AUDIO FOR A MULTIMODAL ASSISTIVE INTERFACE

Demo paper for the ICAD05 workshop "Combining Speech and Sound in the User Interface"

Emma Murphy, Graham McAllister, Philip Strain, Ravi Kuber & Wai Yu

Sonic Arts Research Centre & Virtual Engineering Centre
Queen's University Belfast
University Road
BT7 1NN

e.murphy@qub.ac.uk

ABSTRACT

This paper details the design of an audio interface for a multi-modal content-aware web plug-in. The system aims to provide spatial and navigational information to visually impaired Internet users through speech and non-speech audio with haptic feedback. The web plug-in and audio interface are presented and discussed, along with recommendations for future system development.

1. INTRODUCTION

Despite significant advances in assistive technologies and accessibility guidelines for web design, visually impaired Internet users still experience great difficulties in comparison to sighted users. Commercial screen-reading applications such as JAWS [1] are extremely popular and useful for visually impaired computer users, but do not resolve all of the problems that blind Internet users experience. A user requirement questionnaire with 26 blind and partially-sighted people was carried out to investigate the main problems that visually-impaired Internet users face. Due to the lack of positional feedback offered by a screen-reader and keyboard, gaining an overview of the spatial layout of objects on a web page was found to pose a great challenge. Users indicated that they would like to be more aware of the position of images and links on a web page in order to improve navigation and also to attain a similar perceptual experience to that of a fully sighted user. A multimodal web plug-in, developed to give spatial information through audio and haptic feedback, is detailed in this paper. The web plug-in is described and evaluated with particular attention to audio interface design.

Sound design has been crucial to the development of interfaces for both commercial and academic assistive technology applications. Speech synthesis is fundamental to assistive interfaces for blind users but non-speech audio, such as auditory icons [2] and earcons [3], has also been effectively applied to non-visual interfaces. There has been previous research carried out into the possibility of using speech and non-speech audio specifically to improve web accessibility. Goose et al. [4] and Donker [5] have created audio browsers based on 3D sound spatialisation to convey information on web pages. Patrick Susini et al. have attempted to create *Sonified Hyperlinks* analogous to visual hyperlinks in HTML[6]. Audio Enriched Links software has been developed to provide previews of linked web pages to users with visual impairments [7]. WebSound [8, 9] is a sonification tool where sonic objects are associated with HTML tags, which are then projected into a virtual 3D sound space according to finger position on a tactile touch screen.

Mapping Internet navigational and structural information to sound is a challenging task in that a sound designer cannot have control over the structural design of individual web sites. Furthermore, although there is a small body of research aimed at understanding and formalising the aesthetics of auditory design [10, 11], the field of sound design for auditory display does not have a formalised design framework. The initial design for the auditory interface in this research has highlighted certain issues in sound design particularly in regard to non-speech sounds and how these relate to speech in an auditory interface.

2. ASSISTIVE BROWSER PLUG-IN

An assistive tool, consisting of a multimodal interface with haptic and audio feedback and a content-aware web browser plug-in, has been developed to aid visually impaired Internet users perceive spatial information.

2.1. Content-Aware Web Plug-in

The aim of the content-aware plug-in is to present users with feedback to the nearest graphic or link object, from their relative position on the page. A force feedback mouse is used and its cursor position is constantly monitored by the plug-in that detects the surrounding objects. When the plug-in identifies an object in the vicinity of the cursor, the cursor position, relative to the object, is mapped to audio and haptic feedback. An overview of the system architecture is illustrated in Figure 1. The plug-in uses Mozilla's Firefox browser, which offers advantages for developing extensions through accessibility of source code and cross platform compatibility. The *Microsoft Speech SDK* [12] was utilised to provide speech synthesis via the web plug-in.

The plug-in functions by capturing the current position of the mouse cursor, and then by obtaining the position of each HTML element on the screen. The relative co-ordinates of the mouse pointer are calculated within a distance of an HTML element. The element and surrounding area is divided into nine sections, where each section has a particular coordinate range. Finally the relative coordinates of the HTML element are sent to the haptic and audio components.

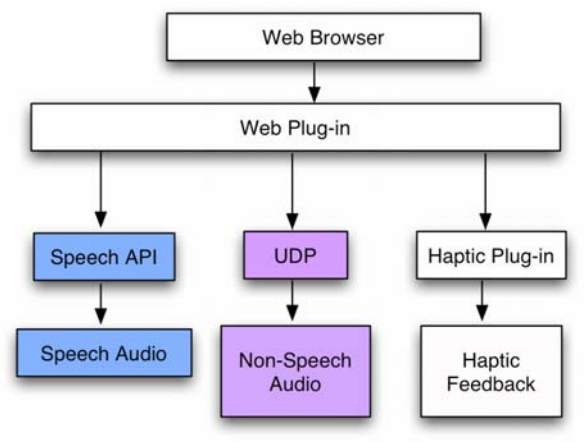

Figure 1. System Architecture.

2.2. Audio Interface

2.2.1. 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 is designed and played back in Max/MSP, a real-time audio programming environment. Netsend, an MSP external object is used to receive x and y location co-ordinates sent via UDP from the web plug-in. Figure 2 illustrates how the element is divided up, and the range of coordinates 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.

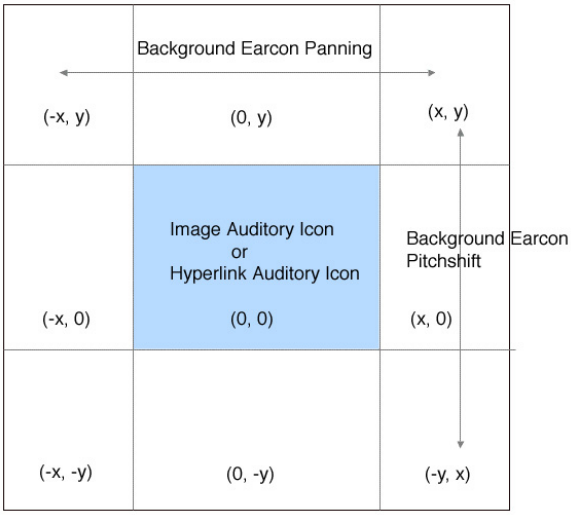

Figure 2. Object Co-ordinates and Audio Feedback

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

2.2.2. Speech Audio

Text on the web page is conveyed to the user through speech audio. As the user rolls over non-link text on a page, the text is read to the user by paragraph. The speech will stop when the user moves off the text onto another object. As the user rolls over an image, alt text 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.

2.3. Haptics

The Logitech Wingman force-feedback mouse was chosen to facilitate on-screen navigation, due to its compact size and compatibility with the Firefox browser. For this system, the Immersion web plug-in is linked to the content-aware web plug-in.

The following haptic primitives were employed in the system; the 'enclosure effect' was coupled with clipping effects bordering the 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' is produced along with force-feedback. The dual effect of audio coupled with force-feedback, is intended to heighten the sense of awareness that the user is directly on the image. The 'periodic effect' is used to provide location awareness of the cursor when directly hovering over a hyperlink. This effect produces a wave that varies over time, depicting a locked sensation, when the user directly hovers over the link.

3. EVALUATION OF CURRENT SYSTEM

Evaluation methods aimed to assess whether a combination of multi-modal cues would improve spatial awareness and navigational skills. After completing a short questionnaire assessing levels of sight loss and prior experience with the web, participants were presented with a brief explanation of the multimodal browser, with optional instruction on how to use a mouse. The main experiment consisted of a series of five search tasks to assess spatial awareness of page objects, and navigation towards images and hyperlinks, following the verbal 'think-aloud' protocol (Figure 3). Participants were then invited to verbally describe the screen layout, and then represent their visualised model either through diagrammatic form or through arrangement of tactual artifacts. Participants were invited to complete a post-task questionnaire, probing perceptions on the use of multi-modality and their confidence in using the browser to locate objects.

As part of an ongoing evaluation process, tasks were conducted with two subjects; one congenitally blind and one adventitiously blind. Participants were encouraged to explore three web pages, of varying length, complexity and accessibility using the multimodal feedback. Results indicated

that both participants could successfully isolate and differentiate between images and hyperlinks on both sparse and crowded web pages. Multi-modal feedback promoted the accurate identification of object boundaries and bodies of the main objects. Short descriptive auditory icons were found to be particularly helpful in establishing meaning, for prompt object identification. This was particularly useful for busier, more complex pages, where care would need to be taken to isolate objects.

Figure 3. User taking part in Evaluation

Observation of hand and cursor movements revealed that the adventitiously blind user made fine, controlled movements with the force-feedback mouse in a slow and careful manner. The congenitally blind user took a longer period of time to adjust to the force-feedback mouse, and develop controlled movements. This could be due to the fact that the adventitiously blind user had experience of using a mouse prior to losing her sight, whereas the other user had never used a mouse before. Both subjects initially moved in a vertical path, down the left-hand side of the page to gain an overview of page layout, omitting one whole side of the screen. This response could have been attributed to the vertical-style mental model of web pages derived from using screen-reading technologies, as highlighted in the data capture stage. As the series of tasks progressed, the user's perception of the page appeared to update, as participants were able to explore all parts of the web page, without any prompting by the researchers. This contributed to more accurate verbal and externalised representations. Diagrammatic representations produced by participants were found to contain minor inconsistencies with the sizing and positioning of objects (Figure 4). 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. Future design would ensure use of tactile paper or concentrate solely on tactile artefact arrangement.

The final two tasks examined the use of non-speech audio for purposes of navigation towards images on a web page. Participants could accurately identify when they were in the vicinity of a nearby image, and could differentiate between the background auditory cues. It was noted that sharper directional cues would be needed, to help facilitate orientation closer to the object in question. Two separate non-speech audio cues should be created to prompt the user to move in a vertical or horizontal direction towards the object.

Analysis of post-task questionnaires revealed that browser functionality was not found to be unduly complex or fatiguing for users. Participants revealed 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 users when interacting with the web. The text-to-speech synthesiser used was found to produce a more pleasant experience than other conventional screen-readers. Users 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.

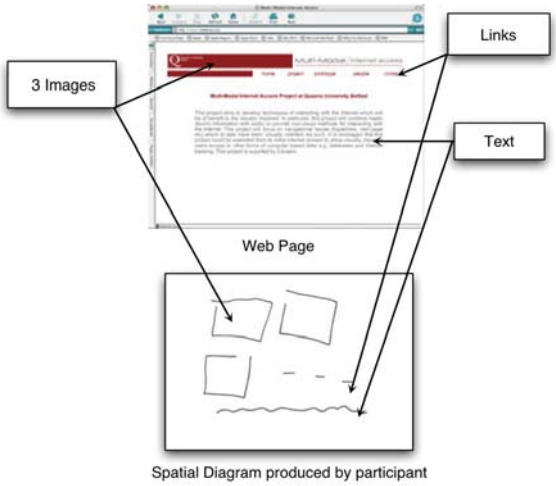

Figure 4. Comparison of web page presented to the participant, using multimodal browser, and the diagrammatic representation consequently produced

Further discussion of the prototype yielded suggestions of additional multi-modal feedback determining whether a user is inside the web page or on the browser toolbar, additional haptic constraints using the force-feedback mouse and a summary of page attributes and 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 three to four hours at a time. Future evaluation would need to focus on whether extended use with the browser would lead to sensory overload effects or increase levels of cognitive workload on the user, and examine ways to minimise the potential risk.

4. FUTURE WORK

In this multimodal system textual information on a web page is conveyed through speech while non-speech audio is used to sonify navigation and semantic information. Vast textual information on web pages can be overwhelming to read for a sighted user but can be particularly daunting for a blind user to hear through speech. Adding semantic information or navigation commands to the speech channel could lead to difficulties for the user to process information. Initial evaluation has revealed that non-speech sounds are more effective at conveying certain object information than speech. Users preferred the use of an auditory icon to convey an object as a hyperlink rather than the method of screen-readers to read "link" before link text.

In future systems we would like to further exploit the use of simultaneous speech and non-speech audio to convey more

complex information. Current functionality of the text to speech plug-in will be extended to provide full control over voice parameters such as timbre and rate. Formats such as headings or italicised text have proved difficult to detect using screen-reading technologies. As a result, users may spend extra time trying to derive the meaning and content of the page. Heading changes could be depicted through a change of speech timbre or by adding a sound to the screen-reading voice.

Initial user evaluation of our system has revealed that users would like more extensive and detailed navigational cues through non-speech sounds. Users have indicated that the current multimodal feedback used in the system is effective to gain an overview of the spatial layout of a web page. In order to gain more feedback from visually impaired users, further evaluation sessions will be conducted with larger samples of visually impaired users, with varying levels of sight loss.

5. ACKNOWLEDGMENTS

This project is supported by Eduserv (http://www.eduserv.org.uk).

6. REFERENCES

[1] www.freedomscientific.com

[2] Gaver, W. W., Smith, R., & O'Shea, T., Effective sounds in complex systems: the ARKola simulation. Proceedings of the SIGCHI conference on Human factors in computing systems, New Orleans, Louisiana, USA, 1991, pp. 85-90

[3] Brewster S., Wright P. & Edwards A., Experimentally Derived Guidelines for the Creation of Earcons, Adjunct Proceedings of the British Computer Society Conference on Human Computer Interaction, 1995, pp. 155-159

[4] Goose, S. & Moller, C., “A 3D audio only interactive Web browser: using spatialization to convey hypermedia document structure”, Proceedings of the seventh ACM international conference on Multimedia, 1999, pp. 363-371

[5] Donker, H., Klante, P. & Gorny, P., “The design of auditory user interfaces for blind users”, Proceedings of the second Nordic conference on Human-computer interaction, 2002, pp. 149-156

[6] Susini, P., Vieillard, S., Deruty, E., Smith, B. & Marin, C, “Sound Navigation: Sonified Hyperlinks", Proceedings of the International Conference on Auditory Display ICAD 2002, July 2-5, Kyoto, Japan, 2002

[7] Parente, P. & Bishop, G., “BATS: The Blind Audio Tactile Mapping System”, 41st Annual ACM Southeast Conference, Savannah, Georgia, 2003

[8] Roth, P., Lori Petrucci, Assimacopoulos, André & Pun, T., “Audio-Haptic Internet Browser and Associated Tools for Blind Users and Visually Impaired Computer Users”, COST 254 Intelligent Terminals, Workshop on Friendly Exchanging Through the Net, 2000, pp. 57-62

[9] Petrucci, L.S., Harth, E., Roth, P. & Assimacopoulos, André, Pun, Thierry, "WebSound: a Generic Web Sonification Tool, and its Application to an Auditory Web Browser for Blind and Visually Impaired Users ", Proceedings of the International Conference on Auditory Display ICAD 2000, Atlanta, Georgia, USA April 2-5, 2000

[10] Adcock, M., Barrass, S. “Cultivating Design Patterns for Auditory Displays”, Proceedings of the International Conference on Auditory Display ICAD 2004, Sydney, Australia, July 6-9, 2004

[11] Leplâtre, G. McGregor, I. “How To Tackle Auditory Interface Aesthetics; Discussion And Case Study", Proceedings of the International Conference on Auditory Display ICAD 2004, Sydney, Australia, July 6-9, 2004

[12] www.microsoft.com/speech/download/sdk51/