FUTURE DIRECTIONS

An important step in this line of research is to perform a series of experiments to quantify the effectiveness of the VR audio system as a function of operating parameters. Experiments are scheduled to determine the localization accuracy possible with the existing Mercator spatialization engine over a range of bandwidths. The results of this work should be helpful in determining the appropriate minimum sample rate for spatial audio. Experiments are also being designed to determine the effects of latency and hysteresis threshold width on a user's immersion in a virtual environment.

A second set of experiments is planned to determine the usefulness of spatial audio in a virtual environment. With its lack of sufficient visual resolution, the virtual reality system should benefit from auditory cues. Auditory cues have several important perceptual characteristics, and we want to find optimal configurations of visual and auditory resolution that improve immersion without slowing down performance. For example, one trade-off could be to use spatialization algorithms for navigational beacons but not for close-range manipulation tasks.

ACKNOWLEDGEMENTS

REFERENCES

1. Blauert, J. (1983) *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press: Cambridge, MA.

2. Borish, J. (1984) Extension of the image method to arbitrary polyhedra, J. Acoust. Soc. Am., 75, 1827-1836.

3. Burgess, D.A. (1992) Techniques for low cost spatial audio, Proceedings of the Fifth Annual Symposium on User Interface Software and Technology (UIST '92), ACM, New York, 1992, pp. 53-59.

4. Gehring, B. (1990) Focal Point^{TM} 3D Sound User's Manual, Gehring Research Corporation, 189 Madison Avenue, Toronto, Ontario, Canada, M5R 2S6.

5. Mynatt, E.D., Edwards, W.K., (1992) Mapping GUIs to auditory interfaces, Proceedings of the Fifth Annual Symposium on User Interface Software and Technology (UIST '92), ACM, New York, 1992, pp. 61-70.

6. Postel, J. (1980) User datagram protocol, RFC 768, Network Information Center, SRI International, Menlo Park, Calif., August 1980.

7. Sheridan, T.B. (1992) Musings on telepresence and virtual presence, Presence, 1, 120-126.

8. Verlinden, J.C., Bolter, J. D., & van der Mast, C. (1993) Voice annotation in virtual environments, (submitted to ESS '93).

9. Verlinden, J.C., Kessler, A., Hodges, L.F. (1993) The Simple Virtual Environment (SVE) Library: Users Guide, Tech Report GIT-GVU-93-24.

10. Wenzel, E.M. (1992) Localization in virtual acoustic displays, Presence, 1, 80-107.

11. Wightman, F.L. & Kistler, D.J. (1989a) Headphone simulation of free-field listening I: stimulus synthesis, J. Acoust. Soc. Am., 85, 858-867.

12. Wightman, F.L. & Kistler, D.J. (1989b) Headphone simulation of free-field listening II: psychophysical validation, J. Acoust. Soc. Am., 85, 868-878.

lacking sufficient optic flow and stationary features. Furthermore, the application provided little interaction. Observations show that the immersion increases as the control on the environment increases (Sheridan, 1992).

There are two simple features that could have been added to this environment that would have greatly improved interactivity and visual cues. The first improvement would be to model the field pattern of the virtual radio. For example, a real radio usually sounds clear from the front, but quiet and muffled from that back and sides. This effect could easily be modelled in the virtual environment using filters in the DSP microcode and source orientation information from the trackers.

A second improvement would be to add animated level meters to the radio. This feature would require an extension to the audio control protocol to allow the spatialization host to send information to the VR host about the current amplitude of the envelope of the sound. Animated level meters would improve the user's association of the sound with its visual cue by means of ventriloquism.

Performance
Previous work with audio in virtual environments suggests that an update period of 90ms is adequate to give an illusion of continuous motion for angular velocities associated with normal head movement (less than 360°/sec) (Wenzel, 1992). Unfortunately, many estimates of this type ignore the difference between update period and update latency. Update period is the amount of time between position updates. Large update periods result in jerky, discontinuous motion of a virtual sound source. Update latency is the amount is time between a user's movement and the resulting relative motion of the sound source. Large update latencies result in the audio component of a virtual environment lagging perceptibly behind the inputs. Update latency and update period are completely independent. In practice, small update periods are easier to achieve than small update latencies because sound source movement requests can be processed in a pipelined fashion.

To update the position of a virtual sound source in a distributed system, the following events must occur:

1.The tracker detects the receiver positions and transmits them to the VR host machine. In our system, this takes 37ms, with most of the delay caused by the serial line between the tracker and the VR host.

2.The VR host formats the position information and writes it to the local area network. If no other application is running on the machine, this operation is expected to take on the order of 10ms, which is the duration of a typical Unix time slice.

3.The position information passes through the network to the spatialization host. In our test implementation, the packet passes through two network gateways[1], and the

transmission latency is about 10ms in a quiet network.

4.The spatialization host reads the information from the network, translates the coordinates, generates a pair of spatialization filters, and sends them to the DSP. In this implementation, all of this takes about 22ms. (The time to send a filter pair to DSP is about 12ms.)

The maximum RS-232 transmission rate of 20 updates per second from the Ascension tracker to the VR host determines the maximum position update rate for the VR audio system. Notice that the latency (80ms) is greater than the period between updates (50 ms), giving the system a pipelined effect (see figure 3). As a result, the system usually generates smooth audio source motion, even though there may be a perceptible lag between a user action and the movement of the virtual sound source.

Reliability
The two weaknesses in the VR audio system were the UDP-based control link and the interference of the headphones with the magnetic trackers.

Software for the UDP control link did not check for out-of-order or incomplete packets. It simply accepted whatever came across the network. Surprisingly, the control link rarely failed, and then only during periods of unusually heavy network use. Because the spatialization host kept no state information, it recovered from errors quickly. This is an important design feature to remember for future implementations—state information should be kept inside the client to the greatest degree possible, otherwise a more reliable (and therefore slower) transport mechanism must be used.

We also found that the magnetic fields generated by the headphones disrupted the operation of the Ascension Bird trackers when a tracking receiver was closer than 10cm to a headphone earpiece (with Sennheiser HD540 headphones). Thin, ferrous shielding plates inside the HMD may provide a practical solution. Electrostatic headphones do not generate these fields, have excellent acoustic properties, and might eventually be considered as an alternative solution, but existing models tend to be too bulky for use in HMDs
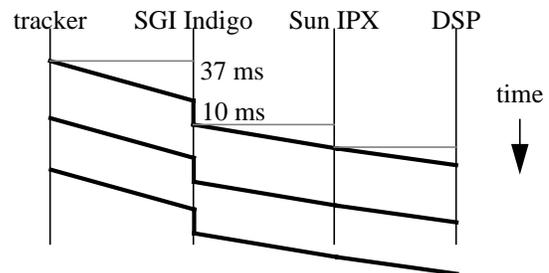

Figure 3. Illustration of pipelined position updates

1. The spatialization host and VR host were in different rooms, separated by a distance of about 30m.

allows for distributed operation, because a comprehensive VR system can easily consume the resources of multiple workstations.

Connecting the Systems

A set of simple UDP-based[1] routines were written to allow the VR host (an SGI Indigo Elan) to transmit tracking information to the spatial sound system (a Sun SPARCstation) through a local area network. The VR host sends a packet specifying a vector from the center of the user's head to the virtual sound source. From this vector, the spatialization host generates a pair of spatialization filters and sends them to the DSP board. The DSP board processes a local audio source (in this case, from an FM radio or CD player) with the specified filters. Processed audio is then sent back to the VR area in analog form through two coaxial cables. Ideally, this audio would be returned to the VR host via the local area network, but analog cables were used to reduce implementation effort and to insure good performance. Once in the VR area, the spatial audio signal is then amplified to a comfortable level

---

2. This filter pair is designed to approximate the Head-Related Transfer Function (HRTF), which is a concise name for a set of effects imposed on the sounds you hear by the shape of the outer ears, head, and upper body (Blauert, 1983). Because these effects vary with the orientation of the head and body relative to a sound source, the brain has evolved means of using them to estimate the direction of a sound. By artificially generating these effects with sufficient accuracy, it is possible to create the sense of a sound coming from a particular direction, even though the listener is actually hearing the sound through headphones.

e, the brain has evolved means of using them to estimate the direction of a sound. By artificially generating these effects with sufficient accuracy, it is possible to create the sense of a sound coming from a particular direction, even though the listener is actually hearing the sound through headphones.

1. UDP is a simple, unreliable datagram protocol which is layered directly above the Internet Protocol (IP). (Postel, 1980)
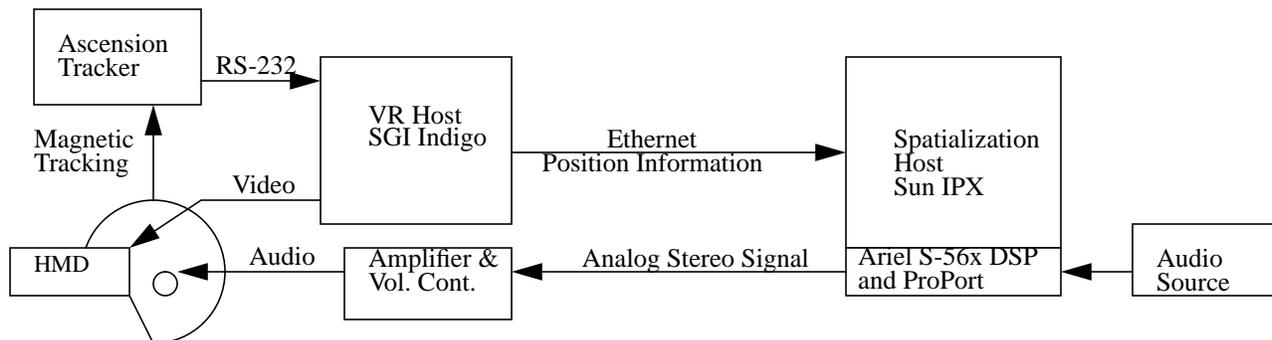
Figure 2. View of the virtual environment

and presented to the user via the headphones in the HMD. The system is diagrammed in figure 1.

OBSERVATIONS OF THE TEST SYSTEM

Based on the resulting Virtual Reality system, a demonstration application was developed. This section presents some observations. First, we will focus on visual cues and immersion. After that, two technical issues are discussed: performance and reliability.

Visual Cues and Immersion

The first testbed application was simple; the virtual world consists of a green field and a hand-held radio (see figure 2). A virtual sound source is attached to this radio. As the user moves the cursor (hand-held tracker), he or she will hear the sound (coming from a real radio channel) moving as well.

When first introduced to this virtual world, a user may not immediately perceive a connection between the visual cue and the virtual audio source. There is a brief training period involved, and experiments are planned to determine the typical length of adaptation.

Our initial hope that the visual cues would improve the effectiveness of the spatial audio is somewhat diminished. Hardware limitations of the helmet mounted display, such as its weight, the low resolution (equivalent to being legally blind), and the narrow field of view seemed to weaken any augmentation of the spatial cues. However, the application provided few visual cues, with a low level of detail, and

Figure 1. Hardware for testbed virtual environment with spatial audio

A First Experience with Spatial Audio in a Virtual Environment

David A. Burgess and Jouke C. Verlinden

Graphics, Visualization, and Usability Center
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
(burgess,jouke)@cc.gatech.edu

ABSTRACT
This report describes a first attempt to use 3-D audio in a virtual environement and the lessons learned from the experience.

KEYWORDS
virtual reality, virtual environments, spatial audio, multimedia, man-machine interfaces, human-computer interfaces, audio choreography

INTRODUCTION
While many Virtual Reality systems include sound as a part of the virtual environment, it never receives as much attention as the video gear; thousands of dollars are spent for head trackers and helmet mounted displays (HMD's), while no special audio equipment is used. Some systems do not even use headphones, though these are considerably less expensive than any visual HMD.

Virtual reality research at Georgia Tech already uses sound, but no spatialization is applied. Auditory cues provide positive or negative feedback while performing an interactive task. With typical VR tasks as grabbing and pointing at objects, or navigating trough space, user performance is restricted by the lack of accuracy of the trackers, lag times, and the limited resolution of the HMD. Auditory feedback proves to be a useful and sometimes even necessary source of information while performing these tasks. Another application of sound within our virtual environments is speech annotation (Verlinden, Bolter, van der Mast, 1993).

For several years, the great expense of the 3-D audio hardware provided a convenient excuse for most VR researchers to largely ignore the auditory components of their virtual worlds. Today, however, the emergence of several new spatial audio systems based on low-cost single chip digital signal processors (DSP's) (Gehring, 1990, Burgess, 1992) leaves no excuse not to have spatial audio in a virtual world.

HARDWARE AND SYSTEM SOFTWARE
The testbed for auditory virtual environments was constructed from existing hardware and software for two ongoing projects in the Georgia Tech Graphics, Visualization, and Usability Center (GVU Center).

Original Systems
The GVU Center VR system consists of a Virtual Research HMD, a Virtual Research Cyberglove, and a dual-receiver Ascension Technologies Bird magnetic tracking system connected to an SGI Indigo Elan workstation. An extensive library was written to support the construction of virtual environments with this hardware (Verlinden, Kessler & Hodges, 1993). This library provides utilities for the loading and saving of objects (hierarchical groups), initialization of special devices, etc.

As part of the Mercator[1] project (Mynatt & Edwards, 1992), the GVU Multimedia Computing Group developed a low-cost spatial audio system. The spatialization system operates by applying a pair of digital filters to an audio stream[2]. By changing these filters in real time, it is possible to simulate the movement of a virtual sound source. The filters used in this system are based on a set of middle-ear recordings provided by Prof. Fredric Wightman, of the University of Wisconson at Madison (Wightman & Kistler, 1988a&b). Additional processing is applied to the sound to generate simple environmental effects, specifically dense reverberation and atmospheric dispersion. These environmental effects help the user get a sense of his distance from the sound source and also give a sense of the size of the listening space. The system hardware consists of an Ariel S-56x DSP board to perform digital audio computation, a SPARCstation IPX host machine, and an Ariel ProPort 656 for digital to analog conversion. DSP microcode and low-level utilities are available for controlling the spatialization engine, but at the time of this work, no high-level programming interface was available. The current system supports only one virtual audio source.

Because the SGI Indigo provides its own DSP engine, it is conceivable that the spatial sound system could be ported directly to the VR machine. We chose not to follow this route, however, because of the difficulty of porting DSP microcode and associated host-side driver software. It is also important that the control interface for the VR audio system

---

1. Mercator is an effort to provide blind users with access to X-windows applications. In Mercator, the components of a graphical interface are mapped into a virtual auditory space.