Auditory browser for blind and visually impaired users

Patrick Roth, Lori Petrucci, Thierry Pun
Computer Science Dpt., Geneva University
24, rue du Général Dufour
CH – 1211 Geneva 4, Switzerland
E-mail: Patrick.Roth@cui.unige.ch

André Assimacopoulos
UCBA/SZB – Swiss Central Union Of
and For the Blind, Schützengasse 4
CH – 9000 St.-Gallen, Switzerland

ABSTRACT

This paper presents our work on the development of a multimodal auditory interface which permits blind users to work more easily and efficiently with GUI browsers. A macro-analysis phase, which can be either passive or active, informs on the global layout of HTML documents. A subsequent active micro-analysis phase allows to explore particular elements of the document. The interface is based on : (1) a mapping of the graphical HTML document into a 3D virtual sound space environment, where non-speech auditory cues differentiate HTML elements; (2) the transcription into sound not only of text, but also of images; (3) the use of a touch-sensitive screen to facilitate user interaction. Moreover, in order to validate the sonification model of the images, we have created an audio “memory game”, that can be used as a pedagogical tool to help blind pupils learn spatial exploration cues.

Keywords

WWW, blind users, virtual sound space, non-speech cues, multimodal interface, pedagogical game.

INTRODUCTION

The World Wide Web has become a key tool to facilitate the autonomy of blind and visually impaired people in everyday activities, such as in education, business, and the home ([7] [8]). Unfortunately, the trend towards privileging graphic layout and objects in WWW pages, rather than plain text, has severely limited blind users access to such information.

A first obstacle to the design of Web browsers for blind people stems from the essentially 2D layout of the HTML documents to be shown. Current browsers for people with disabilities (e.g. [9]) typically transform the 2D content of the document into raw text, which is then translated by an essentially 1D output device such as a Braille line or a text-to-speech converter.

A second difficulty in presenting HTML documents to blind users arises from the presence of embedded images and/or graphical HTML elements. Existing browsers for blind users essentially remove all graphical information. This information often bears essential information, and ought to be preserved and suitably presented.

Computer Human Interaction ’99 – ACM SIGCHI Conference on Human Factors in Computing Systems, Pittsburgh, Pennsylvania, USA, May 15-20, 1999.

PROPOSED SOLUTION

Overview

In order to address these issues, we are designing an augmented browser which preserves both spatial layout and *graphical objects*, by means of a virtual 3D auditory space. Moreover, we strive at designing a spatial metaphor as the basis for the non-visual presentation, that adheres to the GUIB project specifications [6].

Two successive exploration phases, that are both supported by the browser, are performed by users in order to analyze a WWW document. First, a *macro-analysis phase* allows the understanding of the document structure and the element *types* (e.g. text, images, forms) displayed in the browser viewport. Secondly, a *micro-analysis phase* lets users focus on one particular object and thus obtain its information *content*. These phases are iterated: users perform a succession of macro- and micro-analysis explorations of the document in order to understand it.

Collecting document information

In order to analyze the spatial content of Web documents, it is crucial to obtain the attributes of all included elements, such as their type (e.g. text, link, image), position, size and content as they appear in the browser. To this effect, we have developed a proxy server which adds a Jscript program inside the requested document. This Jscript program analyzes the displayed document by retrieving the CSS attributes [2] of each HTML element.

MACRO-ANALYSIS FOR SPATIAL EXPLORATION

In order to allow the examination of the spatial rendering of the document in the user agent viewport, our interface permits two complementary modes of interaction: *passive* and *active*.

For the passive macro-analysis, the boundary locations (in cm.) and type of every element appearing in the viewport are reported by speech synthesis. The boundary location values correspond to tactile Braille rulers that have been taped on the screen frame. In this way, blind users can quickly and precisely obtain the general layout of the displayed document.

For the active macro-analysis, the blind user explores the document by moving his/her finger on a touch sensitive screen. During this exploration, the system responds with an audio feedback that provides a bidimensional “where and what” [5] information.

The sound itself, which corresponds to the type of the HTML element that is touched (the “what” information), is represented by an auditory icon [3]. We chose auditory icons as non-speech sounds since they facilitate an intuitive recognition of the related HTML element. For example, our interface currently employs typewriter and camera shutter sounds to respectively represent texts and images.

Simultaneously, in order to provide information regarding the screen location of an element (the “where” information), the earcon is projected into a 3D sound space [1] which represents the HTML document. This sound space is an immersive environment in which a particular acoustic signal $S(t)$ is perceived as originating from a given spatial location (usually expressed in azimutal angle θ_a, elevation angle θ_e and distance d, all defined w.r. to the listener). To achieve the desired mapping from the screen location of an element to this 3D sound space, $S(t)$ is the desired auditory icon, while $\{\theta_a, \theta_e, d\}$ correspond to the position of the finger on the screen. The 3D audio rendering is accomplished by convolving $S(t)$ with the head-related transfer functions (HRTF) defined for the current values of $\{\theta_a, \theta_e, d\}$. In practice, we have experimented the 3D sound space generation by means of the Microsoft DirectX library, in order to program the new Soundblaster Live card (the latter being used in the final prototype).

MICRO-ANALYSIS OF ELEMENTS CONTENT

The micro-analysis phase allows blind users to retrieve the content of the HTML elements, that have been found during the macro-analysis exploration. If the user points on textual information, speech synthesis with tonal variations and simple sound effects (such as beeps) [4] allow to render the text itself as well as textual differentiation such as links or emphasis.

If the user explores an image, the sound $S(t)$ is related to the pictorial attributes of the touched pixel (color, contour). Simultaneously, the $\{\theta_a, \theta_e, d\}$ values of the location of $S(t)$ in the virtual 3D audio space are related to the finger position in the image [10].

SONIC MEMORY GAME

In order to validate options taken during the system design, we have created a pedagogic game, the “Sonic memory game”. By means of this game, we wish to familiarize blind and visually impaired people with our audio-rendering metaphor.

The game consists of 3 pairs of simple shapes (e.g. square, triangle and circle) randomly hidden “behind” an array of numbers (1 to 6); each shape is hidden at 2 arrays locations. The goal of the game is to find which numbers correspond to the same shapes. To achieve this, the user first learns the audio rendering of each shape, then plays the game. Each shape is rendered by a dynamic acoustic signal, that simulates in the 3D audio space the effect of a finger following the outline of the shape.

EVALUATION AND FUTURE DEVELOPMENTS

A first prototype using simple auditory icons has been implemented and evaluated [10]. We have found that (a) the 3D auditory representation of the HTML document allows a good global layout determination, and (b) the passive mode in which the text-to-speech converter describes the screen spatial organization is useful.

Concerning the future developments of our interface, we are pursuing research on the micro-analysis of images. Another aspect we are addressing is the automated extraction of text from images, in order to render this information by text-to-speech conversion. Finally, we are developing further the “Sonic Memory game”, to address specific needs for the education of blind pupils.

REFERENCES

1. Blauert, J., Spatial Hearing, MIT Press, MA, 1983.
2. Cascading Style Sheets, W3C Recommendation; see http://www.w3.org/TR/REC-CSS2/, 1998.
3. Gaver, W.W., Auditory icons: Using sound in computers interfaces, *Human-Computer Interaction*, 2:167:177, 1986.
4. James, F. Presenting HTML Structure in audio: User Satisfaction with Audio Hypertext, ICAD ‘96 Proceedings, Nov. 1996, pp. 97-103.
5. Kosslyn, S.M., Koenig, O., Wet Mind: The New Cognitive Neuroscience, The Free Press Macmillan, 1992.
6. Mynatt, E., Nonvisual Presenting of Graphical User Interfaces: Contrasting Two Approaches, Proc. ACM CHI ’94, pp. 166-172.
7. New Technologies in the education of the Visually Handicapped, D. Burger, Ed., Les Editions INSERM, Paris, FR, Vol. 237, 1996.
8. Paciello, M.G., Hypermedia for people with Disabilities, ACM SIGLINK, V, 1, Febr. 1996, pp. 7-9.
9. Productivity Works. PwWebSpeak, See http://www.prodworks.com/pwwebspk.htm,