AUDIO-HAPTIC INTERNET BROWSER AND ASSOCIATED TOOLS FOR BLIND AND VISUALLY IMPAIRED COMPUTER USERS

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

[1] Computer Vision Group, Computer Science Department, University of Geneva
24, rue du Général Dufour, CH-1211 Geneva 4, Switzerland
Patrick.Roth@cui.unige.ch, http://cuiwww.unige.ch/~vision
[2] UCBA/SZB - Swiss Central Union of and for the Blind, and
ABA - Association pour le Bien des Aveugles
34, place du Bourg-de-Four, CH-1204 Geneva, Switzerland

ABSTRACT

In the Internet world, the widespread use of graphical user interfaces (GUIs) increasingly bars visually handicapped people from accessing digital information. In this context our project aims at providing sight handicapped people with alternative access modalities to various types of GUIs and graphic intensive programs, in order for instance to facilitate usage of Web services. We describe in this paper **AB-Web** [1], a 3D-audio Web browser that allows blind computer users to explore Web pages, fill in forms, etc., using a 3D sonic rendering. We also present **WebSound** [2], a generic tool that permits to associate with each HTML tag a given sonic object (earcon or auditory icon). Finally, we describe a series of associated programs composed of the family of sonic games **From Dots to Shapes** [3], as well as of **IDEA**, a tool that lets users comprehend simple drawings, as well as to create graphics.

Keywords: WWW, blind users, 3D virtual sound space, multimodal interface, sound and image processing.

1. INTRODUCTION

Internet access, coupled with graphical user interfaces (GUI) and browsers, has become common in education, business, and at home. However, due to the widespread use of GUIs, the so-called enabling technologies for sighted have become disabling technologies for the visually-impaired. Blind users are less and less able to benefit from the enormous wealth of archived digital multimedia that is offered over Internet.

The development of electronic aids for visually-impaired persons has been ongoing for several decades, such as for reading, facilitating mobility, and for educational and occupational; see e.g. [4] for a review. More recently, a pressing need to offer blind users access to Internet-based digital information has been identified ([5] [6]). The major difficulty in designing Internet browsers for blind people stems from the essentially *bidimensional* layout and nature of the textual and pictorial information that has to be presented. This is to be contrasted with the essentially «one-dimensional» nature of existing output devices such as Braille lines or text-to-speech converters. Another difficulty in presenting digital information to blind people users arises from the presence of embedded images; they often bear essential information, and should be suitably presented to users. In this context, our work aims at providing visually handicapped people with a series of tools that will help them to cope with Internet and its accompanying digital information world.

In the sequel, we describe in Section 2 the basic principle of the tactile-auditory interaction model. Section 3 discusses the 3D-audio Web browser AB-Web. AB-Web relies on the sonification tool WebSound that is presented in Section 4. The family of sonic games From Dots To Shapes, as well as the drawings creation and analysis tool IDEA, are addressed in Sections 5 and 6 respectively.

2. TACTILE-AUDITORY INTERACTION

Our driving idea is to map digital documents (e.g. text, graphics) into an immersive virtual 3D audio space, i.e. to transform the elements of the document into sounds that have a precise 3D location. The user interacts via a touch sensitive screen or a graphic tablet, and listens to the system's responses via headphones or loudspeakers. In the basic interaction mode, the user points his/her finger on the graphic tablet. In response to this input, the system generates a sound in the virtual 3D audio space:

- the sound location in the audio space corresponds to the finger position on the tablet. For example, if the finger is at the bottom-left of the tablet, the sound will seem to come from the bottom-left of the audio space;

- the sound characteristics (i.e. pitch, timber, etc.) depend on the nature of the touched element in the display. For example, when exploring a Web page, different sounds are associated with each possible HTML tag.

In this way the user can perform a “tactile-auditory exploration” of a screen document in order to understand its layout, that is “what is where”. In a further exploration phase, the user can then query more detailed information about a particular screen object. For example, touching a text object sends that portion of text to a text-to-speech converter. In case of images the sound corresponds to the local characteristics of the touched part of the image, such being or not over an edge.

Based on these principles, we have developed the following components:

- a 3D-audio Web browser, **AB-Web**, based on WebSound, that allows blind computer users to explore Web pages, fill in forms, etc., using the 3D sonic rendering;
- a Web sonification tool, **WebSound**, that allows to associate with each HTML tag a given sonic object (earcon or auditory icon) which is then projected into the virtual 3D sound space according to the finger position on the screen;
- a family of sonic games, **From dots to Shapes**, to help in a playful manner to teach blind pupils notions such as spatial environment and auditory interaction with computers;
- a tool, **IDEA**, that allows to comprehend simple drawings, as well as to create graphics.

3. AB-Web, A 3D-AUDIO WEB BROWSER

Current browsers for blind users typically transform the bidimensional content of the WWW documents into spoken text, for example by analyzing the incoming HTML source. These browsers often have difficulties in presenting the global layout of documents, and usually remove all pictorial information. We present here the design principles of an augmented WWW browser for blind users that addresses these issues.

The original aspect of our project consists in the creation of an immersive environment (3D sound space) into which the HTML document is mapped. In this environment, each object (text, link, image) is represented by an earcon [7] [8] whose simulated 3D location depends on the location of the displayed element in the Web browser window. The blind user can therefore explore the spatial organization of the document in order to obtain a mental image of what is displayed. Moreover, all images that are included in the document are also translated into sound. This translation should help blind users to discover the global characteristics of the displayed image during the exploration phases. These design principles have led to a working prototype currently being evaluated by visually handicapped people, using HTML documents of varying complexities.

3.1 Macro-analysis and micro-analysis of a document

Two successive exploration phases are performed by users in order to analyze a WWW document. These phases correspond to the “where” and “what” stages in visual perception [9], and are supported by the browser. First, similarly to the “where” stage of vision perception, a macro-analysis phase allows the understanding of the document structure and of the element types (e.g. text, images, forms) displayed in the browser viewport. Secondly, as with the “what” stage, a micro-analysis phase lets users focus on one particular object to obtain its information content. These phases are iterated: users interactively perform a succession of macro- and micro-analysis explorations of the viewport. The content and position of each HTML elements can be retrieved by obtaining their CSS attributes using JScript.

3.2 Active interaction for document exploration

It is well known that interaction between a user and a system is essential for the understanding of the system’s functionalities. Furthermore, the «action-perception loop» linking user actions and system responses, has to be as short as possible. To facilitate interaction, user input is accomplished via a *touch-sensitive screen* or a *graphic tablet*, that eliminates the need for inserting a pointing device (such as a mouse) into the loop. Using fingers, the user actively explores the screen both for macro-analysis, with the system responding with an audio description of the screen layout and elements relationships, and for the micro-analysis, with the system transcribing into sounds the information pertaining to the content of an element of interest (see Figure 1).

Figure 1: system setup, with headphones and touch-sensitive screen (courtesy S. Pereira).

3.3 Audio rendering by means of a virtual sound space

Audio feedback in response to active user exploration is transmitted using a stereophonic headset. We have studied two approaches for this audio transcription of information. The simpler technique consist of using a text-to-speech converter, for reading text and for describing the boundaries positions of a given screen element.

To provide more elaborated feedback allowing e.g. to «display» screen layout and images, we are using a second approach consisting of the generation of a *virtual sound space* [10] [11]. Such a sound space is an immersive environment in which a particular acoustic signal *S(t)* can be perceived as originating from a given spatial location (typically expressed in azimutal angle θ_a, elevation angle θ_e and distance d, all defined w.r. to the listener). This approach relies on Head-Related Transfer Function (HRTF) [12]. HRTFs model the spectral modifications of the source signal *S(t)* due to the listener's external ears, as a function of location $\{\theta_a , \theta_e , d\}$ of *S(t)*. HRTFs depend on each individual, and come in pairs $\{H_{L, \theta a , \theta e ,d}(),H_{R, \theta a , \theta e ,d}()\}$ corresponding to the filtering effects of the left and right ears respectively. The stereophonic effect of a sound source *S(t)* at any given location can thus be simulated by convolving *S(t)* with the appropriate HRTFs.

For audio rendering of an element at a given location in the screen location, *S(t)* will be related in the macro-analysis case to the type of the element, and in the micro-analysis case to its content; $\{\theta_a , \theta_e ,d\}$ will be related to the element's position. In this way, it is possible to generate an immersive virtual sound space related to a particular Web document. Concretely, we use for our project the Intel library RSX which is based on HRTF technology. These principles are implemented in WebSound, described below.

3.4 Evaluation

A number of punctual tests have been performed during all development phases of the program. Then, in November and December 1999, a series of intensive testing has been conducted over several weeks with the help of about 20 blind users, ranging from totally inexperienced with computers, to highly experienced. The main goal of the tests was to assert the usefulness of the spatial localization of screen elements. In summary, it has been observed that almost all users could learn fairly rapidly how to use the program, and more importantly confirmed that being able to build a mental representation of the screen layout was very useful to them.

4. WebSound

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 browser [13]. To reach this goal, the design and implementation of this generic Web sonification tool has been done on the basis of two fundamental ideas (see Figure 2):

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

WebSound will therefore allow to associate with each HTML tag a given sonic object (earcon or auditory icon [14] [15]). Each sonic object is then projected into the virtual 3D sound space according to the finger position on the screen by means of the Microsoft DirectX6.0 sound library. Finally, the audio rendering is done by an audio PC board (SoundBlaster Live).

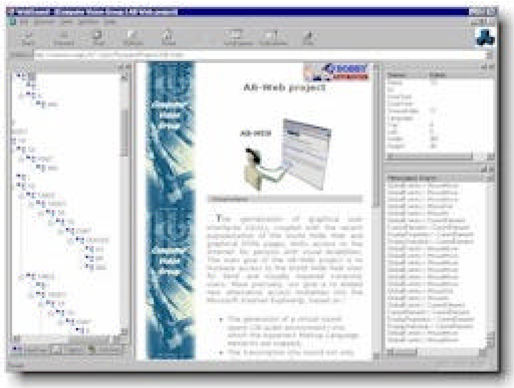

Figure 2: the WebSound interface, showing in particular on the left panel the HTML structure of the Web page that appears in the middle panel. The tool allows to interactively associate with each HTML tag a given earcon or auditory icon.

Two of the most difficult problems to solve have been firstly to find a way to obtain all the internal events that occur in Internet Explorer, in order for example to determine the device pointer's position while moving on the browser and secondly to determine which HTML tag is pointed on. The first problem has been solved using some possibilities that Microsoft offers to get informed of the internal events that occurs in a Windows based software or in an ActiveX such as Internet Explorer. The technical details of this solution are omitted here. To solve the second problem, we need to access the hierarchical structure of HTML documents in order to create an off-screen structure, called *HtmlNode*, of each HTML element that can receive and send messages. 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.

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 regarding of the message type it receive. For example, if the programmer wants to play a sound when the device pointer moves over an *HtmlNode*, he needs to create a *Service* (*MicroSound* in this case) 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 modalities such as reading the HTML document using keyboard, navigating inside HTML tables, associating force-feedback behavior to elements, etc.

5. SONIC GAMES: From Dots to Shapes

5.1 Why sonic games

Computer games are very useful tools to help teach concepts such as spatial organization, mathematics, etc. This is especially true for blind pupils and adults [16] [17]. In their case, in addition to providing an easy familiarization with the computer, the games can be tailored to help them acquire particular skills. Unfortunately, as for all software working with graphical user interfaces (GUI), most games use vision as the principal communication channel. We believe that sonifying games for the blind is essential to help them learn how to cope with computers, as well as to teach them some planar or spatial geometry concepts using the auditory sense.

The three games we have developed correspond to the three stages that blind pupils undergo when learning planar geometry, that is the learning of points, of lines, and of polygons and curves. In this "classical training", the first stage is accomplished by positioning small pawns (the points) on a tactile grid. The second training stage, aiming at teaching the line concept, consists of linking the points defined before. In the final stage, 2D shapes such as circles, squares, etc. are drawn. In a corresponding manner, the first game we have created, Simon, plays a sequence of spatialized sounds at several locations, each sound corresponding to one point. The user must recreate the correct sequence, as well as determine the sound locations. The second game, Between the Dots, asks the user to draw straight lines linking those points; the lines can be heard by means of their 3D sonic rendering. The third game, Concentration (or Memory), is described in more details below since it is the more complex and the one that has been the more thoroughly tested.

5.2 The Sonic Concentration Game

The rules of this game are similar to those of the traditional concentration game. Our game has three different levels, each level offering 4 pairs of geometrical shapes of increasing complexity that have to be matched. These shapes, such as a horizontal line, a square, a circle, a rectangle, are first sonified (see Sonification Principles below) and then randomly hidden "behind" a rectangular array comprising eight cells numbered 1 to 8. The goal is to associate the 4 pairs of shapes together, after the player has heard all the 8 auditory cues. For instance, if our array is configured as {square, rectangle, rectangle, circle, square, triangle, circle, triangle}, then the expected answer will be {(1,5),(2,3),(4,7),(6,8)}.

5.3 Sonification principles

Each shape is rendered by a moving sound that is acoustically drawn within the bidimensional sound space [18] [19]. This sound space can be either horizontal or vertical according to the player's wish. For instance, if the shape to be rendered is a square in a vertical plane, the player will hear in that plane the sound moving horizontally from the upper left to the upper right, descend vertically, move horizontally from the bottom right to the bottom left and finally ascend to reach the initial position.

Concerning auditory perception, much research points out to the difficulty in localizing a sound on the azimutal plane as well as in differentiating the front from the back [20]. Therefore, in order to deal with these problems, a number of functionalities were added such as: a reinforcement of the elevation rendering when using vertical plane, by means of a frequency variation (a decrease in highness will be perceived as a frequency decrease); the use of the Doppler effect to enhance the front/back differences; the possibility for the player to choose their sounds, as either melodic (represented by a sinusoidal timbre) or ecological (such as the sound of an helicopter); a particular coding of line junctions for patterns composed of several lines (e.g. square, triangle) by using alarms such as a "beep".

5.4 Learning modality

It is crucial to embed a learning session in each game that we create for the blind and visually impaired. The training stage that we developed is based on an action-perception loop in which the action is caused by the finger movement and the perception by means of two channels: tactile and auditory. The use of tactile perception therefore allows for a better learnability of the auditory environment.

Concretely, we created for all the geometrical shapes we used their tactile representations using a specialized thermal printer that produces output on swell paper. The swell paper outputs were then placed on a graphic tablet so that the user was able to follow the shape edges while listening to their audio rendering. A graphical tablet for capturing finger movements was preferred to a touch sensitive screen because in addition to provide absolute coordinates, using a horizontal pointing device causes less fatigue than using a vertical touch sensitive screen. For the auditory output, we mapped the graphical tablet into one of the two auditory planes mentioned earlier. In summary, during the training stage, the blind user explores the tactile representation with his/her finger and the system responds with an audio feedback related to the finger position on the graphical tablet.

5.5 Evaluation

We choose to first test the game with adults rather than children because blind adults have a longer experience in using geometrical shape representations, and they can provide more useful feedback to help in the design of the game. The concentration game was tested with four blind people aged 20-35. Two them were totally blind since birth, the third since the age of 8 and the fourth since the age of 13. Three participants have musical experience. Three game levels were defined, using the following shapes: (1) horizontal line, vertical line, diagonal line starting at the upper left corner, diagonal line starting at the bottom left corner; (2) one diagonal, square, triangle, circle; (3) square, rectangle, circle, oval. At each level, the participants were asked to perform the following two tasks: (a) to associate the corresponding auditory cues; (b) to determine to which geometrical shapes the auditory cues corresponded. At the end of the experiment, we asked the users to comment on the difficulties they encountered.

In summary, the results showed that the participants recognized and were able to match all the shapes that we proposed, except for the oval and the circle. After further training on these shapes only, one user was still not able to distinguish between them. We also observed that the training task was more useful for people blind since birth, and that blind people with musical experience had more facility to play with this game. Finally, all users believed that this game could be very useful for young pupils in their education.

6. IDEA, IMAGE DRAWING AND EXPLORATION BY AUDIO FEEDBACK

As we have seen in the section 5.4, the use of tactile material is necessary to facilitate the learning session. This situation implies that we often need to create tactile material corresponding to the displayed geometrical shape. This makes it difficult to easily add new geometrical shapes and sonify them by means of the WebSound application, since we need to create their corresponding tactile representations.

To solve this problem, we propose to substitute the tactile sensation by a combined audio-touch interaction. Using this interaction model, the tactile graphic is substituted by a digital vector graphic (see Figure 3). With this tool, the user can move his/her finger around the digital graphic while the system responds with an audio feedback which informs the user about the feature or kind of feature he has touched. This model correspond to the tunnel vision the blind user has when he/she explores a tactile image with his/her finger. Moreover, the user can not only follow contours but can also at any time ask for a sonic or textual rendering of the contour attributes such as its direction, or the location of its extremities.

We plan to use this tool to teach basic geometry concepts, or to help understand the shape of mathematical curves. Moreover, this tool could also be inserted to the WebSound application by transforming the images into vector graphics exploitable by the visually impaired.

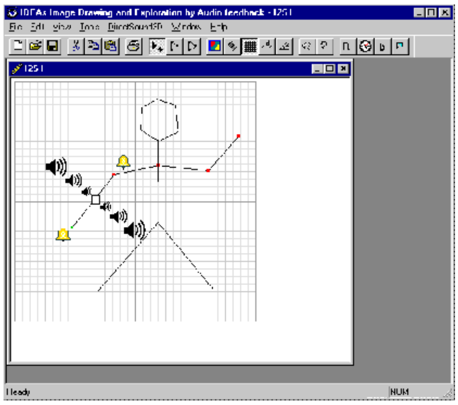

Figure 3: the image drawing and exploration tool. In order to follow a contour, the sound changes when the finger is on top or not of it. Sounds are different within and outside closed elements, and vary according to attributes such as color.

7. CONCLUSION

We have presented in this article our work aimed at providing visually handicapped people with a series of tools that would help them to cope with Internet and its accompanying digital information world. Based on a basic tactile-auditory interaction paradigm, we have developed **AB-Web**, a 3D-audio Web browser that uses 3D sonic rendering, **WebSound**, a generic tool that permits to associate with each HTML tag a given sonic object, **From Dots to Shapes**, a family of sonic games, and **IDEA**, a drawing creation and analysis audio tool. While IDEA is still the subject of ongoing development work, the other three components have been evaluated by up to 20 blind users, to their satisfaction.

8. ACKNOWLEDGEMENTS

This project is financed by the Swiss Priority Program in Information and Communication Structures, by the Swiss Central Union Of and For the Blind and the

Association pour le Bien des Aveugles. The authors are grateful to M.-P. Assimacopoulos, A. Barrillier, A. Bullinger, J. Conti, V. Hintermeister and V. Rossier for their help in the design and evaluation of the prototype.

9. REFERENCES

[1] P. ROTH, L. PETRUCCI, A. ASSIMACOPOULOS AND T. PUN, *Auditory browser for blind and visually impaired users,* In ACM SIGHI Computer Human Interaction 99, Conference on Human Factors in Computing Systems, pp. 218-219, Pittsburgh, PA, USA, 15-20 May 1999

[2] L. PETRUCCI, E. HARH, P. ROTH, A. ASSIMACOPOULOS AND T. PUN, *WebSound: a generic Web sonification tool, and its application to an auditory Web browser for blind and visually impaired users,* In Proceedings of the Sixth International Conference on Auditory Display (ICAD 2000), 2-5 April 2000. (to appear)

[3] P. ROTH, L. PETRUCCI, A. ASSIMACOPOULOS AND T. PUN, *Concentration game: an audio adaptation for blind people,* In CSUN's 15th Annual international Conference "Technology and Persons with Disabilities", Los Angeles, CA, USA, 20-25 March 2000 (to appear).

[4] J.A. BRABYN, *Developments in Electronic Aids for the Blind an visually Impaired,* IEEE Eng. In Medicine and Biology Mag., Dec. 1985. Pp. 33-37.

[5] *New Technologies in the education of the Visually Handicapped,* D. Burger, Ed., Les Editions INSERM, Paris, FR. Vol. 237, 1996.

[6] M.G. PACIELLO, *Hypermedia for people with Disabilities, ACM SIGLINK, V, 1 Febr.1996, pp. 7-9.*

[7] F. JAMES, *Presenting HTML in Audio: User Satisfaction with Audio Hypertext,* ICAD 96 Proceedings, Nov. 1996, pp. 97-103.

[8] M. WINBLATT, *Browsing the World Wide Web in a Non-Visual Environment,* ICAD 97 Proceedings, Nov. 1997.

[9] S.M. KOSSLYN and O. KOENIG, *Wet Mind: The New Cognitive Neuroscience,* The Free Press Macmillan, 1992.

[10] K. CRISPIEN, W. WURZ and G. WEBER, *Using Spatial Audio for the Enhanced Presentation of Synthesized speech within Screen-Readers for Blind Computer Users,* Computers for Handicapped Persons, Lect. Notes in Comp. Sc. 860, Springer-Verlag, 1994, pp.144-153.

[11] E.M. WENZEL, *Three-Dimensional Virtual Acoustic Displays,* in Multimedia Interface Design, ACM Press, New York, 1992, Ch. 15, pp. 257-288.

[12] J. BLAUERT, *Spatial Hearing,* MIT Press, 1983.

[13] L. PETRUCCI, E. HARTH, P. ROTH, A. ASSIMACOPOULOS and T. PUN, *WebSound: a generic Web sonification tool allowing HCI researchers to dynamically create new access modalities,* In Proceedings of the Conference on Human Factors in Computing Systems (CHI 2000), 1-6 April 2000 (submitted).

[14] W. GAVER, *Auditory icons: Using sound in computer interfaces,* Human-Computer Interaction 2, pp. 157-177, 1986.

[15] M. BLATTNER, D. SUMIKAWA and R. GREENBERG *Earcons and icons: Their structure and common design principles*, Human Computer Interaction, 4(1), pp. 11-44. (1989).

[16] T.V. RAMAN, *Conversational Gestures For Direct Manipulation On The Audio Desktop,* ASSETS'98 Proceedings, ACM Press.

[17] M. LUMBRERAS and J. SANCHEZ, *Interactive 3D Sound Hyperstories for Blind Children*, CHI99 Proceedings, May 1999, pp. 318-325.

[18] A.J. HOLLANDER, *An Exploration of Virtual Auditory Shape Perception*, Diploma thesis, University of Washington, 1994.

[19] M. SHIMIZU, K. ITOH and T. NAKAZAWA, *Pattern representation system using movement sense of localized sound*, Proc. HCI'99, Munich, Germany, 22-27 August 1999, 990-994.

[20] F.L. WIGHTMAN and, D.J. KISTLER, *Headphone simulation of free-field listening II: Psychophysical validation,* J. of the Acoustical Society of America, 85, 2, 1989, 868-878.