use blind or visually impaired people in this section of experiment due to the difficulties in applying a controlled evaluation paradigm as a result of the variability between visually impaired users [25] and additional difficulties obtaining a large sample group of representative users. Other researchers' work indicates that there appears to be no significant difference between blind and sighted people's performance for tasks such as locating items when using novel haptic interfaces [8, 26], however it is acknowledged that this may not be the case for all scenarios.

Twelve participants from Queen's University Belfast, aged between 22 and 41 were recruited for the purposes of the experiment. Participants came from a wide range of academic backgrounds, comprising of music, engineering, and life sciences. They had no prior experience of screen readers, the multimodal interface developed

for blind and visually impaired users, or a force feedback mouse. Over half the participants had minor levels of sight loss, corrected through the use of glasses. They mentioned no other auditory impairments or issues with movement or dexterity that would have hindered use of the force feedback mouse. For the purpose of the experiment, the participants were blindfolded to assimilate conditions of being visually impaired.

#### 3.5.1 Section 1: Spatial awareness of object layout on a Web page

The main objective of the section of experiment is to find out whether people can use the multimodal interface to develop a mental image of the spatial layout of a Web page. A group of 12 blindfolded participants took part in the experiment and they were asked to explore two unfamiliar Web pages in 3 min. They were requested to describe the page layout and draw it on a piece of paper after the session. During the task, participants were asked to follow the think-aloud protocol, discussing any strategies that they were using for exploration, the effectiveness of multimodal cues, the size of the objects they were interacting with, and providing any general feedback as to the strengths and weaknesses of the system.

The Web pages used in the experiment are shown in Fig. 6. One of the Web pages is conceptually simpler to explore due to its small number of widely-spaced elements (1 heading, 5 hyperlinks, 2 image-links, 1 image, and text); while the other one is more complex with tighter-packed elements (14 hyperlinks, 9 image-links, 1 image, and text).

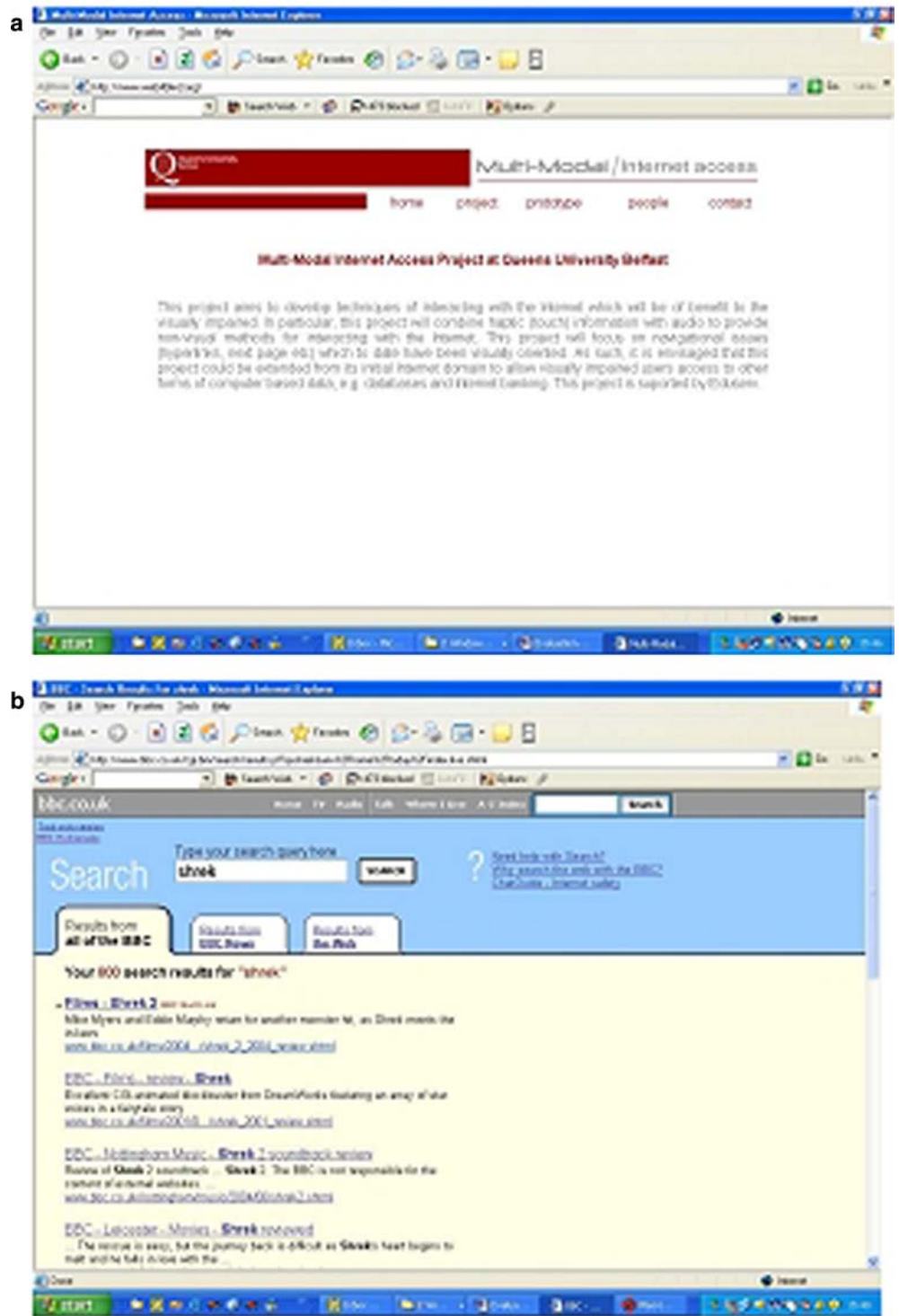
- (a) Simple Web page tested;
- (b) Complex Web page tested.

Before the session, participants were given 5 min of training on a non-complex sample Website to familiarise themselves with the multimodal cues representing images, hyperlinks and text. To improve the learning process of multimodal feedback, cues were introduced unimodally, and then in combination with other feedback. Participants were asked a series of questions during the training session, remarking on the perception and quality of the multimodal cues, and whether distinctions could be made between feedbacks of various elements. This was to ensure participants had acquired the necessary skills to use the multimodal interface.

#### 3.5.2 Section 2: Navigation to target objects on a Web Page

The main objective of this part of the experiment is to compare the multimodal interface with JAWS 5.0 screen reader for Windows [27], in terms of locating interesting items on a Web page. The 12 participants were asked to explore two unfamiliar commercial Web pages in order

Fig. 6 Web page used in experiment Sects. 1 (a) simple web page tested and (b) complex web page tested

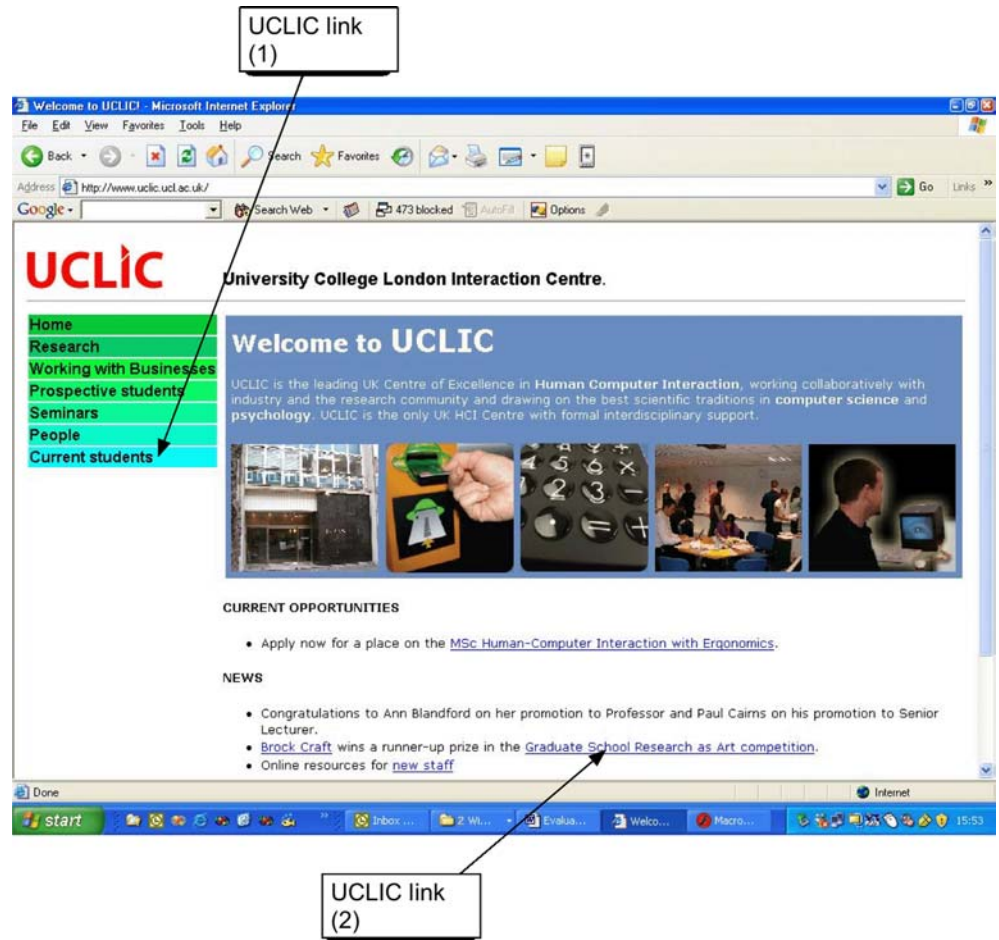


to locate a designated object on each Web page. All participants needed to perform the task using both tools (1) JAWS with Internet Explorer 6.0 and (2) multimodal interface with Mozilla Firefox 1.0. Six participants performed the task-using tool (1) first, whilst the other six used tool (2) first, the randomisation process was performed in order to minimise the learning effect which

might affect the experimental results. A maximum time limit of 5 min was imposed on the participants.

The Web pages used in the experiment are shown in Fig. 7. One of the Web pages is conceptually simpler to explore due to its small number of elements (12 hyperlinks, 1 image-link, 5 images, and text); the other one is more complex with tightly packed elements (1 hyperlink;

**Fig. 7** Web pages used in experiment Sects. 2 (a) target objects on the simpler Web page (b) target objects on the complex Web page



26 image-links, 10 images, and text). Objects to locate on the Web pages included a hyperlink on the simpler Web page, and an image-link on the more complex page. Two different objects were carefully selected on each Web page to ensure that the level of difficulty would be similar for both tools.

- (a) Target objects on the simpler Web page.
- (b) Target objects on the complex Web page.

Again, a 5 min training session was given before the experiment. Participants were introduced to JAWS and the main commands [28] that visually impaired people would use when browsing the Web. They were allowed to practice the commands during the training stage.

### 3.5.3 Section 3: Usability of multimodal browser

At the end of the experiment, participants were presented with a questionnaire probing their perceptions of multimodal interactions. Questions were adapted from usability surveys, soliciting views on the participants' Web experiences using the multimodal interface. These included asking the user if he/she felt confident when accessing the multimodal interface and exploring with the force feedback mouse, whether the system was un-

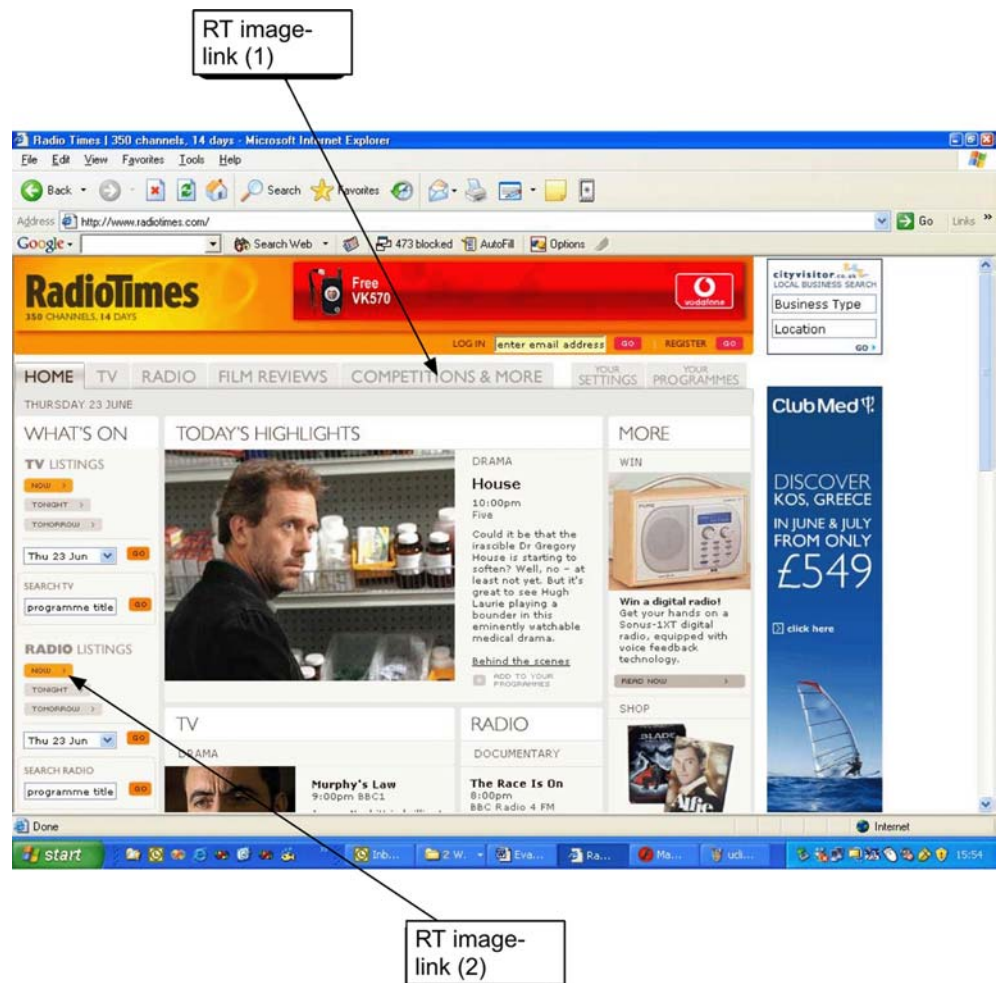
duly complex to negotiate, and whether technical support would be required for future access. The second part of the questionnaire related to issues of engagement and effectiveness of auditory and haptic cues. Data was captured in quantitative format through the use of Likert scales. Range from 1 to 5, with 3 indicating a neutral response, and greater than 3 indicating a positive response. A short follow-up interview was conducted to discuss issues arising from the questionnaire.

## 4 Results and discussion

### 4.1 Section 1: Spatial awareness of object layout on a Web page

All 12 participants were able to provide a verbal account and produce a diagram of their mental model of each page. Verbal descriptions were brief yet yielded rich information concerning the number of elements on a page and their respective locations. On the simple Web page, responses detailing positional layout were generally quite accurate due to the simpler structure of the page. Participants were able to communicate effectively the position of images at the top left of the Web page,

Fig. 7 (Contd.)



along with the position of the text towards the bottom of the page. They managed to communicate the presence of hyperlinks, mapping out the correct position on the Web page. However, participants were uncertain about the number of links on the page, stating that there were maybe two to three links. This sense of occlusion could have been attributed to the relatively small dimensions of each hyperlink, and their relatively close proximity towards one another.

Diagrammatic representations did not always reflect the richer verbal descriptions presented by participants. Representations were in sketch format detailing groups of links, images and text (Fig. 8). Participants were not always able to align the elements on paper, when compared to their verbal descriptions. Errors in alignment could have also arisen from the amount of elements on the page, which users needed to remember. Limited workspace of the force feedback mouse might have also affected participants' perception of object alignment.

Participants indicated that exploring the complex Web page proved to be a challenging task. Verbal descriptions for the task were again informative, but due to the complexity of the page, some of the participants did not feel that they were given adequate time to gain a

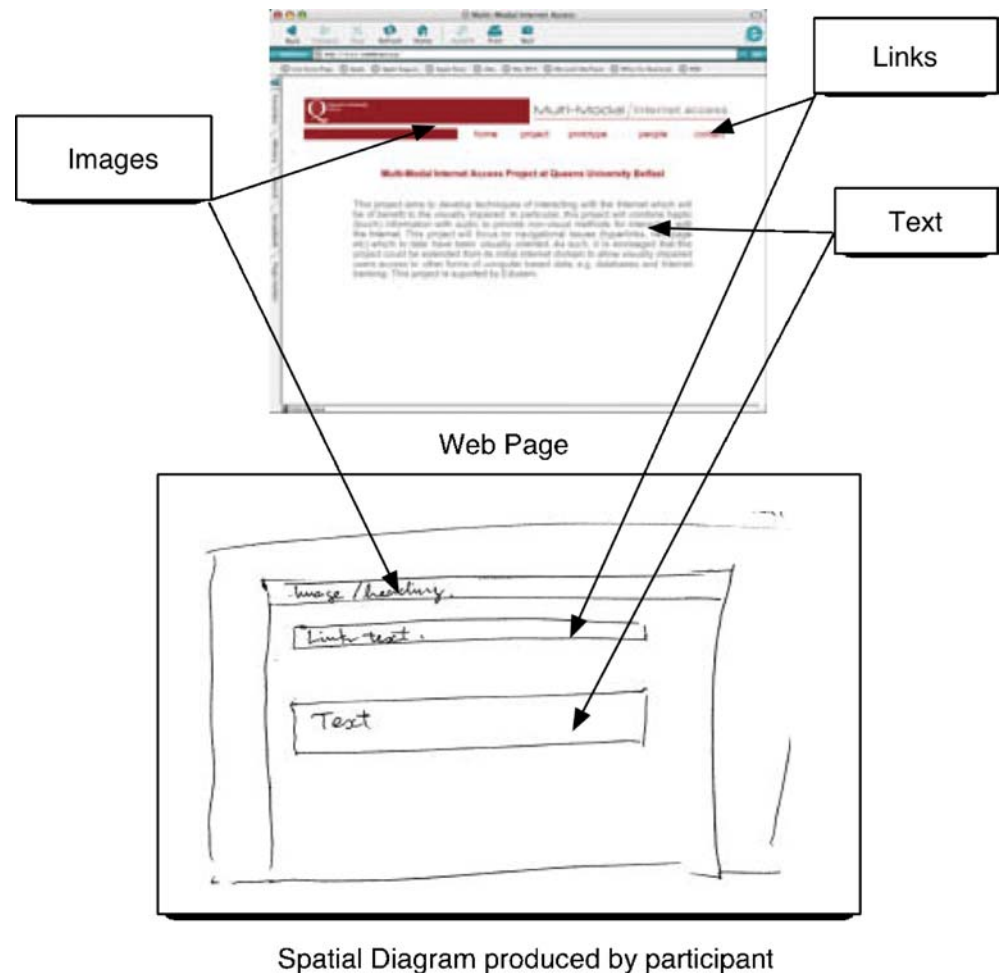
good overview. The position and number of hyperlinks was again found to be difficult for the users to describe, without providing a rough estimation. Reasons for this could have included the long descriptions arising from the speech component of the interface, detailing the rather long search term hyperlinks. Diagrammatic representations were again not as rich as the verbal descriptions (Fig. 9). Participants were able to remember many of the components of the Web page, but alignment on paper was found to be a challenging task.

Participants were observed moving the force feedback mouse at a quick speed, causing them to skip-over visually smaller elements in a page. Many of the fully sighted users were used to moving a mouse quickly around a GUI, in their day-to-day work. The point of confusion was caused as the physical distance moved by the mouse, did not correspond to the distance moved by the mouse cursor on the screen. Slow and controlled movements would need to be made using the force feedback mouse to gain an adequate perception of elements on the screen.

Other points of confusion were attributed to the lack of alternative text presented for larger and smaller images on the Web page. Participants could feel



**Fig. 8** An example of diagrammatic representation of the simple Web page



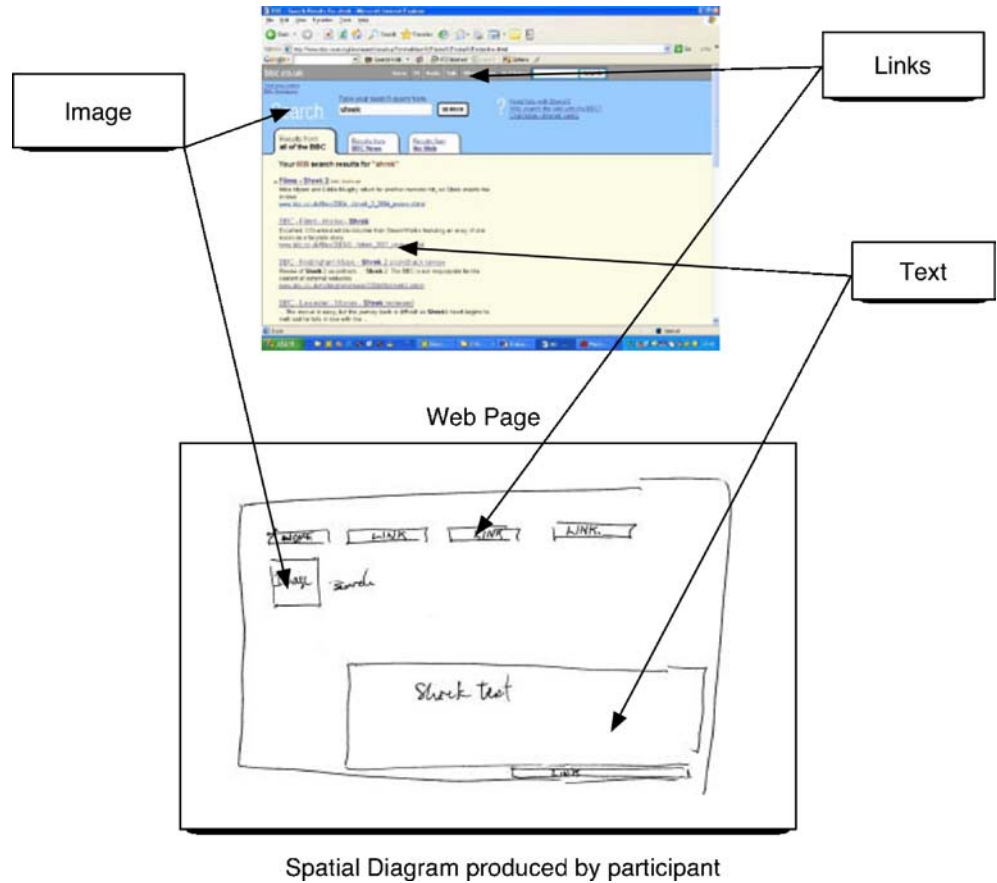
auditory and haptic feedback for these elements, but were unaware of their relevance on the page, without a textual description. Spacer images are often included in Web pages to maintain a standard distance between page elements when viewed through a browser, but if incorrectly labelled, would offer no benefit to a visually impaired user. Image-links were also perceived incorrectly. The participants indicated that they did not find it intuitive to perceive the haptic signals and auditory icons for an image-link. The result of further discussions revealed that participants would benefit from separate cues, distinguishing image-hyperlinks from ordinary images and hyperlinks.

Throughout the spatial awareness of object layout task, participants were encouraged to discuss strategies employed for developing a visualisation of the Web pages, whilst blindfolded. The majority of them were observed initially adopting a trial and error method, to isolate elements on a page. Often a haphazard method was adopted, with the participants randomly moving the mouse around, in the hope that they would find the target. It was obvious after more practice, they seemed to develop a strategy for exploring elements on a page to gain an overview. One music student tended to work in a clockwise motion, moving in a spiral from the outside of

the browser, slowly inwards. When asked about her approach, she mentioned it was a good way to spatialise the information on a Web page. Two other science students from an engineering background adopted another methodical approach. They tended to move to the outside of the browser where an auditory icon was played, to signal the content border. They would then move to the left hand side of the page to detect a reference point, such as an image, link or text. The two participants would move slowly around the reference point to try and detect another object, drawing a map in their minds of the position of elements on the screen. They would then move outside the vicinity of the browser window, to re-orientate themselves on the screen and try to detect other objects in the vicinity. Some participants would also try to move in a vertical line to try and establish an axis in their mind and use this axis to orientate themselves on the page.

Some comments were given by the participants on how to improve the interface. One participant suggested feedback to provide awareness of the mouse cursor position on the Web page. The participant recommended that this feedback could be accessed by clicking the right mouse button, which would provide the user with the option of accessing the position, without receiving

**Fig. 9** An example of diagrammatic representation of the complex Web page



continuous feedback whilst exploring the interface. Another participant suggested a facility to re-position the mouse cursor to the top left of the Web page either by making a keystroke or placing a multimodal icon at the location in question, to give confirmation of position.

#### 4.2 Section 2: Navigation to target objects on a Web Page

The experiment that compared JAWS screen reader and the multimodal interface in target object navigation showed very interesting results. The main measurement was the task completion time. Observations were also made on the strategies that participants adopted in the searching process. Almost all participants were able to complete the task and locate the target objects, with the exception of two participants who failed to find the object (UCLIC link 1) on the simple Web page in 5 min using the multimodal interface. Overall, participants took less time to locate the objects using the JAWS screen reader. Figure 10 shows the comparison of task completion time.

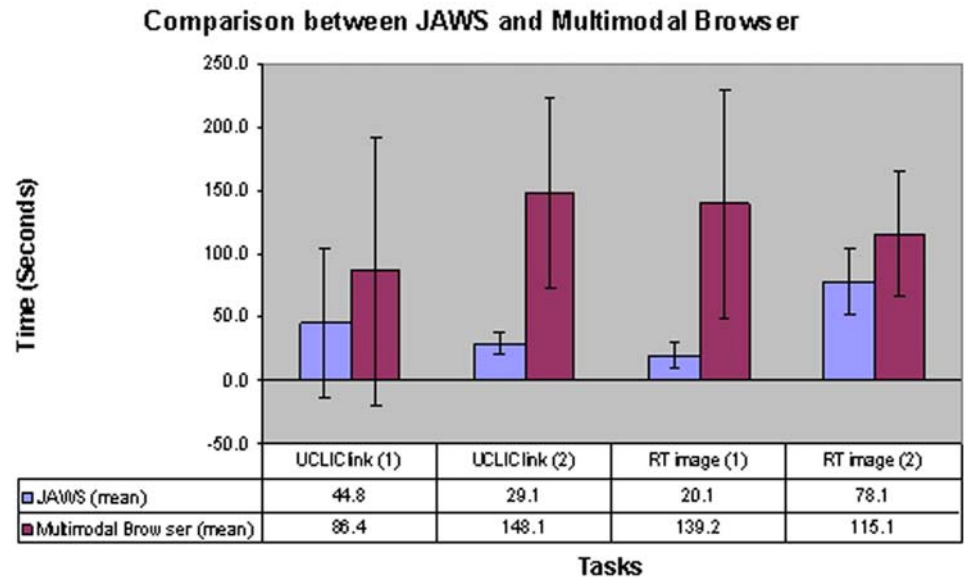
The task completion time in JAWS is consistently lower than in the multimodal interface. The standard deviations (STDEV) are also very low except in the condition of UCLIC link (1) in which the STDEV is 58.3 s. This exception is due to a large task completion

time (163 s) required by one participant. Without taking into account of the result from this participant, the average task completion time would have been 21.1 s with STDEV 7.4 s. This is in line with the figures gained under the other experimental conditions.

The amount of task completion time in JAWS condition increases with the Tab Order value of the target object, which is usually determined by the object's location on the Web page. This is because screen readers read the content of a Web page in a linear fashion. To browse through the objects on the page, participants needed to use the Tab key to go through them one by one. If the target object was placed further down in the page, the time needed to reach the intended object would be longer. As a result, the task completion times are low and consistent (21.1 s without the exceptional case, and 29.1 s) in the simple Web page condition where only a few items are on the page. The complex Web page, on the other hand has more items. The task completion times for the complex page, vary from 20.1 to 78.1 s based on the locations of the target objects.

Participants generally required more time to find the objects using the multimodal interface compared with the JAWS condition. There are also large variations in the task completion times. Some participants can find the target objects in a very short time, for example, 17.7 s on the simple Web page (UCLIC link 1), which is comparable to the shortest time in the JAWS condition,

**Fig. 10** Comparison of task completion time



12 s. On the other hand, one participant spent 148.8 s on the simple Web page (UCLIC link 1) using the multimodal browser, accounting for the larger STDEV of 105.8 s. Two participants could not complete the task on the same page (UCLIC link 2).

The large variations in targeting elements using the multimodal interface seem to be subjected to individual differences. Further study with more participants will be required to obtain conclusive results. However, a number of factors that contribute to the results can be identified in this study. The lack of a visual overview of a Web page presented major difficulties to participants. They needed to adopt a strategy to locate the objects on a page. During the course of the experiment, it was observed that not of all the strategies utilized were effective.

Participants often adhered to their perception of how a Web page should resemble when navigating. Many of them initially directed themselves to the left-hand side of the screen when searching for a hyperlink, intuitively expecting the hyperlink would be located in the vicinity. If the hyperlinks were not present there, they explored horizontally along the top of the page. When searching for images, participants tended to navigate towards the bottom right section of the page, where they believed the main content of Web pages to be. Confusion tended to arise from larger open spaces on a Web page, where participants tended to encircle the area in the hope of locating an object.

Second, besides the location of the object, the size of the object also affects the searching time. Usually, the larger the object, the easier it could be found using the multimodal interface. The results of the experiment do not quite show that participants spent less time on the bigger objects. On the simple Web page, the location of the larger-sized object (UCLIC link 2) seems to be the main reason why participants found it hard to find. The

object is down at the bottom of the page and easy to miss. On the complex Web page, even though the task completion time for the smaller object (RT image 2), is shorter than the bigger object, its time variation is also smaller, 49.8 s compared with 90.6 s for the bigger object. The shortest time in the bigger object (RT image 1) is 28.3 s and the longest time is 275.6 s. Therefore, the task completion time for the bigger object is less consistent and requires more participants to give a conclusive result.

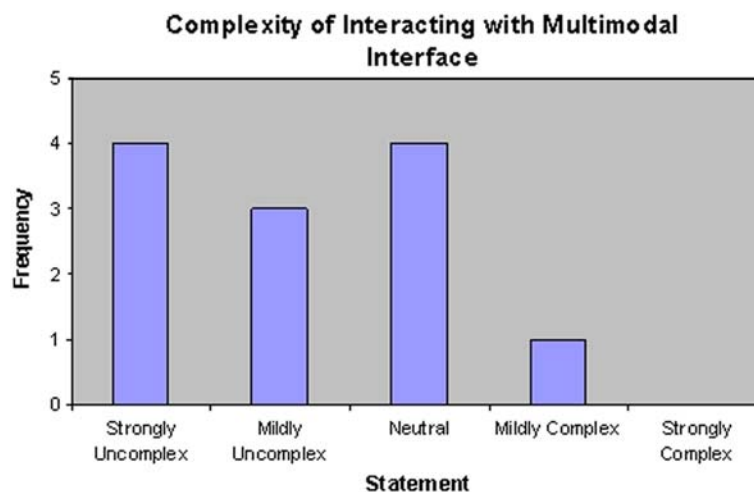
Navigating using the multimodal interface was slower due to the amount of time participants spent exploring each object on a page in order to develop a mental map of the page layout. Participants indicated that when looking for an item on a page, they were more aware of elements and their respective positions, which they could not obtain through the use of JAWS. This resulted in a greater perceptual experience, which was not experienced with screen readers.

Sighted participants did not necessarily have an advantage in the series of experiments undertaken, due to their previous knowledge of working with a mouse. It was acknowledged that just as the experience of mediating a two-dimensional interface with a mouse would be new to the visually impaired participants, the concept of using a screen reader would also be novel to fully sighted users. To account for the learning curves experienced when first interacting with each piece of software, additional training would be provided before future evaluations commence.

### 4.3 Section 3: Usability of system

Participants indicated that they found the multimodal interface to be straightforward to use, after their initial period of training. The majority of participants agreed

**Fig. 11** Feedback on the complexity of the multimodal interface



with the statement that Web pages were not difficult to negotiate using the force feedback mouse, with seven out of 12 participants agreeing that the experience had been non-complex and usable (Fig. 11).

Participants did indicate that they would benefit from additional training when interacting with the interface, to remind them on the meaning of various multimodal cues. They found the system to be learnable but would benefit from additional practice before performing tasks on complex pages. Once the meaning of multimodal cues were clarified in their mind, confidence levels with the interface would improve.

Participants were asked to rate the usefulness of the haptics, speech and non-speech audio in the system. For the force feedback cues, participants rated the feedback for images positively, as the ridge effect around an image seemed intuitive. They were able to develop a spatial representation of the boundary of the image, adding vital context to a page. Feedback for hyperlinks could be sharpened to ensure that participants would not skip over links or could have the option of being constrained to hyperlinks, if they so wished.

In terms of audio, the auditory icons used for indicating the presence of images and hyperlinks were thought to be meaningful. A camera click would automatically conjure the image of a camera. However, participants explained that these icons were too short in duration and easily masked by the pitch and panning. Pitch was thought to be more intuitive if the value increased moving towards an image or hyperlink, rather than reducing whilst moving towards object on a page. Participants could hear the residual noise made by the motor of the mouse and thought that this was an additional source of audio for the interface. Whereas noises from the motor did not cause confusion, participants were finding they were concentrating harder on separating the auditory icons and background sounds from the motor sounds. This could be remedied through the use of stereo headphones, which could also convey the panning in a more effective way, thereby improving spatialisation.

The quality of voice was also opened as an item of discussion, with users stating a preference for the softer tone used by the multimodal interface. Five out of the 12 users asked for improvements with the technology. Further discussions revealed a preference for a more humanised voice that did not mispronounce names and words. They speculated that listening to the synthesised voice for prolonged periods would lead to eventual overload, also attributed to the verbosity of information read out. The option of customising the amount of information that the interface could read out was also considered as a viable method of designing an inclusive system.

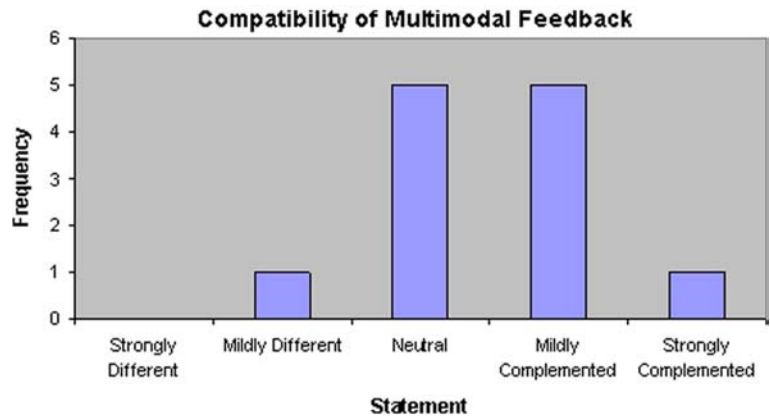
The majority of participants indicated that the multimodal cues worked well in conjunction with each other (Fig. 12). On further discussion, it was revealed that accessing the Web using the multimodal system was initially slightly overloading, as they had not been practiced in processing simultaneous sounds. However, during the course of the tasks, they were able to surpass this barrier and processing feedback more effectively. This resulted in a more natural and enjoyable experience, compared to using a conventional screen reader. Participants were able to maintain engagement of all their senses, without the fear of sensory overload.

In terms of future improvements for the system, participants suggested that a larger work space for the mouse would pose fewer constraints. By being able to navigate in a space roughly the same size as the screen, participants would have increased awareness of the mouse cursor location on the screen.

#### 4.4 Assessment with visually impaired people

In order to assess the overall accessibility and usability of the multimodal interface, a second experiment was conducted with visually impaired people. The experiment was divided into two sections, Sect. 1 to examine whether visually impaired users could obtain spatial awareness of positional layout and Sect. 2 to reveal

**Fig. 12** Feedback on the compatibility of multimodal feedback



whether the interface would provide an accessible and usable means to exploring a Web page.

Seven participants from the Royal National Institute for the Blind Youth Group aged between 14 and 25 were recruited for the trial. All participants identified themselves as visually impaired or blind, with sight loss ranging from being able to see larger blurred images on the screen with the help of magnification software to total occlusion. Three of the seven participants mentioned that they had a stronger level of sight in younger years. Six of the seven participants had knowledge of screen reading technology, and had Internet training in the past. Three of the six participants described themselves as beginner to intermediate level and three at a more advanced stage. Two participants had previous experience with a computer mouse in their younger years. None of the participants has experienced a force feedback mouse before.

#### 4.4.1 Section 1: Spatial awareness of object layout on a Web page for visually impaired users

All participants were asked to explore two unfamiliar Web pages without the help of the evaluators (Fig. 13). The experiment set up and procedures were similar to the one used for sighted people. Participants were given 5 min to explore on each Web page. The increased time is to accommodate participants' unfamiliarity with the use of computer mouse in the multimodal interface. After the session, participants were given the choices of describing the Web page layout using either the pen and paper or tactile objects (Lego). A slightly different complex Web page was used (13 hyperlinks; 3 image-links; 1 image; text) as one of the visually impaired users was already familiar with the complex page in the first experiment.

Participants were provided with 5 min of training on a non-complex sample Web site, using Mozilla Firefox 1.0, to familiarise themselves with the multimodal cues representing images, hyperlinks and text. As most of the subjects had not previously accessed a mouse, users were offered additional instruction.

#### 4.4.2 Section 2: Accessibility and usability of multimodal browser

All participants were presented with a series of questions aiming to assess their perceptions of multimodal interactions with the browser. An open-ended style questionnaire was used to solicit views on the benefits and disadvantages that the multimodal interface offers in comparison to current assistive technologies. Views could be followed up during the questionnaire. Participants were encouraged to discuss their abilities to process simultaneous sources of auditory feedback, usability of the force feedback mouse and enjoyment arising from using the system. They were asked on their ability to rate the perceptual experience offered by the interface, in comparison to existing assistive solutions.



**Fig. 13** Blind user taking part in evaluation

## 5 Results and discussion

### 5.1 Section 1: Spatial awareness of object layout on a Web page

Participants were able to explore the simple Web page and provide a relatively good description of the page layout, whilst five participants were not able to explore the whole complex Web page within the time limit given. Participants were presented with a choice to either align tactile objects or create a diagrammatic representation in order to recreate the location and layout of objects on the Web page. Two users chose to draw the location of objects. Diagrammatic representations produced by participants were found to contain minor inconsistencies with the sizing and positioning of objects (Fig. 14). This could have also been attributed to the fact that both users were unfamiliar with drawing skills and could not mark points on the diagram, which they could later use for reference.

Using tactile objects, participants were able to align artefacts representing images and hyperlinks in a given order. Tactile artefact representation varied from participant to participant; some similar to the visual layout of the page, some radically different. Alignment of hyperlinks and images using the tactile objects was often poorer, particularly for the “busier” Web page. Many of the participants described the process of visualising a

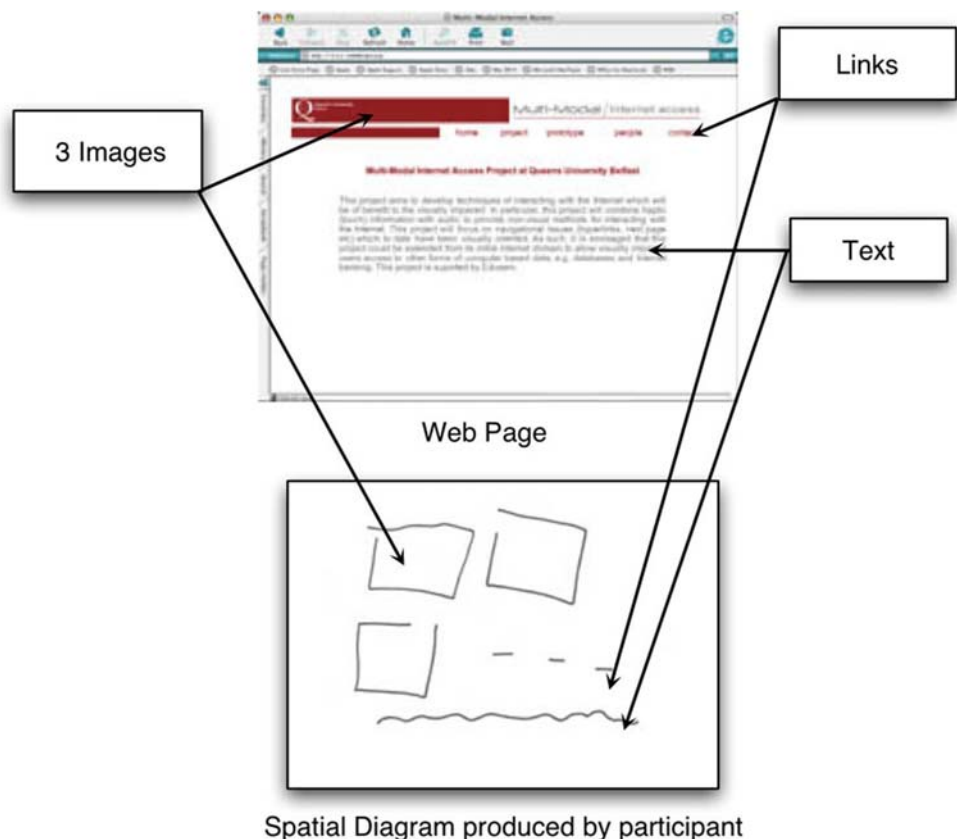
Web page the way that fully sighted people would visualise a page, quite difficult. Individual differences, including experience with tactile arrangements and age could have also been grounding factors.

Observations were made on the methods adopted by participants, to explore the Web pages. After initial cautiousness using the mouse, many of the participants spent most of the time period, navigating vertically, orientating themselves towards the left-hand side of the Web page. Further discussion with participants revealed that participants expected the pages to resemble a vertical list of text and links. This model had been formed, due to the sequential format offered by screen readers. Exploration patterns also appeared to be more strategic for visually impaired people. Many of the participants remarked on being able to find a large reference point on a Web page, such as an image, and trying to move slowly around it to find another reference point. This way, they could draw a virtual map in their mind. When asked to verbalise a description of the Web page, the participants were able to provide a fairly clear representation.

### 5.2 Section 2: Accessibility and usability of multimodal browser

Analysis of post-task questionnaires revealed that interface functionality was not found to be unduly

**Fig. 14** An example of diagrammatic representation of the simple Web page



complex or fatiguing for the participants who stated that the system provided benefits for visualisation, and they expressed confidence in being able to use the system in the future, unaided. Multimodal cues were found to complement each other, providing a novel, engaging experience for the participants when interacting with the Web.

Initially, most blind participants found the mouse difficult to control. One stated that he could not visualise the speed of the mouse, suggesting that he would like a sense of how fast the cursor was moving. Three participants suggested that the base of the force feedback mouse should be larger, almost the size of the actual screen so that physical movements could be closer to cursor on the screen.

Visually impaired participants found the process of hovering over hyperlinks and clicking the mouse to select the hyperlink quite difficult. This could be due partly to inexperience working with a mouse and difficulties visualise the physical position of the mouse cursor over the narrow hyperlink body. Reduced vibration force feedback over the hyperlink could improve the situation, stopping the user from moving off. Generally blind participants indicated that they would like haptic effects to constrain their cursor movements within the page. The concept of a haptic groove for a link or a list of links that would make hovering on a link easier was considered to be beneficial. One participant felt that constraining the cursor could be confusing for some visually impaired users, therefore this should be an optional feature.

Visually impaired participants generally appeared more confident than their sighted counterparts at processing sounds simultaneously. Most of them could perceive changes in pitch, panning and could make use of auditory icons. They were able to identify the sound of the camera click and metallic chains, as representing images and links on a Web page. They considered the metaphors to be appropriate in that they understood the fact that the camera noise implied that they were about to enter an image, even though some of them had never experienced a visual image. The locational earcon was considered useful to identify the proximity of a link or image although participants felt that this should be developed further to convey more information about the cursor position in relation to the image or link. One participant suggested the use of more descriptive auditory icons to provide information on the direction of cursor movements. For example auditory icons could be designed specifically to evoke upwards and downwards and sideways movements.

In terms of future development, participants suggested that other parts of a page should be rendered to offer additional feedback as feedback of moving in and out of the browser was not found to be effective enough. Additional auditory icons to mark direction and position on the page, would offer clues to rectifying moving away from the main body of the page in error. A haptic

barrier may act as one method of preventing users from leaving the browser, until they wanted to transfer to another application.

The text-to-speech synthesiser used in the plug-in, was found to produce a more pleasant experience than other conventional screen readers. Participants considered it to have a softer, more human-like tone. This is an important feature for visually impaired Internet users when listening to synthesised speech for prolonged lengths of time. Participants found it difficult to compare the multimodal interface with JAWS screen reader in terms of navigation as the two systems were so different. Experienced visually impaired screen reader users felt that they could navigate links faster using JAWS but could not compare the interfaces in terms of spatial awareness as this is not a feature of JAWS or any other screen reader. Participants said that they would need more experience with the multimodal browser for a realistic comparison to the screen reading technology that they were familiar with in terms of speed of navigation. However, participants stated that spatial information conveyed by the multimodal interface provided a much richer navigation experience than that possible with a conventional screen reader.

In the current prototype, text is read by paragraph. In future systems, participants would like to have more control over synthesised speech for non-link text on a Web page in terms of speed, volume and duration.

The multimodal interface has not taken into account the issue of scrolling through a Web page. Information would need to be conveyed about the existence of elements currently occluded from view, and allow for participants to orientate their position when scrolling within the page. A future version of the prototype may examine the adoption of this feature, which could help visually impaired participants to navigate through a page.

Further discussion of the prototype yielded suggestions of additional multimodal feedback for (1) determining whether a user is inside the Web page or on the browser toolbar, (2) additional haptic constraints using the force feedback mouse, (3) a summary of page attributes, and (4) spatial positioning to be presented when the user arrives on a Web page. These would culminate in greater levels of usability as less time would be wasted when moving from page to page. No effects of sensory overload were reported in the trials. According to our user requirement capture, many visually impaired Internet users interact with the Web for periods of 3–4 h at a time. Future evaluation would need to focus on whether extended use with the multimodal interface would lead to sensory overload effects or increase levels of cognitive workload on the user, and examine ways to minimise the potential risk.

For the purpose of conducting evaluations on future prototypes, we would like for both sighted and visually impaired participants to follow the same conditions within experiments. However, the variability inherent in blind and visually impaired users may make this a

complex process. Factors include the length of time since the onset of blindness, differences in the levels of education and technical skills, physical and cognitive disabilities, acknowledged by Stevens and Edwards [25]. The authors have subsequently recommended using a method of co-operative evaluation as an alternative to the strict controlled experiment paradigm to evaluate assistive technologies, which may help to overcome the issue of variability discussed, which could be implemented in our future studies.

---

## 6 Conclusion

This paper has described a novel technique, which can determine when a user's cursor is in close proximity to a region of interest on a Web page, e.g. a hyperlink or image. By rendering the spatial visual information via the multimodal interface, visually impaired people are not only informed of these regions of interest, but are also guided to them by the audio and haptics. The evaluation of the multimodal interface has shown that it can assist users in the construction of a mental map of the Web page layout, which is impossible with the current screen-reading software. The experimental results have also revealed that using the multimodal interface is slower to search and locate an object on a Web page compared to the screen reader. However, after experiencing the multimodal interface, the visually impaired users in our study, showed their appreciation for the sense of spatial awareness and navigational information provided by the haptic and audio features. They have provided valuable feedback on the advantages and limitations of the multimodal interface, which will be taken into consideration for future implementations.

---

## References

1. The Web Access and Inclusion for Disabled People, A Formal Investigation conducted by the Disability Rights Commission. <http://www.drc-gb.org/publicationsandreports/report.asp>
2. Hakkinen M, Dewitt J (1996) WebSpeak: user interface design of an accessible web browser. White Paper, The Productivity Works Inc., NJ
3. Zajicek M, Powell C, Reeves C (1998) A web navigation tool for the blind. In: Proceedings of the 3rd ACM/SIGAPH on assistive technologies, pp 204–206
4. Donker H, Klante P, Gorny P (2002) The design of auditory user interfaces for blind users. In: Proceedings Nordichi, pp 149–156
5. Goose S, Moller C (1998) A 3D audio only interactive web browser: using spatialization to convey hypermedia document structure. In: Proceedings of the 7th ACM international conference on multimedia, pp 363–371
6. Roth P, Petrucci LS, Assimacopoulos A, Pun T (2000) Audio-haptic internet browser and associated tools for blind users and visually impaired computer users. Workshop on friendly exchanging through the net, pp 57–62
7. Betacom ScreenRover. <http://www.elkshelp.org/screenrover.html>
8. Ramstein C, Martial O, Dufresne A, Carignan M, Chassé P, Mabilieu P (1996) Touching and hearing GUI's: design issues for the PC-Access system. Proc Assist Technol 2–9
9. Brewster SA (2001) The impact of Haptic 'Touching' technology on cultural applications. In: Proceedings of the electronic imaging and the visual arts, pp 1–14
10. Yu W, Reid D, Brewster SA (2002) Web-based multi-modal graphs for visually impaired people. In: Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology, pp 97–108
11. Parente P, Bishop G (2003) BATS: The blind audio tactile mapping system. In: Proceedings of the ACMSE, Savannah, GA
12. IFeelPixel: Haptics & Sonification <http://www.ifeelpixel.com/faq/#whatitwill>
13. Caffrey A, McCrindle R (2004) Developing a multimodal web application. In: Proceedings of the 5th international conference disability virtual reality and associated technologies, pp 165–172
14. Lahav O, Mioduser D (2000) Multisensory virtual environment for supporting blind persons' acquisition of spatial cognitive mapping, orientation and mobility skills. In: Proceedings of the 3rd international conference disability virtual reality and associated technologies, pp 53–58
15. Jansson G, Monaci L (2005) Haptic identification of objects with different numbers of fingers. In: Ballesteros S, Heller MA (eds) Touch, blindness and neuroscience. UNED Press, Madrid
16. Sharmin S, Evreinov G, Raisamo R (2005) Non-visual feedback cues for pen computing. In: Proceedings of the World Haptics, pp 625–628
17. Internet Explorer. <http://www.microsoft.com/windows/ie>
18. Mozilla Firefox. <http://www.mozilla.org>
19. Web Content Accessibility Guidelines 1.0. <http://www.w3.org/TR/WAI-WEBCONTENT>
20. DOM: Document Object Model. <http://www.w3.org/DOM/>
21. Turner D, Oeschger I (2003) Creating XPCOM components, Open Source
22. Maclean K (1999) Application-centered Haptic interface design. In: Srinivasan M, Cutkosky M (eds) Human and machine haptics. MIT Press, MA
23. Representations of Visual Geo-Spatial Information. <http://cnice.utoronto.ca/guidelines/geospatial.php>
24. Microsoft Speech SDK. <http://www.microsoft.com/speech/download/sdk51>
25. Stevens RD, Edwards AN (1996) An approach to the evaluation of assistive technology. In: Proceedings of the 2nd annual ACM conference on assistive technologies, pp 64–71
26. Yu W, Ramloll R, Brewster S, Riedel B (2001) Exploring computer-generated line graphs through virtual touch. In: Proceedings of the 6th international symposium on signal processing and its applications, pp 72–75
27. Freedom Scientific JAWS 5.0 Software. [http://www.freedomscientific.com/fs\\_products/software\\_jawsinfo.asp](http://www.freedomscientific.com/fs_products/software_jawsinfo.asp)
28. JAWS commands help guide. Freedom Scientific Software