The Audio Browser: An Audio Database Navigation Tool in a Virtual Environment

John F. Whitehead

During this research: 
Human Interface Technology Laboratory 
Washington Technology Center 
University of Washington 
Seattle, Washington, U.S.A.

Currently: 
Department of Computer Science 
Duke University 
Durham, North Carolina, U.S.A. 
jfwhite@duke.edu or jf@woc11.sf.ca.us

Abstract

Many audio displays do not use a person's natural, three-dimensional perceptual and cognitive skills, which is particularly unfortunate for a musician who must sacrifice inspiration and spontaneity in order to use state-of-the-art equipment. The Audio Browser shows that with a visual and immersive "virtual environment" one can audition a library of sounds more effectively than with conventional auditory displays. This virtual environment can present sounds arranged in three dimensions (assisting comprehension of the database as a whole), coordinate the sounds with visual reinforcement (aiding perception and memory), and present many sounds together (allowing one to find a desired sound more efficiently). Auditioning sound samples can become less like changing computer settings and more like using a paint palette, which gives the composer an environment that is more conducive to creativity. This prototype indicates the promise that virtual technology holds for accessing large sound databases for effective music composition and performance.

1 Objective

Many artificial presentations to the senses do not make good use of a person's natural, three-dimensional perceptual and cognitive skills. A musician often must sacrifice some inspiration and spontaneity to incorporate state-of-the-art equipment into the creation of music [Oppenheim, 1986], such as when assembling sound samples for the orchestration of an electronic composition. This "object definition" task [Buxton et al., 1985] has become more involved as digital instruments have increased their capacity of stored sound samples. Despite the benefits that this increased variety brings, it has become more difficult for the composer to navigate a library of sounds easily to find one that is desirable. The composer can only listen to one sound at a time, must remember previous sounds when trying to compare the current one, and has little sense of the total available selection. When auditioning samples stored on a compact disc, it is difficult to get a sense of the sounds that are in hidden directories or may be related but are stored in different physical locations. The benefit of an immersive environment can bring to these situations is the nearly simultaneous presentation of many sounds, which allows a user to find a desired sound more easily and with a better understanding of the options available. The Audio Browser demonstrates this with a three-dimensional graphic and auditory display in virtual reality. It allows a user to browse a large library of sounds by moving within an organized database intuitively instead of by auditioning a poorly structured list sequentially. This virtual environ-

ment can present sounds arranged in three dimensions (assisting comprehension of the database as a whole), coordinate the sounds with visual reinforcement (aiding perception and memory), and present many sounds together (allowing one to find a desired sound more efficiently).

The intention of the Audio Browser is to provide a fully immersive virtual environment that gives the user the sensation of being inside the database of sounds, with the ability to move around its elements and to hear them from different perspectives. Although this does not easily (or cheaply) integrate with existing music studio equipment, it could be an effective prototype for an ideal audio database navigation tool.

2 Equipment

The Audio Browser runs on a network of four computers: a Macintosh to play the sounds, an IBM-PC clone to localize them, an SGI workstation to provide the graphic rendering, and a DEC workstation to run the controlling program. Digidesign's SampleCell on the Macintosh is used to store 8 MB of sounds (about 80 short, single-note, monaural selections from Digidesign's SampleCell CD-ROM, sampled at 22 kHz). The sounds are localized in three dimensions around the listener [Wenzel et al., 1991; Wenzel et al., 1993] with a Convolvotron from Crystal River Engineering running in an MS-DOS machine. The graphic rendering hardware is a Silicon Graphics Iris 4D/20VXG with custom NTSC converters to VPL EyePhones, and the graphics software is a custom imager that uses Iris libraries. User
movement and location are controlled by a custom handheld "wand" and are monitored by Polhemus 3Space Fastrak sensors. The controlling program is written in XLISP and runs on a DEC 5000-240. It makes use of virtual environment software from the Human Interface Technology Laboratory, including Mercury v. 1.5, VEOS, Sound Rendizer, and to a slight extent, FERN v. 2.0.

3 The Program

The primary goal of this Audio Browser is to study the feasibility and basic operations of an audio database navigation tool, not to test the most effective layout of sounds. For this reason, a simple structure is used for the database: about 80 samples, representing 54 different instruments, are arranged in a two-dimensional tree (though the user can move in three dimensions). It appears as about twenty groups of two to four cubes, each group colored differently, with lines linking each cube to its child and parent nodes, and spread across a plane in front of the user. (Cubes are used for this first Audio Browser to minimize the impact on the performance of the graphics processors.)

Related sounds are grouped together and become as specific as one travels up the tree. For instance, a snare drum rim shot sample is grouped with other snare rim shots, which are part of the traps family, which is part of the percussion family:

percussion:

hand drums: 
miscellaneous:

tuned percussion:

trap:

rotocon:

snare hits:

snare rolls:

snare rim shots:

snare rim shot sample 1
snare rim shot sample 2
snare rim shot sample 3

The program implementing the Audio Browser is designed to be independent of the audio data. In this way, a user can call up a different set of sounds without needing to change any code. Currently the only assumption about the sounds is that they are in a tree structure, but there is no restriction on the number of children for each node or on how deep the tree is. This reinforces the idea that the Audio Browser is a tool for examining a database of any sounds, not just the instruments used for this experiment.

The database of sounds comprises a list of information about each node in the data tree. An example for the snare rim shot part of the example database is

```
(xim #56 0 68) #56 2) trams rim3)
(xim #56 0 64) #56 0) rim ml)
(xim #56 0 64) #56 1) rim ml)
(xim #56 0 64) #56 2) rim ml)
```

The first field is the name of the sound and the second is an array of the X Y Z that corresponds to the virtual world coordinates. Since the coordinates of the database are stored with the sounds and not in the controlling program, the visual layout can also be altered fairly easily.

The third field denotes the MIDI note and channel corresponding to how the sound is set up in SampleCell. The fourth field is the name of the parent node and is used for setting up and displaying the tree (arurally and visually). The last field is the name of the child node and indicates which child's sound is represented by the parent (or if it is nil if there is no child node).

The controlling program sets up the sounds in the virtual environment and monitors the user's actions. Sensors check the user's head and control which parameters of the sound are changing. As well as whether a movement is forward or backward) or selection button has been pressed. At regular intervals, the user data are processed. If the user moves to a new group of sounds or if the user presses the selection button, the closest objects are sounded and change color as a visual reinforcement of their location. The objects are sounded by a call to the Sound Rendizer, which triggers SampleCell to play each sample through the Convolvotron to localize the sound according to each sound's location in relation to the position and orientation of the user's head.

4 The Audio Browser

As an example of how the Audio Browser can be used is shown in a videotaped navigation of the database. After the sound samples have been loaded into SampleCell, the user starts up the program, puts on the stereo headphones, dons the three-dimensional goggles, and maneuver by moving her head and body naturally or by pressing the buttons on the joystick-like wand.

The user has the sensation of being fully immersed inside the database of sound objects and has the ability to move around to see and hear sounds from different perspectives. The sounds are spatialized to seem to originate from their coordinating visual objects' locations and change color when selected. A user can browse by following the visual links between groups, traveling from an auditioned sound to a related sound in a group of sounds.

For instance, if the user wants to hear the piano samples in the database and is near a part of the structure that sounds like a keyboard instrument, he or she can follow that sound object's link to a group of several keyboard sounds. If this group consists of an organ, a piano, and a synthesizer sample, she would know to follow the middle link to the group of sounds seen just beyond it. She could observe that the four remaining nodes (samples of a Bosendorfer, a Yamaha, etc.) are the only piano samples in the database contains, and move over to and audition them.
5 Results

Overall the Audio Browser is effective in enabling a person to have access to a variety of sounds and explore them naturally instead of having to worry about typing in numbers or guessing what poor verbal descriptions signify. The perception of the sounds and ability to compare them is greatly aided by their sonic spatialization and the graphic cues of seeing objects light up in the corresponding visual location. The utility is that the user does not need to know in advance what sounds are available; she can just move herself toward sounds of interest and know quickly the extent of the contents of the database. She can go to a region of interest and know that all the objects she sees in her vicinity are the ones that will be useful to her.

Another aspect of the Audio Browser that works well is that it is easy to get a good sense of the number of sounds available. One can back up a little or fly up to see the entire database—even staying close and looking around provides a good sense of the overall structure.

Similarly, one can tell if no other sounds of a certain type are available. For instance, while in the acoustic guitar section, one can see easily that there are four and only four acoustic guitar samples available. On a sampling keyboard it is sometimes difficult to know what sounds are not available, and this can be important to a composer trying to find the right one.

6 Future Research

In addition to the basic features implemented in this prototype, there are several interesting avenues for continued research. One option is to design a database layout that is organized by sonic qualities such as timbre (Wessel, 1979), perhaps arranged in three dimensions. A future Audio Browser might also allow alterations to its database by letting a user arrange sounds into a layout according to personal preference. It would also be advantageous to be able to create new sounds by grabbing, combining, and manipulating the existing ones. Improved interfaces could allow the composer to load the sounds directly into MIDI scores or sampling instruments for complete (instead of single-note) auditions.

Unfortunately, two of the strengths of the Audio Browser—its intuitive navigation through a sound database and its convenience for musicians—are severely limited by today's very expensive, complex, and bulky virtual reality equipment. The goggles are uncomfortable and the computers are dedicated workstations that are not easily integrated into an audio studio. The work required to initialize SampleCell with MIDI data for each sample in a sound library and to create the physical structure of the database is considerable. Improvements in these areas could allow access to larger, more complex audio data structures which perhaps could be generated automatically. These advances could allow an Audio Browser to be available for semi-professional and professional composers and would be less likely to interfere with the composer's concentration.

7 Summary

The advantage of implementing the Audio Browser virtual environment is recognized when the user no longer needs to search a linear representation of the sound library, but can explore the database itself at its lowest level, as a collection of sounds. The musician does not need to know the content of the database in advance—he can quickly learn what options are available for her composition by exploration. Unlike other auditory displays, the user has a sense of the overall structure of the database and can easily examine it at the micro- or macrolevel.

Creating a database of sounds that can be displayed and navigated is difficult because, unlike still visual images or text, sound changes with time. However, by presenting sounds in an artificial environment more similar to that of our real and visual three-dimensional world, one can create an audio database tool that benefits from a person's natural, three-dimensional skills. With careful design, a user can experience an increased sense of an audio database as a whole, a better knowledge of the relations of sounds to each other, and an improved method of finding sounds with desired characteristics. This work on a prototype Audio Browser indicates that effective use of virtual environment technology can provide a more natural interface for audio database navigation, which will allow a composer to write with fewer impositions on his or her creativity.

8 Acknowledgments

I am very grateful to Brian Karr for assistance in planning the Audio Browser and to Max Minkoff for programming help. I also would like to thank Katherys Best, Colin Bricken, William Bricken, Marc Cygnus, Toni Emerson, Tom Furness, Ari Hollander, Andy MacDonald, Dan Pirone, Squish, and everyone else at the HIT Lab for their support of and feedback on the Audio Browser.

References


Oppenheim, D.V. (1986). The Need for Essential Improvements in the Machine-Composer Interface used for the Composition of...
