A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure

Stuart Goose
Multimedia Department
Siemens Corporate Research
Princeton, NJ 08540, USA
Tel: 1-609-734-3391
sgoose@scr.siemens.com

Carsten Möller
Multimedia Department
Siemens Corporate Research
Princeton, NJ 08540, USA
Tel: 1-609-734-6500
cmoeller@scr.siemens.com

ABSTRACT

Interactive audio browsers provide both sighted and visually impaired users with access to the WWW. In addition to the desktop PC, audio browsing technology can be deployed that enable users to browse the WWW using a telephone or while driving a car. This paper describes a new conceptual model of the HTML document structure and its mapping to a 3D audio space. Novel features are discussed that provide information such as: an audio structural survey of the HTML document; accurate positional audio feedback of the source and destination anchors when traversing both inter-and intra-document links; a linguistic progress indicator; the announcement of destination document meta-information as new links are encountered. These new features can improve both the user's comprehension of the HTML document structure and their orientation within it. These factors, in turn, can improve the effectiveness of the browsing experience.

Keywords

3D audio, spatialization, document structure, Hypertext, WWW, browsing.

1. INTRODUCTION

The World Wide Web (WWW) has enjoyed phenomenal growth over recent years and now accounts for a significant proportion of all Internet traffic. The unmitigated success of the WWW bears testimony to the previously unsatisfied need for a system able to integrate and deliver distributed information. The profile of hypermedia has been raised significantly by the WWW, which has endorsed hypermedia as an appropriate technology for accessing and navigating information spaces. Users can access a wealth of information and associated services over the WWW, ranging from international news to local restaurant menus.

At Siemens Corporate Research, one research focus has been the investigation and development of interactive audio browsers that enable users to process their electronic mail and browse the WWW while driving a car [22] and using a telephone [10]. This technology also further democratizes the WWW by providing improved browsing capabilities for the visually impaired community.

ACM Multimedia '99 10/99 Orlando, FL, USA
© 1999 ACM 1-58113-151-8/99/0010...$5.00

This paper describes several innovative ways in which the efficacy of the audio browser can be improved through the judicious application of spatialized, or three-dimensional (3D), audio technology.

A survey of the related work is discussed in section 2. A justification of the importance of document structure is outlined in section 3. An overview of audio spatialization is provided in section 4. Reported in section 5 are some preliminary experiments with 3D audio toolkits. Section 6 discusses the conceptual model and how an HTML document is mapped to the 3D audio space. The new features of the audio browser that exploit the 3D audio technology are presented in Section 7. Section 8 proposes areas for further research and provides some concluding remarks.

2. RELATED WORK

Much hypermedia research has focused on the seamless integration of media within a unified framework. Due to the application scenarios and the delivery devices targeted, our emphasis is exclusively on the audio medium. Little work has been conducted on interactive audio-only hypermedia systems. The Hyperspeech system [1] was the first to demonstrate such an approach. Arons manually transcribed several recorded interviews, analyzed their structure and generated corresponding audio nodes and links. Unlike our system, HyperSpeech [1] requires that documents be pre-recorded in audio prior to use.

To access computer-mediated information blind people, until recently, largely relied upon Braille output devices and software known as *screen readers*. A screen reading program applies various techniques to gain access to the textual content of application software and employs speech synthesis technology to speak this information to the user. This approach allows screen readers to be application software independent, and hence can be used to read the text displayed within a visual WWW browser. In this case the screen reader extracts only the text, as it is not concerned, or aware, of the underlying HTML. As a result, the speech output generated communicates the raw content to the listener but fails to impart any information regarding the structure of the document. The importance of document structure as an aid to understanding is addressed in section 3.

Several researchers have since attempted to address this shortcoming. Asakawa *et al* [3] explains how the Netscape browser supplies Home Page Reader data from which descriptions of the HTML document structure are appropriately interspersed in the speech output generated. Designed specifically for visually impaired computer users, the pwWebSpeak browser [19] parses the HTML document in order to augment the audio rendering with structural descriptions. Djennane [8] describes a system for rewriting structured and semi-structured documents to meet the input schema of a hyper-audio hierarchical browser. In addition, these systems provide a range of features one would expect of a WWW browser. Although the audio browsers reviewed [3, 8, 19] analyze and describe the HTML structure, unlike our system they do not exploit spatialization or simultaneity in their audio rendering or orientation aids.

Although far from perfect, screen readers provided visually impaired people with a tool for hearing the content of the screen until graphical user interfaces (GUIs) became commonplace. The advent of the GUI made the task of screen reading more complex, thus inspiring research into GUIs for the blind [15, 17]. Petrie *et al* [18] have conducted preliminary evaluations on input and output schemes to identify favorable hypermedia system interfaces for blind users. Many of the recommendations have been incorporated into our system.

Ubiquitous computing is the attempt to break away from the traditional desktop interaction paradigm by distributing computational power and resources into the environment surrounding the user. Goose *et al* [10] describe a proxy-based interactive service (DICE) into which users can telephone and use touch tones or voice commands to browse dynamically generated audio renditions of both email and WWW documents. Due to its audio nature and the minimum interaction required to operate it, the car radio was selected as the interface metaphor by Wynblatt *et al* [22] as the basis of an audio browser (WIRE) for providing drivers with access to email and WWW services. An analogy is drawn between selecting a bookmark and selecting a radio channel. Once selected, the driver listens to the audio rendering of the document as if listening to the radio, but with the option to issue voice commands for features such as following a link. The research presented in this paper builds on the foundation of these projects [10, 22].

Sawhney *et al* [20] describes a nomadic application for presenting email, voice and reminder messages. This application employs a clock metaphor and new messages are presented to the user at the position in the 3D audio space corresponding to their time of arrival. The aim of Schmandt *et al* [21] with AudioStreamer is to enhance the effectiveness of simultaneous listening by exploiting the human ability of separating multiple simultaneous audio streams. An interface allows the audio source of greatest importance to the listener to be made more prominent. Kobayashi *et al* [13] describe a system for relating each section of a document to a point on the perimeter of a circle in a spatialized audio interface. This approach allows the human spatial memory to compensate for the weakness of temporal recall. Our system, in common with the systems reviewed, makes extensive use of spatialized audio, but none of these systems are concerned with the interactive browsing of HTML documents. Our system dynamically generates audio renditions of both email and WWW documents, as opposed to [13, 21] which require the documents to be pre-recorded prior to use.

Although much progress has been made in the area of audio browsing, as this literature review confirms, to the authors' knowledge this is the first 3D audio only web browser to be reported. The main contributions of this paper are a new conceptual model of the HTML document structure and its mapping to a 3D audio space; the application of audio spatialization and simultaneous voice streams to a number of novel aids to enhance browsing. These new features can improve the user's comprehension of the HTML document structure, content and their orientation within it.

3. DERIVING STRUCTURE FROM DOCUMENTS

The native document description language of the WWW is called Hypertext Markup Language (HTML). At a quick glance, a sighted user can assimilate the document structure of a richly graphical HTML page as rendered by a visual web browser. This is possible as much of the context is conveyed implicitly through the document structure and layout of the information. A user can then apply their understanding of the HTML document structure to aid orientation, navigation, and ultimately, the location relevant information.

Given that structure is obviously a key aid to the comprehension of a visual document, it is of paramount importance to the user of an audio browser. It is clear that most of the context would be lost if a document were "rendered" by simply sending the raw text of a document to a text-to-speech synthesizer. The telephone and automobile browsers [10, 22] apply an analytical algorithm to an HTML document, or frame set, to elicit both the structure and context.

Although intended to represent document structure, HTML has also evolved to include constructs for visual specifications. Consequently, no clear distinction exists between the document structure and its presentation view. Many authors strive to design aesthetic and intuitive graphical HTML pages. In order to achieve this goal some authors purposefully select alternative HTML constructs to fashion a custom view of the structure of the page, as opposed to using the HTML constructs originally designated for specifying the logical structure. One typical example of this is the selection of a large font to customize the appearance of a section heading in favor of the standard HTML header construct. While entirely legitimate, the algorithm that analyzes the HTML document structure must identify such behavior to determine the author's actual intent.

Once analyzed, an audio rendering is then produced which combines the use of earcons [4, 9] and the features of an underlying text-to-speech synthesizer, such as multiple voices, intonation, announcements and pausing, to make structural elements of the document explicit to the listener. The aesthetics of the audio rendition can simultaneously help reduce the monotony factor and enhance comprehension [12].

4. BRIEF OVERVIEW OF AUDIO SPATIALIZATION

Cognitive psychologists have conducted many experiments to achieve a deeper understanding of the way in which humans interpret through their ears the cacophony of audio signals encountered. There are many cues in the natural environment that facilitate human spatial audio perception. The primary cues are described below:

- **Volume**: the farther away an object is from the listener, the quieter is the sound. This phenomenon is called *roll-off.*

- **Interaural Intensity Difference (IID):** a sound emanating from the listener's right will sound louder in the right ear than in the left ear.
- **Interaural Time Difference (ITD):** a sound emitted by a source to the listener's right will arrive at the right ear approximately one millisecond before it arrives at the left ear.
- **Muffling:** the orientation of the ears ensures that sounds emanating from behind the listener are slightly muffled compared with sounds coming from the front. In addition, if a sound is coming from the right, the sound reaching the left ear will be muffled by the mass of the listener's head.
- **Reverberation:** sound reflections from surfaces are known as reverberation. The listener perceives different effects dependent upon the size and shape of the room and the absorptiveness of the surfaces.

Synthetic sound spatialization is the processing of sound in such a way that when it reaches our ears it reproduces the characteristics of a sound located in a 3D-space external to the listener. The effects described above, and many more, can be modeled by a Head-Related Transfer Function (HRTF) [11]. A digitized mono audio stream can be convoluted with an artificial HRTF to create a stereo audio stream that reproduces the timing, frequency and spectral effects of a genuine spatial sound source. For ideal results an HRTF needs to be tailored to an individual, but a generalized HRTF can still produce pleasing results.

5. PRELIMINARY SPATIALIZATION EXPERIMENTS

Our aim was to use regular PC audio hardware with a commercially available 3D audio software toolkit. Prior to design, we conducted experiments to establish the limitations of the technology and whether it was perceptually feasible to utilize the entire 3D audio space. The 3D audio toolkits that we evaluated were Microsoft DirectSound, Intel RSX and Aureal. All of these audio toolkits by necessity use generalized HRTFs.

We began by examining each axis in isolation. For each axis we played different types of sounds at both static and moving positions. The results of these tests concurred with the literature, and are summarized in table 1.

Axis	Static sound	Moving sound
X	Participants were able to identify accurately the position	Participants were able to identify accurately and track the position
Y	Participants were not able to identify accurately the position	Participants were only able to track the position with a low degree of accuracy
Z	Participants were not able to identify accurately the position	Participants were only able to track the position with a low degree of accuracy

Table 1: Limitations of using the entire 3D space.

The different types of sounds employed did not have any discernable effect on our results.[1]

An initial idea for evaluation was to simulate the notion of reading down a physical page, from top to bottom. Speech synthesis would be employed to speak the document, as described in section 3, but the y-coordinate in the audio space at which it is spoken would be varied to represent the position through the document, akin to reading down a page. The results above confirmed that with the toolkits we selected, humans cannot accurately perceive the position of sound sources in the y-axis. This deficiency was overcome by instead projecting onto the x-axis a moving speech synthesis source. This made it possible to harness the listener's intrinsic ability to track accurately the moving position of a sound source that conveyed simultaneously the current position in the document and its content.

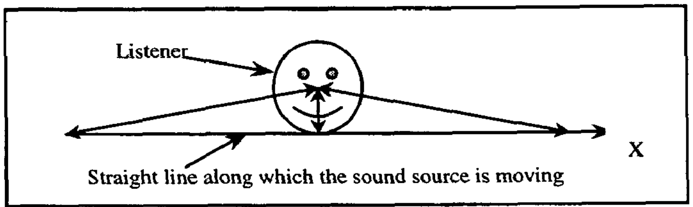

Figure 1: The distance from sound source to the listener's head is not constant.

Although it is possible to migrate a sound along the x-axis in front of the listener's head, the distance from the sound source to the listener's head is not constant, as illustrated in figure 1. *The roll-off* effect causes a sound to be louder in the center and quieter on each side, thus having a negative influence on the listener's ability to identify accurately the position of the sound source. These effects were mitigated by projecting the moving sound source along a semi-circle around the front of the listener's head, as shown in figure 2.

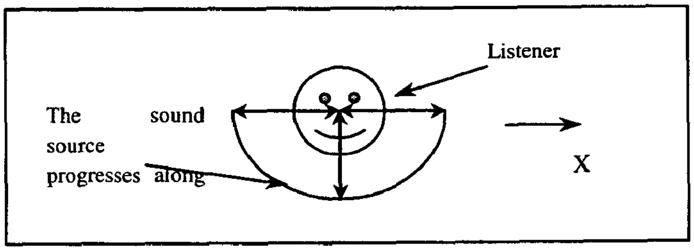

Figure 2: Constant distance from the arc to the listener.

A subsequent experiment highlighted that although the listener is able to track accurately the position of a sound source along the majority of the x-axis, the accuracy is significantly depleted at the extremes of the semi-circle as illustrated in figure 3. The inaccurate regions were eliminated by reducing the semi-circle to that of an arc, called the Stage-Arc, which can be appreciated in figure 4.

As can be seen in figure 5, the listener is positioned in the center while the current position through the document corresponds to the position along the arc from which the document content is spoken. The validity of this approach was evaluated using three documents of differing lengths. The results are summarized in table 2. Although suitable for small documents, these results indicate that this approach did not scale well to arbitrarily large documents.

[1] The participants were equipped with high quality headphones.

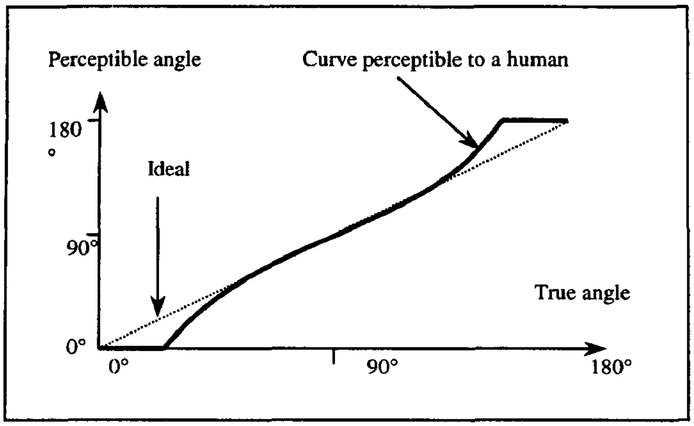

Figure 3: The curve perceptible to a human.

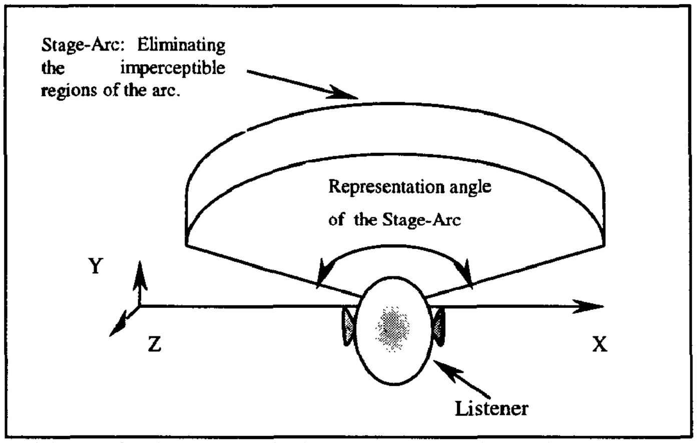
Stage-Arc: Eliminating the imperceptible regions of the arc.
Representation angle of the Stage-Arc
Y
Z
X
Listener

Figure 4: Eliminating the imperceptible regions of the arc.

The "cocktail party effect" [2], the ability of a person to focus their listening attention primarily upon a single talker while surrounded by others also holding noisy conversations, has been acknowledged for some time. Another experiment was conducted to investigate how this human ability could be exploited within the context of this application. The results of the previous experiments led to a refinement of the Stage-Arc. The main voice for speaking the document content now remains at a fixed location at the mid-point of the arc. A second voice was introduced that periodically announces the current position through the document as a percentage. The position along the arc at which the second voice can be heard is relative to the position through the document. Although this simultaneity worked, it was significantly improved upon by:

- employing distinct voices, e.g., a male voice for the document content and a female voice for the percentage position
- increasing the volume of the document content voice and decreasing the volume of the position voice
- raising the y-coordinate of the document content voice to avoid impinging upon the trajectory of the position voice

The use of 3D audio in computer games is becoming increasingly common, but these sound effects are often for atmospherics and seldom to impart precise positional information. Later in the paper we describe how the entire 3D audio space is used effectively to convey additional information and cues for which accurate positional perception is of less importance. Although the results of these preliminary experiments yielded results that dampened our initial ambitions, they provided practical input to the conceptual model of the auditory interface.

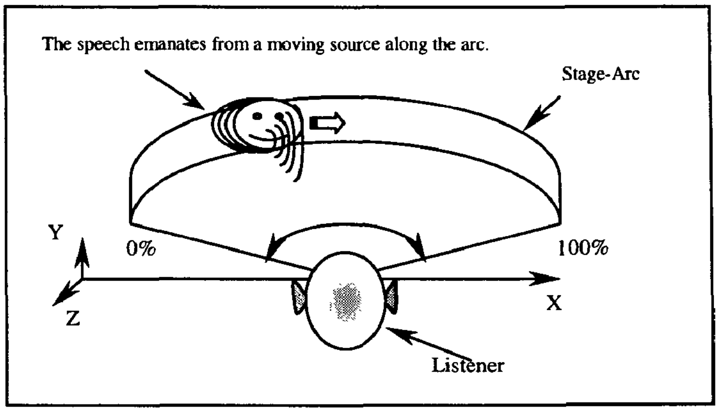

Figure 5: Experimenting with a moving speech source to convey the current position through the document.

Document length	Participant's qualitative feedback
34 seconds	- The speed of the speech position along the arc was fast. - The movement detracted from the ability to comprehend the speech. - Identification of the position of the speech along the arc was easy.
1 minute 40 seconds	- The speed of the speech position along the arc was fine. - It became disturbing to hear the speech for too long in only one ear. - Identification of the position of the speech along the arc was very accurate.
7 minutes	- The speed of the speech position along the arc was very slow. - It proved very unpleasant to hear the speech for too long in only one ear. - Identification of the position of the speech along the arc was poor due to the slow movement. The ears became "tired" continuously attempting to interpret this cue.

Table 2: Evaluating the suitability of a moving speech source.

6. CONCEPTUAL MODEL OF AUDITORY INTERFACE

Our preliminary experiments proved the ineffectiveness of the y-axis in the audio space of conveying an accurate position cue to the listener to simulate reading down a page. By instead projecting the document onto the x-axis an effective solution has been identified, which affords accurate positional cues to the listener.

Once parsed and analyzed, the audio browser commences the sequential acoustic projection of the richly structured hypermedia document onto the Stage-Arc. Three distinct male synthesized voices are designated respectively to speak the headers, content and hypermedia link anchors. These three voices emanate from the static position in the center of the Stage-Arc. At periodic intervals the percentage position through the document is announced by a synthesized female voice located at the relative position along the arc. When a hypermedia link is encountered in the audio rendering an earcon is sounded at the relative position along the Stage-Arc. The mapping of the various structural elements in the HTML document to the Stage-Arc can be viewed in figure 6.

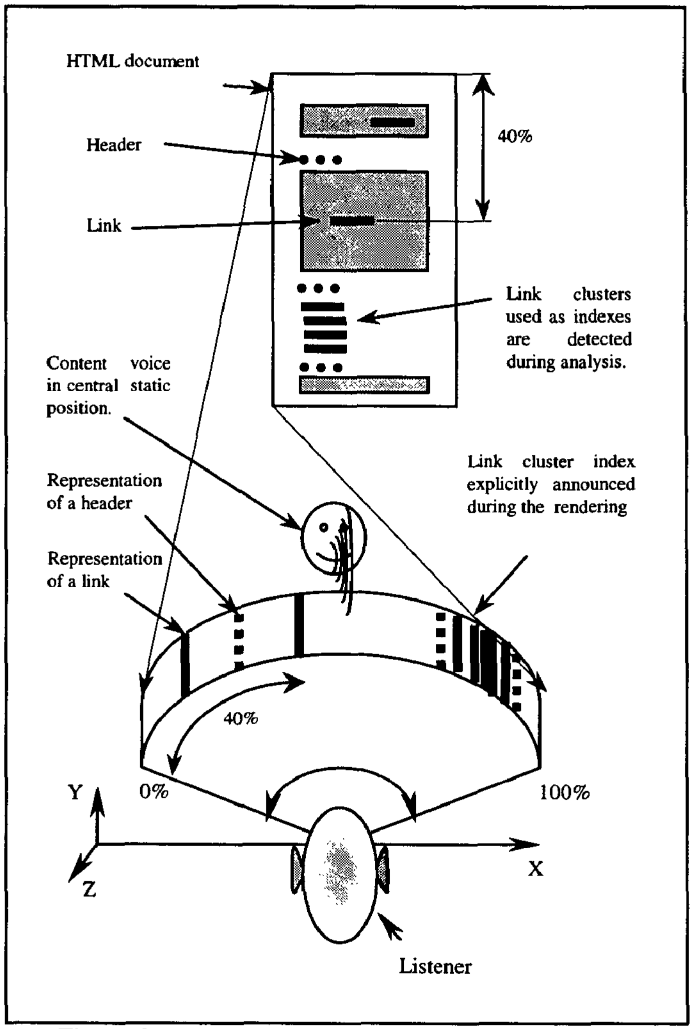
HTML document
Header
Link
40%
Link clusters used as indexes are detected during analysis.
Content voice in central static position.
Representation of a header
Representation of a link
Link cluster index explicitly announced during the rendering
40%
Y
0%
100%
Z
X
Listener

Figure 6: Mapping an HTML document to the Stage-Arc.

7. ENHANCED BROWSING WITH 3D AUDIO

The conceptual model of the auditory interface described above provides an infrastructure upon which a number of novel browsing aids have been layered. A description of these aids is given in this section.

7.1 Link Traversal

An essential ingredient of hypermedia documents is the link, and, in the context of the WWW, a link can either point to another place within the same document (intra-document link) or another document entirely (inter-document link). Petrie *et al* [18] note that visually impaired users can become disorientated during navigation without a mechanism for disambiguating these two link types. Moreover, Landow [14] advocates the use of a "rhetoric of arrival and departure" when authoring hypermedia documents for mitigating the effects of disorientation during navigation. Solutions to these issues using spatialized audio are described below.

In order to navigate through a hypermedia web the user must be cognizant of the links. Once aware of the convention, empirical tests indicated that the combination of a distinct earcon, followed by a special voice reserved for announcing the anchor text, enabled users to identify the presence of links correctly every time. The audio browser reserves two sonically related earcons for link notification thus enabling the listener to distinguish easily between the two link types.

When an intra-document link is traversed, three transitional sounds are seamlessly combined to convey the impression of being catapulted at low altitude to a specific destination location. A *take-off sound* rises a short distance in the y-axis above the source of the link. A *flying sound* travels along the Stage-Arc to the link destination position. The effect is completed by a *landing sound*, which descends the short distance to the original height on the y-axis. This sequence is illustrated in figure 7.

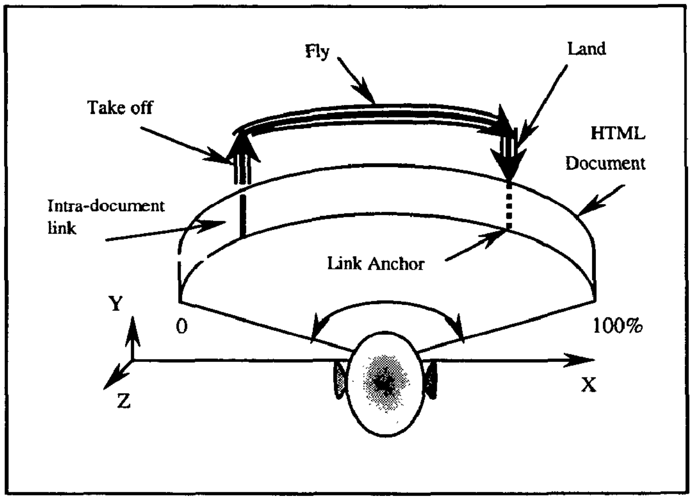

Figure 7: 3D audio intra-document link traversal.

A seamless combination of transitional sounds is also used when an inter-document link is traversed. The effect created by these sounds is that of a spaceship being launched high into orbit and then descending to land on another planet. A "beaming" sound that oscillates and rises in pitch proved to be easily associated by listeners with that of a *launching sound.* In parallel to the rising pitch, the sound rises in the y-axis above the source of the link. The decrease in its volume is attributable to the *roll-off* factor as the sound source gains altitude. This effect is visualized in Figure 8a. Upon retrieval of the destination document, the listener hears the inverse of the oscillating "beaming" sound now with a falling pitch that descends in the y-axis to represent a *landing sound.* The sound approaches the listener with increasing volume, again due to the *roll-off* factor. A small explosion is sounded to signify touchdown. Although it is possible to specify an anchor in a destination document as part of a URL link, it is most common for links to point to the beginning of the destination document. The position on the Stage-Arc at which the touchdown occurs is dependent upon the anchor position within the destination document, as illustrated in figure 8b.

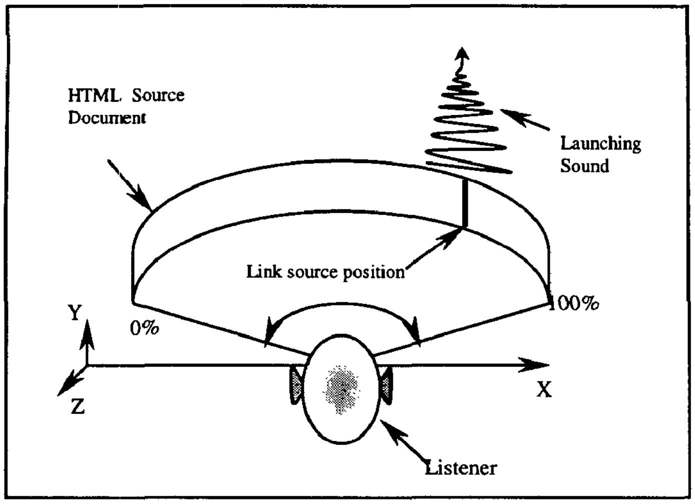

Figure 8a: 3D audio inter-document link traversal source.

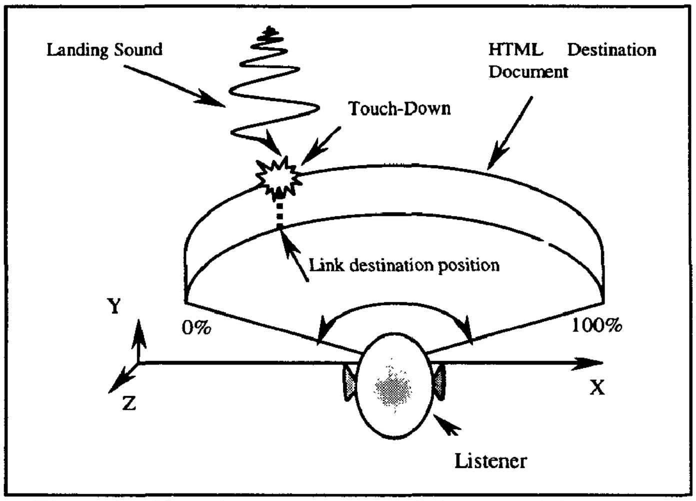

Figure 8b: 3D audio inter-document link traversal destination.

Sighted users can establish their current position in a document by checking the scroll bar offset in the browser. Brewster *et al* [5] argue that the use of an auditory scrollbar is of significant benefit, providing useful sonic feedback and alleviating the user from regularly glancing at the scrollbar thumb position. Although the two link traversal aids described here are not auditory scrollbars, they deliver similar positional information. Brewster *et al* employ an ascending or descending musical scale to provide an impression of scrolling. A user with a very good "musical ear" is necessary for this technique to be useful in imparting precise positional information. Our approach for conveying the position in the document exploits the listener's ability to accurately identify sound sources along the x-axis. Results offered by Oldfield *et al* [16] indicate that humans can identify the location of a sound source along the x-axis to within nine degrees.

7.2 Document Sound Survey

A classic and well-recognized problem with hypermedia is that of users becoming disorientated during navigation, or "lost in hyperspace." This problem is exacerbated when navigating hypermedia documents using an audio browser, as the user can soon become lost even within a single document. During the act of browsing, the user is often seeking new avenues of investigation via the links, and link clusters, in addition to assimilating information. Improved support for this activity through the use of spatialized audio is described below.

As audio is a serial medium, it is difficult with audio interfaces, unlike their graphical counterparts, to present persistent information without the risk of being intrusive and annoying. The problem faced by the user is of how to identify rapidly the presence of interesting structural elements in their vicinity of the document. For example, the user may wish to know if any links exist within the next few sentences, or even to recall the positions of links that have recently been heard. Our solution to this problem is to generate a 3D "sound survey" of the area surrounding the user's current position in the document. The user can request a sound survey at any time, but, in addition, the audio browser can be configured to initiate automatically a sound survey on completion of a link traversal. After a link traversal the user might well appreciate a brief survey of the surrounding region to aid orientation and provide additional structural context.

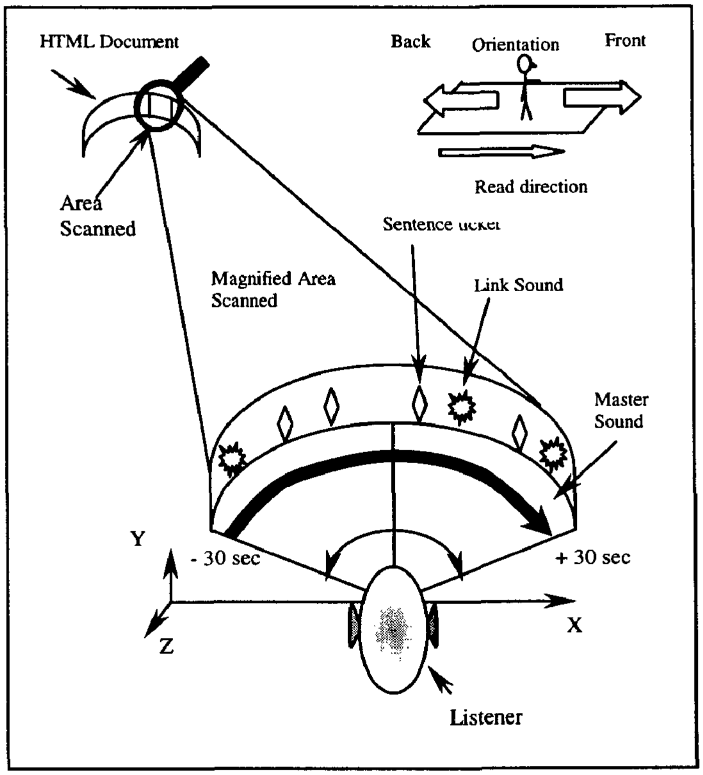

Figure 9: 3D sound survey of a region of the document.

The user can configure the audio browser to specify, in units of time, how far in both directions the sound survey should extend. From figure 9 it can be seen that when this feature is invoked the immediate vicinity (in this example defined to be ± 30 seconds) of the current position is magnified and projected onto the entire Stage-Arc. This allows for a more accurate rendering of the sound survey. The sound survey is performed in faster than real time, for example, allowing a 60 second region to be heard in 15 seconds.

Although many of the structural elements in an HTML document are candidates for featuring in the sound survey, currently we identify only the presence of links. Four distinct earcons are employed in the sound survey, thus allowing both position and context information to be conveyed, succinctly and in parallel, without compromising meaning [6]. The four sounds are:

- **Time ticker:** to signify time moving by, a "tick" sound can be heard moving along the Stage-Arc from left to right. Every fifth tick is emphasized to provide a coarser granularity of time, with which the listener can identify and synchronize.
- **Link indicators:** two earcons representing intra-document links and inter-document links are sounded at the appropriate positions along the Stage-Arc.
- **Sentence boundary:** another sound is used to denote each sentence boundary that is encountered. The listener can use this feature to identify the links and their relative distances as measured in sentences.

If the sound survey was manually selected by the user, as opposed to being triggered as the result of a link traversal, then as the survey sweeps over the first half of the arc this should reaffirm the structural elements recently heard. The second half of the survey introduces what is yet to be heard. As the sound survey sweeps from left to right, the relative volume of the link and sentence earcons is increased and decreased to simulate the relative distance from the current position of the user.

Crease *et al* [7] describe an audio progress indicator able to combine four earcons in parallel to denote current status. An ascending musical scale in conjunction with tempo variation provides specific feedback. The sound survey has similarities to the audio progress indicator, but is also able to exploit the ability of the listener to identify accurately the position of sound sources along the x-axis.

7.3 Audio Preview of Link Destination Document

The use of thumbnail images and moving icons has become commonplace for providing visual abstracts of graphical documents. The rationale being that a user can gain an appreciation of the content and evaluate whether the original document should be downloaded and browsed. An equivalent abstraction mechanism that uses audio spatialization is described below.

A solution was sought which, when a link was encountered as the document was being rendered, could convey to the listener succinct key information concerning the destination document. An automatically generated document summary would undoubtedly be too verbose and distracting, but some basic meta-information of value to a listener about a destination document might be the title and the time required for listening. A multithreaded component was implemented that pre-fetched every link destination document, extracted the title and computed the time required for listening. An additional synthesized voice was introduced to announce the title and the time duration when a link was encountered during the document rendering. As links are often positioned adjacent to one another, a technique was required for delivering this meta-information without creating a cacophony of overlapping and disruptive announcements. A variety of approaches to this problem were evaluated with the most successful illustrated in figure 10.

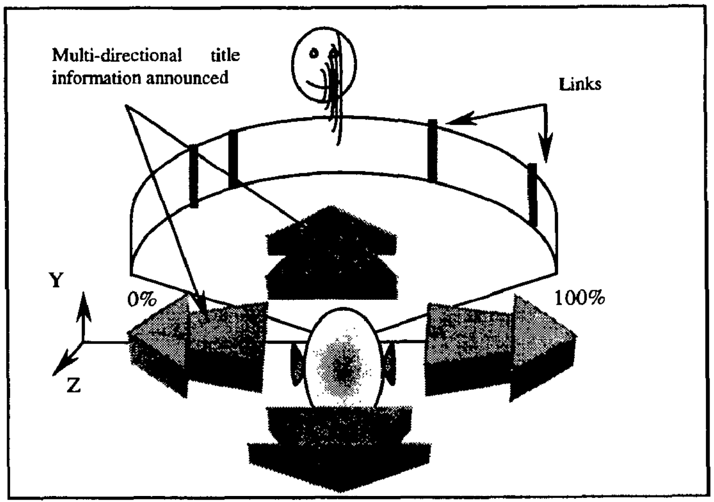

Figure 10: Multiple trajectories used to announce destination document meta-information.

As each link in turn is encountered, the destination document meta-information is announced at one of the four positions, in rotation, in the 3D audio space as indicated by the large arrows. For a couple of seconds the announcement remains at a constant volume before the position of the voice travels on a trajectory away from the listener in the direction of the arrow. This trajectory causes the relative volume to decrease due to the *roll-off* factor. Spatial overlap is avoided as the meta-information is announced at the position of one of the four large arrows in rotation, while the "cocktail party effect" explains the ability of the listener to switch and selectively attend to the most relevant voice stream.

7.4 Structural Navigation

To alleviate the user from listening to every document in its entirety, a selection of rendering modes are supplied. In addition to the entire document being rendered, a second mode announces each link anchor; useful if a document is frequently used as an index to subsequent documents. A third mode announces only the section headings; a convenient mechanism for rapidly scanning a document. A fourth mode skips over structural descriptions and renders the content. As the browsing mode can be changed dynamically, a user can combine these approaches to navigate more efficiently within and across document boundaries. Although these four modes continue to prove useful, a more interactive approach, able to facilitate a finer degree of control over the structured browsing of HTML, was sought.

The user can now select to browse the document in terms of:

- **Links:** once this mode is selected, the user can skip both forward and backward through the document listening to the text of the link anchors being announced. The opportunity to traverse the link is available.
- **Grammar constructs:** for this mode the concept of *granularity* was introduced. The user can select whether they wish to travel around the document in terms of words, sentences or paragraphs.

- **Section headings:** when using this mode, the audio browser can be considered as an audio equivalent to a folding editor. This allows the user to conceptualize the document as a collection of lanes around which they can maneuver, as visualized in figure 11. This mechanism allows a rapid overview of the document to be gained by skipping across the top-level section headings and hearing them announced. Alternatively, if a specific section captures the attention of the listener, they are completely at liberty to *dive* down and hear the sub-section headings announced, and so on to the content within.

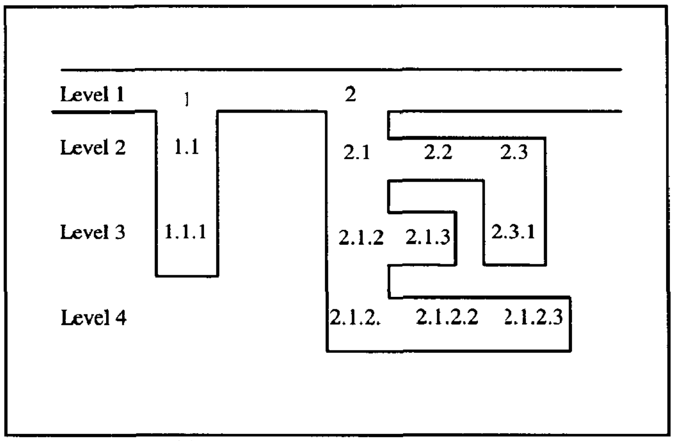

Figure 11: Navigating the document structure.

When used in conjunction with the sound survey, this improved structural navigation can be particularly useful when advancing to the desired entity in the document. An additional related feature is that the user can, at any juncture, request the time required to listen to the current unit of granularity selected.

8. FUTURE WORK AND CONCLUSIONS

Although the technique described for delivering the audio previews of link destination documents works well, new approaches able to impart more information than just the title and the time duration would be valuable. Embellishing the sound survey with additional key constructs may also yield benefits. Iterative informal studies were conducted as an integral aspect of this work, but a more comprehensive user study would undoubtedly provide valuable insight.

A new conceptual model of the HTML document structure and its mapping to a 3D audio space is reported in this paper. It has been discussed how this auditory infrastructure has been augmented with a number of innovative browsing aids. These aids include: an audio structural survey of the HTML document; accurate positional audio feedback of the source and destination anchors when traversing both inter-and intra-document links; a linguistic progress indicator; the announcement of destination document meta-information as new links are encountered. Early results indicate that these new features can improve both the user's comprehension of the HTML document structure and their orientation within it. These factors, in turn, can improve the effectiveness of the browsing experience for visually impaired and sighted users alike.

9. REFERENCES

[1] Arons, B., Hyperspeech: Navigating in Speech-Only Hypermedia, Proceedings of the ACM International Conference on Hypertext, pages 133-146, December 1991.

[2] Arons, B., A Review of the Cocktail Party Effect, Journal of the American Voice I/O Society 12, pages 35-50, July 1992.

[3] Asakawa, C. and Itoh, T., User Interface of a Home Page Reader, Proceedings of the ACM Conference on Assistive Technologies (ASSETS), Marina del Rey, USA, 1998.

[4] Blattner, M., Sumikawa, D., and Greenberg, R., Earcons and Icons: Their Structure and Common Design Principles, Human-Computer Interaction, 4(1), pages 11-44, 1989.

[5] Brewster, S., Wright, P. and Edwards, A., The Design and Evaluation of an Auditory Scrollbar, Proceedings of the ACM International Conference on Computer Human Interaction, Boston, USA, April 1994.

[6] Brewster, S., Wright, P. and Edwards, A., Parallel Earcons: Reducing the Length of Audio Messages, International Journal of Human-Computer Studies, 43(2), pages 153-175, 1995.

[7] Crease, M. and Brewster, S., Making Progress with Sounds: The Design and Evaluation of an Audio Progress Bar, Proceedings of the International Conference on Auditory Display (ICAD), November 1998.

[8] Djennane, S., HTML Document Integration in a Hyperbase and New Access Methods for Web Navigation, Proceedings of the ACM International Workshop on Open Hypermedia Systems, pages 1-1, April 1997.

[9] Gaver, W., Auditory Icons: Using Sound in Computer Interfaces. Human Computer Interaction, 2(2), pages 167-177, 1986.

[10] Goose, S., Wynblatt, M. and Mollenhauer, H., 1-800-Hypertext: Browsing Hypertext with a Telephone, Proceedings of the ACM International Conference on Hypertext, Pittsburgh, USA, pages 287-288, June 1998.

[11] HRTF Measurements of a KEMAR Dummy-Head Microphone, http://sound.media.mit.edu/KEMAR.html

[12] James, F., Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext, Proceedings of the International Conference on Auditory Display (ICAD), pages. 97-103, November 1997.

[13] Kobayashi, K. and Schmandt, C., Dynamic Soundscape: Mapping Time to Space For Audio Browsing, Proceedings of the ACM International Conference on Computer Human Interaction, Atlanta, USA, March 1997.

[14] Landow, G., The Rhetoric of Hypermedia: Some Rules for Authors, Hypermedia and Literary Studies, MIT Press, Cambridge, 1991.

[15] Mynatt, E. and Edwards, W., The Mercator Environment: A Non-visual Interface to the X Window System, Technical Report GIT-GVU-92-05, February 1992.

[16] Oldfield, S. and Parker, S., Acuity of Sound Localization: A Topography of Auditory Space. I. Normal Hearing Conditions, Perception, 13, pages 581-600, 1984.

[17] Petrie, H. and Crispien, K., Providing Access to GUIs for Blind People: Using a Multimedia System - Based on Spatial Audio Presentation, Proceedings of 95th Convention of the Audio Engineering Society, New York, 1993.

[18] Petrie, H., Morley, S., McNally, P, O'Neill, A and Majoe, D., Initial Design and Evaluation of an Interface to Hypermedia System for Blind Users, Proceedings of the ACM International Conference on Hypertext, Southampton, UK, pages 48-56, April 1996.

[19] Productivity Works Inc, http://www.prodworks.com

[20] Sawhney, N. and Schmandt, C., Design of Spatialized Nomadic Environments, Proceedings of the International Conference on Auditory Display (ICAD), pages 109-113, November 1997.

[21] Schmandt, C. and Mullins, A., AudioStreamer: Exploiting Simultaneity for Listening, Proceedings of the ACM International Conference on Computer Human Interaction, Denver, USA, May 1995.

[22] Wynblatt, M., Benson, D., and Hsu, A., Browsing the World Wide Web in a Non-Visual Environment, Proceedings of the International Conference on Auditory Display (ICAD), pages 135-138, November 1997.