3-D Interactive Percussion: The Virtual Drum Kit

David Burgess and Elizabeth Mynatt

Graphics, Visualization, and Usability Center
Georgia Institute of Technology
Atlanta, GA 30332
[beth,burgess]@cc.gatech.edu

ABSTRACT

This interactive experience places the user in an auditory virtual environment which combines computed models of natural sound sources with spatial audio. The user wears headphones and a lightweight head-tracking receiver. A second receiver is held in the user's hand and used as a *mallet* for striking various virtual objects, or *drums*. When struck, each drum generates a particular percussive sound. Each sound is spatialized to appear to come from the direction of the virtual drum that generates it. The drums are also responsive to the manner (velocity, angle, etc.) in which they are struck and may sound different when heard from different angles.

KEYWORDS

Acoustic displays, spatial sound, auditory perception, virtual displays

INTRODUCTION

This interactive exhibit is designed to demonstrate some of the many possible properties of sound in three-dimensional space, and to allow the user to experiment with how these properties affect the listener's perception of the sound source and the listening space.

The exhibit also demonstrates a number of human factors and electrical engineering principles which can be applied to improve the user's level of comfort in an acoustic environment and to allow the generation of complex environments at modest computational costs. The exhibit is designed to produce a compelling, realistic experience with a simple set of techniques.

MODELLING PERCUSSIVE SOURCES

In this environment, a virtual drum has three parts: a sound generator, an auditory field pattern, and an algorithm for controlling user interaction.

CHI94 Companion-4/94 Boston, Massachusetts USA
© 1994 ACM 0-89791-651-4/94/0045...$3.50

Modelling Sound Generation

Although the use of sampled sound is popular, we have chosen not to use sampled sound for this environment. Synthetic sound affords us much greater flexibility—we can generate whatever we need, not just what we could find to record. Synthetic sound also allows us to model user interactions with the sound source that are natural and consistent with everyday experience. With sampled sound, you get the interaction you recorded, and that is all.

Our models of percussive sounds are built from sets of dampened oscillators (Gaver, 1993). The sound source is modelled as a bank of filters[1] and the interaction with the source is modelled as an input to the filter bank. For impact sounds, the inputs are short pulses. Other types of inputs can be used to model other types of interactions. For instance, we can model scraping by sending noise through the filter bank. Possible "drums" based on this model might include:

a block of wood, tapped as the hand passes
a metal rod, struck as the hand passes
a taunt string, plucked as the hand passes
a ball of water, floating out in space

We can also build more complex instruments from collections of these models. For example, a rattle might be modelled as a collection of tiny percussion sounds, all shifted from each other slightly in time. In this exhibit a large region is defined in the area above and behind the user's head where the hand acts as a rattle instead of a mallet. In the actual implementation, the rattle is a single filter bank, but each excitation pulse is expanded into a packet of closely-spaced impulses with slight random variations.

Modelling Field Patterns

Real sound sources have field patterns, which means that the volume and timbre of the source are different in different directions. The virtual drums also have field patterns, produced by digital filters whose parameters are varied with the orientation of the source relative to the listener.

1. A dampened oscillator and a second order infinite impulse response linear filter are equivalent.

Interacting with the Drums

In order to have a natural, familiar interaction with the virtual drums, there must be a carefully designed set of rules to map the path and velocity of the hand to the force and hardness of the virtual mallet. The drum is mapped to a surface in space. As the hand passes through this surface, a pulse is generated an passed through the filter bank. The faster the hand is moving, and the closer it passes to the center of the drum's spatial region, the shorter and more intense the pulse, corresponding to a harder strike with a harder mallet.

Interacting with the Rattle

The stream of pulses driving the rattle is based on the fourth derivative of the tracker's motion with respect to time. This fourth derivative is sometimes called *jerk*, and it indicates changes in the acceleration of the hand. The larger the jerk, the louder the rattle.

LOW-COST AUDIO SPATIALIZATION

The virtual environment uses 3D binaural sound, similar to that produced by the Convolvotron (Wenzel et al, 1988) or Focal Point (Gehring, 1990) systems, but without the use of any special signal processing hardware. This low cost 3D audio is made possible by both an improvement in computer technology and by an analysis and modification of 3D audio techniques.

Filter Manipulation

Sound is spatialized by the application of a pair of digital filters, one for each. This filter pair is designed to approximate the Head-Related Transfer Function (HRTF), which a set of effects imposed on sounds by the shape of the outer ears, head, and upper body (Blauert, 1983). By artificially generating these effects with sufficient accuracy, it is possible to create the sense of a sound coming from a particular direction. An analysis of filter sets used in existing applications showed that a majority (75%) of the computational effort of applying the filters was wasted on the manipulation of noise (Burgess, 1992). Controlled tests of the new low-cost filter sets show that their effectiveness is comparable to previous high-cost military systems (Lee & Burgess, 1993).

Design of the Environment

Although this virtual environment contains a large number of sound sources, only a single source is active at any given time. Because only one source is active at a time, computing power is only needed to spatialize a single source. This is an important feature in the design of the exhibit that allows us to present a complex environment with relatively modest equipment. This design feature is also consistent with our experiences in designing acoustic displays—most users prefer quite environments, and are annoyed by constant sounds.

MODELLING THE LISTENING ENVIRONMENT

The audio processing for the exhibit includes simple models of the listening environment. Dense reverberation and atmospheric dispersion are included in the model.

Dense Reverberation and Distance Perception

The level of direct (or *dry*) sound relative to dense reverberation is particularly important to the perception of the distance of a sound source. The lower the level of direct sound, the greater the sense of distance for the listener. The amount of power received from a direct sound source varies according to the inverse square law (approximately). If direct sound is attenuated in this manner while the level of reverberant sound is kept constant, the user perceives the distance of the source.

Importance of Stereo Reverb

When generating dense reverberation, it is important that the two ears receive slightly different versions of the sound, otherwise reverberation appears to be coming from the center of the head. The reverberation in the exhibit is generated by a pair of conventional banks of allpass filters. The filter banks for the left and right ears differ slightly to create the effect of reverberation coming from outside of the head.

Atmospheric Dispersion

In most media, high frequency sound waves travel slightly faster than low frequency waves and are attenuated more severely over a distance. This effect is called dispersion. In this virtual environment, dispersion is modelled by a low-pass filter with parameters that are adjusted as the distance of the source changes.

REFERENCES

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

2. Burgess, D.A. (1992) Real-time Audio Spatialization with Inexpensive Hardware, Georgia Tech GVU Center Technical Report GVU-92-22.

3. Gaver, W.W. (1993) Using and Creating Auditory Icons, to be publish in the Proceedings of the 1993 International Conference on Acoustic Displays (ICAD '93).

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

5. Lee, M.D., Burgess, D.A., (1993) The Perception of Location Using Synthetic Auditory Localization Cues: Accuracy and the Effects of Stimulus Bandwidth, Georgia Tech GVU Center Technical Report, GVU-93-32.

6. Wenzel, E.M., Wightman, F.L., Foster, S.H. (1988) A Virtual Display System for Conveying Three-Dimensional Acoustic Information, Proceedings of the 32nd Annual Meeting of the Human Factors Society