Browsing the World Wide Web in a Non-Visual Environment

Michael Wynblatt, Dan Benson, Arding Hsu
Multimedia / Video Technology Department
Siemens Corporate Research
755 College Road East
Princeton, New Jersey 08540
{wynblatt, dbenson, *ahsu}@scr.siemens.com*

ABSTRACT
We have been investigating building a non-visual browsing environment for the WWW, specifically one that relies heavily on audio for information delivery and feedback. The idea is that with an unobtrusive browser, a user may listen to the WWW in the background, just as many people now listen to the radio while doing other tasks. This paper discusses two of the major issues which are faced in such a system, how to render HTML documents using audio, and how to provide a non-visual browsing interface for the WWW. We have implemented a prototype system as part of the *Web-based Interactive Radio Environment*, which includes our approaches to these issues.

MOTIVATIONS AND RELATED WORK
The World Wide Web (WWW) is rapidly becoming the single most important source of information for businesses and consumers. With any source of information, a trade-off is involved between the value of information discovered and the opportunity cost of the time spent discovering it. Recent advances in technology, like the cellular phone, have helped people to improve the result of this trade-off, allowing them to better utilize time that would otherwise be unproductive, such as the time spent commuting to work, or exercising. The WWW, however, is difficult to use in such situations, because existing WWW browsers require significant visual attention and a high degree of interactivity.

At Siemens, we have been investigating a solution to this problem by building a non-visual browsing environment for the WWW, specifically one that relies heavily on audio for information delivery and feedback. The idea is that with an unobtrusive browser, a user may listen to information (or entertainment) from the web in the background, just as many people now listen to the radio while doing other tasks. This paper discusses two of the major issues which are faced in such a system, how to render HTML documents using audio, and how to provide a non-visual browsing interface for the WWW. We have implemented a prototype system as part of the *Web-based Interactive Radio Environment* (WIRE) which includes our approaches to these issues[1].

There have been several proposals to provide WWW access in limited visual environments, notably the HDML language proposed by Unwired Planet [7]. These proposals support their target applications, but still make the assumption of a visual representation (albeit a small one), and keypad entry capability. This kind of interface is not suitable for truly non-visual applications such as a browser for automobile drivers, visually impaired users, or screenless phones.

There have also been several attempts to provide audio representations of WWW documents. A notable early attempt, WebOnCall by NetPhonic [4], offers telephone access to WWW sites rendered using audio. However, the WebOnCall system requires WWW documents to have special preparation on the server side, and therefore users may only access subscribing web sites. As a result, browsing of the WWW is not really supported, since only a small number of sites actually make the effort to provide the service. A more general solution is to provide a client-side system, so that any existing web site can be accessed.

The Productivity Works has developed WebSpeak [5], an audio output browser intended for use by visually impaired users. This browser, however, is not suited for passive browsing as its interface requires a very hands-on approach: the user must advance the rendering manually from item to item. The interface is also very keyboard oriented. Neither of these properties is appropriate for the automobile environment we are targeting. Moreover, although text-to-speech synthesis is provided for text within HTML documents, there is little attempt to render the HTML markup; most HTML tags are simply announced, with

[1] US Patent Pending, Application Serial Number 08/768,046

the resulting rendition left to the user to interpret mentally. A preferred approach would be to actually interpret the HTML, and render the document in an audio-oriented way.

In the time since our prototype was developed, work by James [1,2] has employed "Wizard of Oz" experimentation to try to determine what kinds of HTML audio might be preferred by novice users. We have not yet performed a formal user study with the WIRE rendering and interface, but the informal observations of our experimentation with the system support several of James' findings, which we will mention during the discussion.

RENDERING HTML USING AUDIO

Some data on the WWW is already provided in audio form, much of it in the form of Progressive Network's popular RealAudio streamable format [6]. In an audio-only rendition, this data can be rendered directly. Most WWW data, however, is in the form of HTML documents. HTML was originally intended to be primarily independent of presentation, stressing a document's abstract structure. It has evolved, however, to include many tags which make explicit visual specifications, and these tags enjoy particularly wide use since they lead to more polished looking WWW documents. This section discusses how the WIRE system generates an audio rendering for documents that are tightly coupled to an explicit visual representation.

One of the difficulties with HTML documents is that much information is conveyed implicitly through layout. A group of links that allow the user to navigate throughout a site might be clustered together in a corner or in a side panel. A quick glance tells the user the common purpose of those links. Dividing bars, cell backgrounds, headings, blank spaces and margins alert the user to conceptually separate items on a page through the use of *context*. Although it is a fairly straightforward process to "render" a document by sending the text passages to a text-to-speech synthesizer, most of the context is lost if the text is simply read in this way. In order to create an audio representation of an HTML document, it is necessary to convey some of this context to the user along with the rendered text.

One way to convey visual context is to break the audio rendition of the document into sections. The sections are based on the visual boundaries within the document, as determined by an analysis of the document specification. Often, a heading for a section can be determined by looking for a few words with a larger font at the beginning of the section. During rendering, the section boundaries can be used in several ways. First, sections can be announced to the user to give the context of the position within the document, as in "Section 2.1 … Local Events". Second, announcements can be made to describe the content of the section. WIRE distinguishes between "navigation sections" and "content sections" based on the *link density* [8] of the section. This technique allows the user to understand whether the section is mainly a link menu or contains some text as well, a distinction that would be clear in a visual representation. Finally, sectioning is useful in enhanced navigation techniques as described in Section 3.

Other kinds of announcements, such as the verbal enumeration of list elements, and the explicit identification of titles and captions, can also help to convey context to the user.

Another technique to convey context is the use of audio effects. Our experience supports James' findings that most users prefer a rendering with a minimal use of traditional "sound effects" like beeps and tones. Two effects that we have found useful are to vary the speaking voice of the speech synthesizer and to introduce pauses into the rendition. Varying the speaking voice can convey a difference in the function of the text content. Titles, headings and structural announcements, for example can all be rendered in a different voice than the main text, to distinguish them, and to give the user the context of their use. Since not all HTML authors use the more abstract `<h>` tags to indicate headings, a certain amount of engineering work goes into successfully recognizing headers. For text which appears inline with normal text, but from which the specification requires emphasis, the standard speaking voice can be varied subtly depending on parameters offered by the speech synthesizer. Pauses help to make clear separation between text blocks, and to emphasize speech immediately following them.

An especially important feature of an HTML documents are its hyperlink anchors. In order for a user to navigate through a hypermedia web, she must be able to recognize the occurrence of links. Reserving a special voice for hyperlink anchors helps significantly in this regard. James found that users were distracted by voice changes within a text passage [2]; indeed, the abruptness of a speaker change within a passage seems ideal for alerting an otherwise occupied user, such as an automobile driver, to the presence of a hyperlink anchor. We have found that the addition of a sound effect, such as a simple bell, along with the change in voice, and a slight pause before the anchor, enables the user to identify hyperlinks correctly essentially all of the time. This is true even for novice users once the convention has been explained.

AN AUDIO-TACTILE BROWSING INTERFACE

Another challenge with a non-visual WWW browser is to provide a browsing interface that can be used in the absence of a computer monitor or keyboard. Although "non-visual" does not necessarily imply "non-keyboard", our applications for occupied users (e.g., driving a car, exercising, writing at a desk, etc.) effectively require this restriction. Traditional WWW browsers use mainly GUI input and feedback, including such conventions as following a link by clicking it, using pull-down menus for favorites and history lists, entering search terms and URLs via keyboard, and using scrolling bars for moving within a single document. The WIRE system has an interface using physical inputs (buttons and knobs) and audio feedback, which allows equivalent functionality, and some extra features suited for the environment. Following the technique suggested by Mynatt [3], we modeled the traditional interface components without modeling their visual behavior.

The simplest feature of the interface is a set of *favorite buttons*. Just like bookmarks in a traditional browser or presets on a car radio, these buttons allow the user to jump immediately to a WWW document that they have selected beforehand. The mappings between documents and buttons are stored at a friendly remote site, which allows the buttons to be programmed just as easily from home or office if a relevant page is found.

Once a page is reached, it is rendered as described in Section 2. WIRE uses a technique called *active link* to allow the user to follow hyperlinks in a document. In this technique, the system renders hyperlink anchors in a notable way, as described in Section 2. As each link anchor is played, it becomes the active link. The user may follow the active link at any time by pressing the follow button. The active link remains active until a new link anchor is rendered.

In many circumstances, a user may wish to replay part of the document. This might especially be true if a hyperlink anchor that the user would like to follow was "missed", that is, was replaced by a new active link. To this end, WIRE features a *rewind* button, which allows the user to scan backwards through the document. *Pause* and *fast-forward* buttons are also provided, and together these three features allow the user to navigate through a single document in the familiar way that they might navigate through an audiotape. Since moving through an HTML document is a discrete process, the fast forward and rewind buttons produce a beep for each unit that they pass. In this way, the user is able to gauge their progress by the number of beeps.

Another important part of browsing is keeping track of where you have been and being able to get back. The *history list* is a common tool toward this end. WIRE has an analogous *history knob*, which allows the user to dial back to the previous pages; as the knob is turned, a beep is provided for each step back in the list that has been made. Since the first part of a document to be rendered is always the title, users can stop at any point in the dialing and quickly reorient themselves. The knob can be turned in the opposite direction to the forward feature offered by traditional browsers.

When browsing the WWW, a user does not always wish to read the entirety of every document they access. WIRE offers a number of features to help them to reach the relevant material more quickly, and the most important of those is that several different *browsing modes* are offered. It has been assumed in the preceding discussion that an entire WWW document is rendered, and complete rendering is the first browsing mode offered by WIRE. The second browsing mode is called navigation mode. The user might choose this mode when they are using a document as a stepping stone to reach another document. In navigation mode, only hyperlink anchors are rendered. The third mode in called content mode. On many web sites, every single page begins with a long set of links to help you reach other parts of the site. Although these are useful, many times the user is more interested in the content of the page itself. Content mode allows the user to skip over any navigation sections that were identified during the document analysis phase; the navigation sections are announced, but not rendered. The final mode is headline mode, in which only section titles are rendered. This mode can be useful for scanning quickly through a page for the main ideas. The user can change modes at any time.

AN EXAMPLE

Figure 1 shows a traditional rendering of a webpage. The WIRE system would begin its rendering with an analysis of the document. The document would be automatically sectioned into 6 sections, corresponding to regions on the page, as is also shown in Figure 1. WIRE would identify the topmost section, and the "Cool Links" section as navigation sections, that is, sections containing primarily links. During rendering, all sections would be announced by number and name, if a name for the section could be identified. For example, the bottommost section would simply be "Section 6", while the second section from the top would be rendered as "Section 2: Contact". During the rendering, the two navigation sections would be announced as such, for example as "Navigation Section with 6 links", in addition to the normal section announcement. Browsing in content mode would allow the user to skip over these sections, although their presence would still be announced. Browsing in navigation mode would cause the user to jump from section 1 directly to section 5, skipping the content in between. Browsing in headline mode would result in only the names of the sections being read. The page title, section titles,

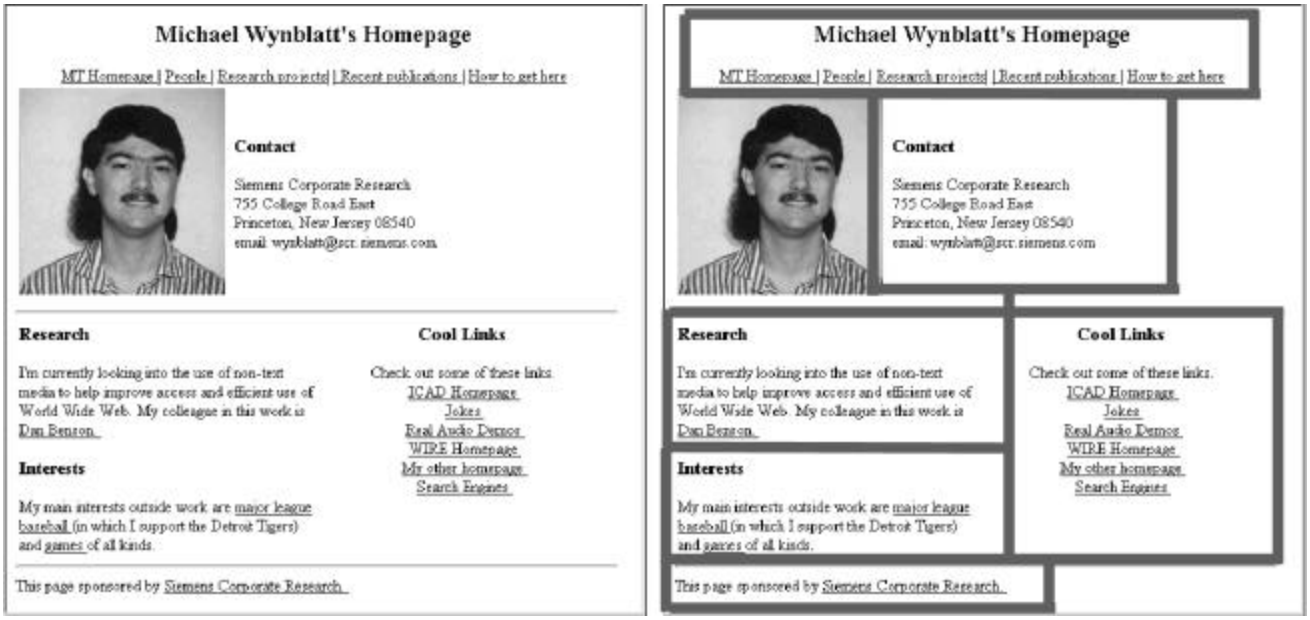

Figure 1. An example webpage rendered traditionally (left) and as segmented by WIRE (right).

and other meta-information such as the presence of navigation sections would be announced by one voice, hyperlink anchors by a second voice, and other text in a third voice.

CONCLUSIONS

The WIRE system developed at Siemens Corporate Research offers access to the World Wide Web in environments where the user is generally occupied with another task. Just as one might listen to the radio or a recorded program, one may now listen to the Web. As part of this system, we have developed a process for rendering HTML documents using audio, and a WWW document browser which does not require a computer screen or keyboard but only a simple set of buttons and knobs. A tool like this, which allows Internet access in an automobile or access to those engaged in other activities, offers to integrate the information-rich Internet even more into our daily activities.

Our continuing work in this area involves further improvements to our rendering and browsing techniques, and attempts to integrate WIRE with other automobile information services such as GPS, radio broadcasts, and cellular phones.

REFERENCES

[1] James, F. *Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext*. ICAD '96 Proceedings, November 1996. pp. 97-103.

[2] James, F. *Presenting HTML Structure in Audio: Stanford Digital Libraries Working Paper* SIDL-WP-1996-0046, 1996.

[3] Mynatt, E., *Auditory Presentation of Graphical User Interfaces*, in Auditory Display Sonification, Audification, and Auditory Interfaces, G. Kramer Ed., 1994 pp. 533-555.

[4] Netphonic Communications, Inc., *Web On Call Product Information*, http://www.netphonic.com/product/woc/wocprod.htm

[5] The Productivity Works, Inc., *The pwWebSpeak Project*, April 1997, http://www.prodworks.com/pwwebspk.htm

[6] Progressive Networks, Inc., *RealAudio Home Page*, http://www.realaudio.com/

[7] Unwired Planet, Inc., *Proposal for a Handheld Device Markup Language*. May 1997, http://www.unwiredplanet.com/pub/hdml_w3c/hdml_proposal.html

[8] Wynblatt, M. and Benson, D., *Web Page Caricatures: Visual Summaries for WWW Documents*. Submitted to 7th International WWW Conference, 1997.