WebSound: a generic Web sonification tool, and its application to an auditory Web browser for blind and visually impaired users

Lori Stefano Petrucci[1], Eric Harth[1], Patrick Roth[1], André Assimacopoulos[2], Thierry Pun[1]

[1] Computer Science Dept.
CUI, University of Geneva
CH – 1211 Geneva 4, Switzerland
+41 22 705 76 32
Lori.Petrucci@cui.unige.ch

[2] UCBA
Swiss Central Union of and for the Blind
Schützengasse 4, CH- 9000 St.-Gallen,
Switzerland

ABSTRACT

The inherent visual nature of Internet browsers makes the Web inaccessible to the visually impaired. Although several non-visual browsers have been developed, they usually transform the visual content of HTML documents into textual information only, that can be rendered by a text-to-speech converter or a Braille device. The loss of spatial layout and of textual attributes should be avoided since they often bear visually important information. Moreover, typical non-visual Internet browsers do not allow visually impaired and sighted individuals to easily work together using the same environment. This paper presents *WebSound*, a new generic Web sonification tool and its application to a 3D audio augmented Internet browser (Internet Explorer 5.0) developed by the Computer Vision Group at the University of Geneva.

Keywords

Audio-haptic interface, non-sound speech, 3D virtual audio environment, earcons, blindness and visual impairment.

INTRODUCTION

In the past, computer interfaces mostly consisted of simple raw-text displays of information. Providing simple access solutions to blind and visually impaired computer users was then a straightforward task, since the use of text-to-speech output devices were sufficient to restitute the displayed information.

Nevertheless, the generalization of Graphical User Interfaces (GUIs) and of the World Wide Web in particular which heavily rely on the visual presentation of information, leads to new accessibility problems for users with visual deficiencies. In fact, more and more digital information becomes available daily on the net and can be accessed by everyone and everywhere. Hypertext documents – HTML in particular and probably XML in the future – have become the most used support to present such information. However, the generalization of visual documents has modified the way information should be restituted to blind users. In addition to textual information, it is also important to take into account textual attributes such as boldface, italic, underline, color, size, as well as the spatial layout itself relating the elements that appear on the screen. Those new problems have to be solved with new alternative non-visual display techniques in order to permit blind users to take advantage of new technologies.

A preliminary study done at the University of Geneva has shown that the use of non-speech sound in graphical interfaces, particularly for an Internet browser, can considerably increase the "bandwidth" of computer output [1,2]. This is also shown in the work of Stephen A. Brewster [3,4] which provides guidelines for incorporating earcons into the design of human-computer interfaces, as well as in the studies of Frankie James [5]. In a first project, named AB-Web, we have developed a prototype of an augmented audio Web browser which transforms the two dimensional visual representation of an HTML document into a 3D immersive virtual audio navigable environment (soundscape) [6].

The need for a generic Web sonification tool which allows to easily create, test and validate new sonification models has become evident. For this reason, we have developed such a tool, WebSound, which combines the haptic sense with an audio output. Our approach tries to validate the hypothesis that a 3D immersive virtual sound navigable environment combined with haptic manipulation of the audio environment can enable blind user to construct a mental representation of the spatial document.

USING AN AUDIO-HAPTIC INTERFACE TO CONVEY VISUAL AND SPATIAL INFORMATION

The visual channel has a tremendous capacity for information transfer. As a result, it is heavily relied upon in normal human-computer interfaces. Internet browsers (for example Internet Explorer or Netscape Navigator) do not escape from this consideration. Most of the information that is presented, however, is not visual in nature, and could easily be presented in an

alternate verbal form. This consideration is certainly the basis for non-visual Internet browsers (for example IBM Home Page Reader, Lynx, pwWebSpeak, etc.) that typically transform the two-dimensional visual representation of HTML documents into a one-dimensional representation, that is simple raw text which can be presented by a text-to-speech converter.

The need expressed by visually impaired people to use standard visual Web browsers such as Internet Explorer, has led us to develop new alternative display techniques to convey the spatially and visually related information that is present in an HTML document. Since the new challenge is to present the spatial relationships between HTML elements (tags) and their visual attributes, it immediately suggests the use of techniques and senses which are commonly used by blind and visually impaired people to find locations and to perceive their environment. These include both auditory and tactile approaches. A possible technique is to provide different auditory attributes (earcons / auditory icons) in association which each different tag. Moreover, the use of a 3D immersive audio environment, which permits to make a sound appear from a given position, may give blind users the sense of object location.

To enhance the mental representation of the virtual sonic environment, we propose to add to the hearing sense the haptic modality, that is the sense of where one's hand or arm is in space. Using a haptic tablet or a touch-sensitive screen would allow an individual to move his/her finger about while at the same time keeping track of his/her position on the device. The system would then respond with auditory feedback while the user moves his/her finger (device pointer) around the virtual screen. This approach, as shown by M. Lumbreras et al [7] and in our work [2] validates the hypothesis that a 3D immersive virtual sound environment combined with haptic manipulation of the audio environment can enable blind users to construct a mental representation of the spatial environment (in our case, the spatial layout of HTML documents).

WEBSOUND: ARCHITECTURE, USER INTERFACE, AND IMPLEMENTATION

One of the goals of *WebSound* is to provide researchers of HCI working on non-visual Web interfaces, with a tool that allows them to easily add new accessible functionalities to a standard visual Web browser. This doing, blind and sighted computer users would use the same environment to browse the WWW. To reach this goal, the design and implementation of a generic Web sonification tool (*WebSound*) has been done on the basis of two fundamental ideas:

- use of the same augmented visual browser based on the standard Internet Explorer 5.0, by "Internet surfers" (visually impaired and sighted users) as well as by programmers of new behavioral models;
- creation of an add-on tool, the *Workspace*, allowing to dynamically and visually create new access modalities.

Architecture

The WebSound application has been developed for two categories of users: the programmers of behavioral models and the "Internet surfers". This consideration has lead us to develop not only a new augmented visual browser for blind and sighted users but also a generic Web sonification tool that will be useable by HCI researchers.

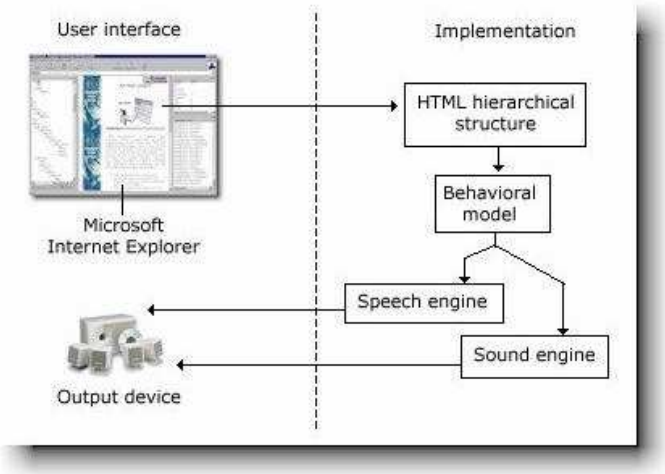

Figure 1. WebSound global architecture.

The WebSound components shown in figure 1 have been developed in order to furnish a set of functionalities that will permit programmers to easily add new behavioral model onto the application. This set of functionalities primarily consists of:

- the Internet Explorer ActiveX component that will be responsible for the Internet browsing;

- a component that will be responsible for representing the internal hierarchical structure of the HTML document;
- a component, called the *Workspace*, allowing the definition and creation of new behavioral models;
- a component that will be responsible for the restitution of the audio output in a 3D audio environment;
- a component that will be responsible for the textual output using a text-to-speech converter.

User interface

The user interface (Fig. 2a,b) has been divided into 6 different sub-windows, each displaying a different type of information.

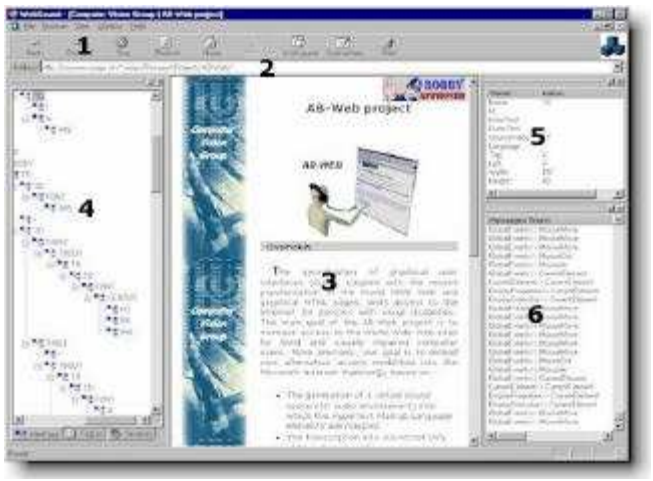

Figure 2a. WebSound interface displaying Internet Explorer.

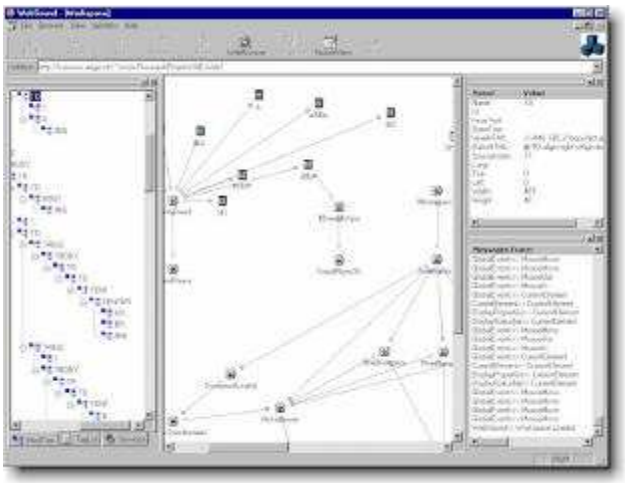

Figure 2b. WebSound interface displaying the Workspace.

The views (1,2,3) are common to the two categories of users mentioned earlier, since they are responsible for the Internet browsing. If the user is an Internet surfer, the windows (4,5,6) will be hidden. In this case, the WebSound application will look like the standard Internet Explorer, except that it will propose new access modalities. Nevertheless, the views (4,5,6) are very useful for the developer of new behavioral modalities since they display technical information such as the hierarchical structure of the HTML document (Fig 2a, view 4) or the internal content of each HTML element (Fig 2a, view 5).

Implementation

The implementation of a generic Web sonification tool has compelled us to address some difficult issues. The first problem to solve has been to find a way to obtain all the internal events that occur in the Internet Explorer ActiveX component, in order for example to determine the device pointer's position while moving on the browser. The second problem has been to determine which HTML tags the device pointer points on. Finally, we needed to create a component, the *Workspace* (Fig 2b) that was to be responsible for managing the creation of new behavioral models.

Figure 3. A HTML document and its hierarchical representation

The first problem has been solved using some possibilities that Microsoft offers to get informed on the internal events that occurs in a Windows based software or in an ActiveX application such as Internet Explorer. The technical details [8] of this solution are omitted here. To solve the second problem, we need to access the hierarchical structure of the displayed HTML document (Fig. 3, Fig. 2.a, view 3). This can be done using the Document Object Model (DOM) of an HTML document.

Finally, an off-screen structure that can receive and send messages, called *HtmlNode*, is created for each HTML element.

In order to create and test new behavioral models, all the events that are produced during the Internet browsing need to be detected and processed by the WebSound application. For example, if the user moves his/her finger across an image present in the HTML document, the *HtmlNode* associated with the image tag will send messages of type *MouseIn*, *MouseMove* (repeated while the device is inside the image border) and *MouseOut* to the WebSound application. In the same way, if the user moves around a header, the *HtmlNode* associated to this element will be responsible to send the corresponding messages.

Figure 4. An example of a dynamic sonic model.

Those messages have no behavior until they are connected, via the *Workspace,* to entities called *Services.* A *Service* is a piece of code that performs a particular task depending on the message type it receives. For example, if the programmer wants to play a sound when the device pointer moves over an image, he needs to create a *Service* (*MicroSound* in the figure 4) which can take a sound and play it when he receives a *MouseMove* message from the *HtmlNode* associated with the image.

This mechanism can be extended to any accessibility modality such as reading the HTML document using keyboard, navigating inside HTML tables, filling HTML forms, etc.

CONCLUSION

In this article, we present a new application, *WebSound,* that allow sighted and visually impaired "Internet surfers" to navigate on the same visual browser. It also describes a new concept that allows HCI researchers to dynamically create alternative access modalities on top of a standard visual Web browser (Internet Explorer 5.0). Finally, it also proposes a possible way to permit visually impaired individuals to explore spatial information by the way of audio-haptic interface.

FUTURE WORK

We are currently working on a speech recognition module that will allow blind as well motor impaired users to exploit our WebSound browser using the voice as input device. Moreover, we are also integrating a new haptic device (WingMan Force Feedback mouse from Logitech) to enhance the perceptual feedback loop. This new sense will allow blind users not only to listen to the content of each HTML element but also to feel those elements. For example, the blind user will be able to sense the contour of an image; it will also be able to feel the associated texture of an element instead of the associated sound.

ACKNOWLEDGMENTS

This project is financed by the Swiss Priority Program in Information and Communication Structure and by the Swiss Central Union Of and For the Blind. The authors are grateful to A. Barrillier, V. Hintermeister and V. Rossier for their help in the design and evaluation of the prototype.

REFERENCES

1.Petrucci L., Roth P., Assimacopoulos A., Pun T. *An Audio Browser for Increasing Access to World Wide Web Sites for Blind and Visually Impaired Computer Users*. HCI'99, Munich, pp. 995-998.

2.Roth P., Petrucci L., Assimacopoulos A., Pun T. *AB-Web: Active audio browser for visually impaired and blind users*. In International Conference on Auditory Display (ICAD'98). Glasgow, UK, November 1998.

3.Brewster S.A. *Using Non-Speech Sounds to provide Navigation Cues.* ACM transaction on Computer-Human Interaction 5(3), pp. 224-259.

4.Brewster S.A. *Providing a structured method for integrating Non-Speech Audio into HCI.* Doctoral Dissertation, University of York, 1994.

5.James F. *Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext.* ICAD'96 Proceedings. Xerox PARC, 4-6 Nov. 1996, pp. 97-103.

6.Blauert J. *Spatial Hearing.* The MIT Press, Cambridge, Massachusetts, 1997.

7.Lumbreras M., J. Sanchez. *Interactive 3D Sound Hyperstories for Blind Children*. CHI'99 Pittsburgh PA, USA.

8.Harth E., *Plate-forme de Sonification Dynamique de Pages Web Pour Aveugles et Malvoyants.* Diploma thesis, University of Geneva, 1999.