A 3D sound hypermedial system for the blind

Mauricio Lumbreras[1], Mariano Barcia[2], Jaime Sánchez[3]

[1,2]LIFIA - Laboratorio de Investigación y Formación en Informática de Avanzada
Fac. Cs. Exactas - Universidad Nacional de La Plata
Calle 50 y 115 - 1er Piso - (1900) La Plata
ARGENTINA

[3]Department of Computer Science, University of Chile
Casilla 2777, Santiago
CHILE

[1]*mauricio@info.unlp.edu.ar,* [2]*mbarcia@info.unlp.edu.ar,* [3]*jsanchez@dcc.uchile.cl*

ABSTRACT

It has been said that quite often a hypermedial application running over a GUI is somehow inappropriate or unusable. This is the case for end-users with little or no visual capabilities. In this paper we present a conversational metaphor to try to ameliorate this problem. We propose a framework in which the interaction is rendered using 3D sound. Several voices used as media to convey information are placed in the space through this technology. By using a glove the user controls the system manipulating a special version of 3D auditory icons, called *audicons*. Moreover, we introduce a technique called grab-and-drop, equivalent to the visual drag-and-drop. We show that this framework permits the building and adaptation of current hypermedia interfaces in a way that can be used without visual cues.

Keywords: drag-and-drop, 3D sound, aids for the visually-impaired, auditory I/O, virtual reality, metaphor, hypermedia

1. INTRODUCTION

It is widely known that hypermedial applications, in particular those accessed via CD-ROM, are becoming prevalent in current domains such as education and on-line documentation. Unfortunately, the interface metaphors emphasizing graphical images, icons, menus, and the like, do not consider visually-impaired people. Several initiatives enable the visually-impaired to have generic access to computing systems, such as the Mercator Project (Mynatt et al., 1992) and the European GUIB Project (Weber et al., 1993). In this paper we present a rather different and more specific approach by trying to answer questions such as:

- how visually-impaired people can take advantage of 3D sound technology?
- how we can produce suitable metaphors to browse information?, and
- what is the best modality in which to interact with a system without visual cues?.

New virtual reality technologies enable us to produce a spatialized kind of sound over headphones, called 3D sound. This sound processing is achieved by convolutioning the desired sound with certain FIR (Finite Impulse Response) filters, called HRTF's (Head Related Transfer Function), that are calculated with special acoustic recordings taken from real human ears (Wenzel, 1992). As a result, we hear through the use of the headphones a sound in a specific position of the space.

The advantages of these capabilities prompted us to produce some kind of virtual aural environment in order to generate a user interface to be used with information systems for visually impaired people. But nice sounds floating in the space do not provide a usable and friendly information system. How can we produce a useful framework? Several studies talk about the use of 3D sound , but very few deal with models and metaphors to exploit these capabilities.

 ISBN 0 7049 1140 X

2. THE METAPHOR

This study is based on a special version of a well-known hypertext model (Conklin, 1987). The basic architecture consists of a direct graph of nodes and links. Nodes represent certain documents. Links reflect semantic relationships between documents. Usually links can be displayed both explicitly, as a menu of choices on the screen, or implicitly, as an action taken when the user does something to an identifiable target object in the interface. The hypertext model offers many advantages such as the management of linked references, a context in which the content is not bound to a fixed structure, and dynamic interaction.

We adapt the conventional hypermedial model to fit our needs. Each node belongs to some type that reflects the kind of information contained. Each type of node is mapped to a speaker in a determined position in the space. The role of speakers is to engage a conversation, allowing the user to interact with participants by browsing and managing the flow of talking. The most interesting feature is the link selection. If during the talk there is a link to another concept, the speaker in charge of talk, makes a short comment. If the user is interested he or she may indicate the direction of the next desired speaker by using interfaces such as pressing a button in a joystick or grabbing-and-dropping some audicon in a position in the space. The user can do this, because she or he recognizes two cues: the voice, and more critically, the position of the speaker in the space. The voice of each person was sampled from different real subjects, and processed in order to generate 3D voice. Through the assigned type, each speaker plays a particular informational point-of-view and an anthropomorphical characteristic is assigned to the information content (Muller et al, 1992).

Unconsciously, the user is using a hypermedia system, because a hypertext structure was mapped to the conversation. Moreover, each link may reflect several conversational characteristics, such as requesting, acknowledging, counter-offering, etc. As a result, the information presented is fine-grained, nonlinear, and highly interconnected. Additionally the structure of the system and type of information are dependent on the application domain, because different domains may present different classes of information or point-of-view assignments.

3. THE ENVIRONMENT

The environment is controlled by manipulating a special version of 3D auditory icons, called *audicons*. An audicon is an entity that represents functionality. It is rendered with 3D sound without visual cues. Audicons have behavior, a position in space, and a set of sounds that can be played by the user. Thus we extend the Handy Sound environment (Cohen, 1995). Audicons are presented in space within a customizable horizontal plane, and selected by using a grab-and-drop technique. As a result, we present some kind of direct manipulation, a new concept for running a system without visual cues. By taking advantage of 3D sound, the user can select one of several simultaneous audicons. This is also referred to as the "cocktail party effect". In order to avoid annoyance, the user can select either sequential or simultaneous audicon notifications.

In addition, a static surrounding can be simulated by using an auditory version of a room metaphor enabling the modeling of the static environment architecture. Thus the user could move between rooms and a corridor. Rooms are organized along the corridor to provide a type of spatial index. In each room we can have a conversation related to the whole content.

In order to carry out task control , there is a special speaker, the assistant, that consistently remains in a fixed position in relation to the user. The assistant manages tasks such us backtracking and user orientation through context dependent advice. As a result, control task and information content are presented homogeneously. In other words, the user interact with different people.

When desired, the space aural simulation of the environment is reinforced by means of verbal descriptions presented by the assistant. There is evidence that this mode creates isomorphism between the mental model and the simulated space (Denis, 1993).

3.1 The grab-and-drop technique

Usually hypermedia applications are navigated by using the "point and click" technique. Arons (1991) states "...one cannot 'click here' in the audio world to get more information, by the time a selection is made, time has passed, and `here' no longer exists...". Keeping this idea in mind, it seems almost impossible to find a solution.

The well known drag-and-drop technique implies direct manipulation, but it is difficult to apply to a transient medium such as sound. In our approach by using a spatial metaphor, the clickable speaker exists all the time regardless of whether it is speaking or silent. This concept can be of great value when applying the proposed grab-and-drop technique.

3.2 Why Grab-and-Drop?

Usually sound is used to create an awareness of something happening or existing. We extend this notion in order to fully control the system by grabbing-and-dropping audicons, a special type of 3D auditory icons.

Our approach extends the ideas that have appeared in the literature (Gaver, 1986). Actually, in hypermedia applications it is very common to use a pattern of interaction, i.e. to apply an action over a target to backtrack, to stop, to play, and to get the next information chunk. The "drag-and-drop" technique is a powerful tool that matches this pattern of interaction very closely. If we provide the user with a way to pick up a certain entity and associate it with the target, we are offering the same functionality as offered by the drag-and-drop technique.

If the user can control the audicons in a kinesthetic way, by means of a glove, we are taking advantage of one of the major input modalities available to the visually-impaired. Therefore, this design of the user interface is based on the possibility of allowing users to actively interact with the environment instead of wandering around menus and prompts.

3.3 Grabbing audicons vs. dragging icons

Basically, the underlying idea is simple: the user, wearing a VR glove, closes his hand to pick up an audicon, grabs and drops it over the target object. The 3D sound offers the spatial attribute that makes possible the spatial manipulation of the audicon and the glove represents the means of interaction.

3.3.1 Similarities and Differences

In GUIs, the icon representation changes when it is dragged, notifying the user that the drag mode is on. Usually Virtual Reality gloves do not provide kinesthetic feedback. For this reason, the feedback must be rendered in a synesthesic way, replacing the tactile sensation by an adequate sound of something that is closing.

When a standard icon is dragged, its graphical representation is updated in the display to show the new position. When an audicon is grabbed, due to the transient nature of sound, a looping sound characteristic of this audicon is played to replace the update functionality. Here, we have two main ideas. On the one hand, the sound that is played in loop when the audicon is grabbed serves as grabbing feedback. On the other hand, it is also a notifier that the audicon grabbed is the right one. Another point to consider in the graphical domain when the user is dragging an icon that has been erroneously chosen. To cancel this activity, the user drops the icon in a void area of the desktop. Without visual cues, the user can lose confidence since he can not know if he is dropping the audicon in an empty space. To solve this problem, we provide a special audicon called "the trash" that absorbs an audicon dropped near it, and returns the audicon to the rest position, see Fig. 1.

Usually, in a GUI, the user scans the desktop to find the right icon. In the aural environment the user can drop a special audicon, the scanner, over the assistant, a special speaker, in order to get into scan mode. When the user passes the hand near the rest position of some audicon, she or he hears its characteristic sound with an intensity that is proportional to the distance between the hand and the audicon rest position. To cancel this mode, the user again grabs the scanner audicon over the assistant. To recognize or refresh one´s memory of which audicons are in the surrounding environment, the user can grab one of the refreshers over the assistant in order to hear sequentially or simultaneously the different audicons by their characteristic sound.

In a graphical-based hypermedial application the user navigates by clicking buttons. In the acoustic-proposed context, if the user wants to listen to a chunk presented by one of the speakers, he grabs and drops an audicon which is placed in the direction of the desired speaker.

Audicons are manipulated by grabbing and dropping. Each step is rendered with different sounds: 3D auditory icons for either synesthesic feedback or the icon´s own characteristic sound, and a looping *earcon* or musical sound for the grabbing process. Since this interaction belongs conceptually to the same entity, it has a behavior, and is conceptually perceived as a part of a whole, then we call this *audicon*. Thus the term audicon represents both a spatial/functional concept and its behavior.

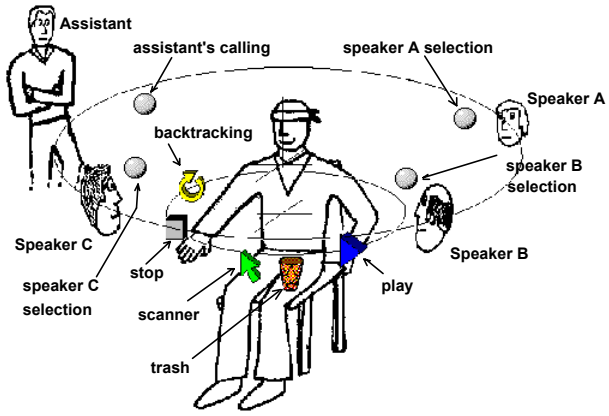

Figure 1. *Artistic version of the audicon´s environment. The audicon´s graphical representation is only illustrative because they are rendered with 3D sound.*

4. IMPLEMENTATION

We built several prototypes testing hardware alternatives. All versions were implemented using a PC clone as the underlying hardware. The first version tested was developed using a Gravis Ultrasound Max sound board as 3D sound generator. This board has a limited capability to render 3D sound, because the sound must be preprocessed and the interpolation technique used limits 3D positioning. The sound is produced by mixing four or six preprocessed sounds that represent above, below, right, left, front and back positions. In the same prototype to manipulate the environment we include a low cost version of a VR glove: the Power Glove. This glove, originally manufactured to be used with Nintendo games, was adapted to be plugged into a PC. The fact that the Power Glove has limited finger flexion detection, allows the accurate detection of two simple gestures like opened and closed hand. A software written in Borland C++ was in charge of querying the glove, playing the 3D sound, and carrying out the interface management and control of the hypermedial engine.

The final prototype was an upgrade of the first versions, because the Ultrasound board was replaced by one Alphatron 3D sound card (Alphatron, 1995). This piece of hardware has the capability to compute 3D sound in real time, by taking advantages of a Motorola DSP 56001. With this configuration the system is capable of rendering 2 or 4 sources simultaneously, depending on the playback frequency (44 or 22 Khz). Head tracking reinforces notably the sensation of immersion and offers a powerful opportunity to eliminate the ambiguity of the sound source position, but we have not included head tracking as yet.

5. DISCUSSION

It is premature to say that this system is fully usable. Initial testing shows that the cross modality could be interesting as a sound-position correlation training application, but long sessions with the system can cause fatigue due to continuous arm movement. Initial results indicate that a joystick or even a keyboard could be suitable devices to interact with the system, but always in the presence of 3D sound.

In addition, it is not currently possible to determine if it is necessary to render 3D sound in real time to build an usable system. This idea is supported by the fact that 3D sound can be computed off-line. Thus we can produce self-contained material packaged in a CD-ROM. With certain trade-offs related to real-time manipulation, the final product will contain a large set of 3D audicons to produce the environment simulation as well as the voice of the different speakers. This option requires basically an inexpensive platform: a personal computer, a CD-ROM player and a sound card. By using a set of HRTF's provided by Dr. Fred Wightman, we are testing the viability of our idea.

When entertainment is concerned, immersion, interactivity, and physical user activity are factors in the acceptance or rejection of a game. The grab-and-drop technique is a good choice for initial production of an unpublished type of entertainment for the visually impaired, taking into account the use of 3D sound technology with glove-mediated interaction.

Another interesting idea is the access to the WWW. It is possible, by using the proposed metaphor, to render synthetic 3D voice in real time, by passing the synthetic voice through a 3D sound processor. Mappings between VRML and an acoustic modality is not an unreachable idea.

Proc. 1st Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Maidenhead, UK, 1996

At the moment testing is still ongoing. For this reason we have no polished methodologies and well grounded results yet. In the near future we will probably have more data to share and discuss. For now, we believe that the metric organization of the audicons and their modalities suggests a new research dilemma.

Acknowledgment: We would like to thank to Dr. Fred Wightman of the University of Wisconsin for the HRTF set. This work was partially supported by the Chilean Science and Technology Fund(FONDECYT) grant No. 1950584

6. REFERENCES

Alphatron (1995), *Alphatron User's Manual*, Crystal River Engineering Inc., 490 California Ave, Suite 200, Palo Alto, CA 94306

Arons B. (1991), Hyperspeech: Navigating in speech-only hypermedia. *In Proceedings of Hypertext '91*, pp. 133-146.

Conklin J. (1987), Hypertext: An introduction and Survey, *Computer*, September 1987, pp. 17-41.

Cohen M. & Wenzel, E.(1995). The Design of Multidimensional Sound Interfaces. *Unpublished Technical Report 95-1-004*. Human Interface Laboratory, The University of Aizu, Japan.

Denis M. (1993), Visual Images as Models of Described Environments, in *Proceedings of the INSERIM-SETAA conference Non-Visual HCI*, (Paris, March 1993)

Gaver W., (1986) Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, **2 (2)**, pp. 167-177.

Lumbreras M., Rossi G. (1995), A Methapor for the Visually Impaired: Browsing Information in a 3D auditory Environment, *in Companion Proceedings of CHI '95*, (Denver CO, May 1995), ACM Press, pp. 261-262.

Muller M., Farrel R. et al (1992), Issues in the Usability of time-varying Multimedia, In *Multimedia Interface Design, ACM Press*, pp. 7-38.

Mynatt E., Edwards W. (1992), The Mercator Environment: A Non Visual Interface to X Windows and Unix Workstation, *GVU Tech Report GIT-GVU-92-05*

Weber G., Kochanek D. et al (1993), Access by blind people to interaction objects in MS Windows, *Proc. ECART 2 European Conference on the Advancement of Rehabilitation Technology* (Stockholm, may 1993), pp. 2.2.

Wenzel E.M. (1992), Localization in Virtual Acoustic Displays, *Presence*, vol. 1 number 1, pp. 80-107.