Center for Research in Electronic Art Technology
University of California, Santa Barbara
http://www.create.ucsb.edu

Research on Spatial and Surround Sound at CREATE

Researchers at the UCSB Center for Research in Electronic Art Technology (CREATE) have been developing spatial sound performance systems and multichannel surround sound rendering software for several years. We use these systems as components of immersive user interfaces for a variety of applications, as well as for the performance of spatialized music. This white paper surveys our previous work in the field and describes our plans for the future.

Background

Today's music listener takes a multichannel spatial sound "image" for granted, whether listening to an MP3 player over headphones, playing a multimedia computer game, sitting before an audiophile music playback system, or in a concert or theater.

There are also an increasing number of computer applications that require accurately localized spatial sound for an enhanced user experience, for example *virtual environment* (VE) simulations, "eyes-free" displays, aural rendering or "auditory display" (e.g., in airplane cockpits and automobiles), and spatialized music performance. As immersive user interfaces become more common, spatial sound rendering will also become more and more an assumed component of advanced human-computer interfaces.

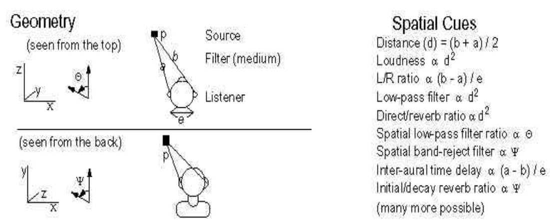
In the past, systems for multichannel sound playback were generally fixed in terms of the number of channels they supported and the format they used for representing spatial sound. Simple examples of this are traditional stereo (e.g., AIFF or WAV files), binaural headphone recording, "quad," and more recently 5.1-channel (and 7.1-channel) theatrical surround sound.

For the applications we are building at CREATE, we need more flexible and scalable spatial sound systems—including both flexible real-time surround sound rendering software, and configurable multichannel performance systems. This is the goal of our spatial sound R&D. This white paper introduces our two primary projects in the field of spatial and surround sound, and presents our plans for future research on this topic.

Surround Sound for Immersive User Interfaces and Virtual Environments

In order to use immersive user interfaces and virtual environments (VEs or "virtual reality" systems) more effectively, we need to integrate immersive output devices such as head-mounted stereo-optic displays and flexible spatialized sound rendering. Several solutions to the first problem (visual output) are available off-the-shelf with the advent of low-cost head-mounted stereo-optic displays.

For increased realism of VE worlds, we require real-time multichannel spatialized sound processing so that sounds appear to emanate from their "virtual positions" and can be tracked by the user in harmony with the visual cues provided by the optical renderers. This kind of "aural renderer" has been difficult in the past because of the different latency introduced by the visual and aural output systems, the network overhead of multichannel sound streaming, high levels of network latency jitter, and the complexity of high-quality models of sound spatialization.

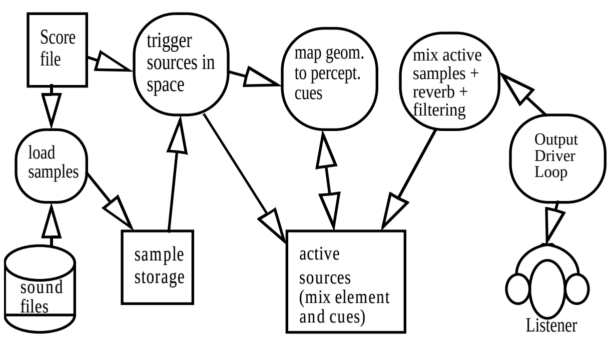


Basic Source/Listener Geometry and Psychoacoustical Spatial Cues

For maximum flexibility and scalability, we have developed a stand-alone spatial sound output system (called the CREATE auralizer, see the Web site *http://www.create.ucsb.edu/aton*), and provide several kinds of interfaces to it, including CORBA IDL, direct API calls from VE delivery systems, and a socket-based protocol for use by other programs.

The auralizer is a configurable multichannel geometry-based sound spatializer program. It plays back stored or streaming sounds and can synthesize any of a number of psychoacoustical cues that allow users to localize the virtual sound sources. The system is scalable, in that one or many auralizers can work in parallel (for the case of many sources and many speakers), and that each auralizer can dynamically shift the number of cues it synthesizes and thus the complexity of the processing it performs.

The simplest cues that the auralizer can render include the distance / amplitude cue, inter-aural amplitude difference (panning) and inter-aural time delays (precedence effect). More sophisticated scenarios can add local / global reverb, ray-tracing-derived early reflections, head-related spatial filtering (on the direct signal and optionally even the early reflections), and other more compute-intensive cues such as diffusion from (and even refraction around) solid objects.



CREATE Auralizer Architecture

The auralizer is a stand-alone program that can generate one or two channels of output for given source / listener / speaker positioning. For many-channel or many-user situations, several auralizers are used in parallel, with each one generally processing the sound for a single loudspeaker, or a stereo mix for a single listener. This allows us to scale the system up either to support many listeners as different positions in the same virtual acoustic world (as in multiuser VE systems), or to process sound for many-speaker concert situations (see below)

To achieve the flexibility we require, the auralizer is designed as a set of “pluggable” modules that perform time-domain and frequency-domain signal processing to synthesize the desired localization cues. The geometry module maps the relative positions of source and listener to a set of sound source properties; it uses parameters that can be set on a per-user or even per-source basis. This facilitates experimentation with different parameter settings for different scenarios.

The core of the auralizer is the configurable mixer / reverberator / filter that reads one or more sound sources and performs the digital mixing with a selectable set of cues being generated. The cues we use, and their weights, can all be configured and even changed at run-time, for example by a load-balancing network software manager.

We have implemented and tested several generations of auralizers, and driven them from a variety of VE systems and other input sources. We are still developing new spatial cues and refining the auralizer architecture for better scalability, easier network distribution, and more flexible I / O (see below).

The Creatophone Pluriphonic Sound Projection System

For the performance of larger-scale spatial music and sound environments, CREATE has been experimenting with “pluriphonic” (i.e., many-channel) sound performance systems for a number of years.

Creatophone Configuration 1

Our *Creatophone* system consists of anywhere between 12 and 32 loudspeakers grouped as between 8 and 20 channels (see *http://www.create.ucsb.edu/htmls/c_phone.html*).

As the figures above and below show, we have tested a number of different frontal and surround speaker configurations, we've used the vertical dimension as a parameter of spatialization, and we've used a variety of speaker types in our experiments. We install the Creatophone three times a year in the main concert hall at the UCSB campus.The research component of the Creatophone effort consists of developing the practise of large-scale surround sound performance, including the issues of wide and deep frontal "stereo" sound imaging, convincing use of height in spatialized music, and addressing the continuity of the lateral (side) spatial images in multichannel systems.

It is our goal to eventually drive the Creatophone with a high-speed network of computers running auralizer programs in real-time. this will allow composers to produce their music independent of the actual speaker configuration used in the concert. The auralizer programs will run from the pre-determined geometrical description of the music, and will mix and process the sounds at performance-time.

Creatophone Configuration 2

Spatial Sound Renderers and Auralizers

To allow game and VE developers, as well as music composers to use such flexible spatial sound tools, we need flexible editing and playback systems that represent sound mixing and spatialization parameters on their pure geometric form. This way the run-time surround sound renderer can choose the appropriate signal processing parameters and localization cues to reproduce; the same "medium" (consisting of sound content and geometry) can be used to play back over everything from stereophonic headphones to a 32-channel Creatophone set-up, including 5.1- and 7.1-channel home theater systems.

Research Issues in Spatial Sound

Spatial and surround sound are the subjects of much current research. The project topics range from applications-oriented system development projects to on-going basic research in human auditory perception. There have been several specialized conferences on spatial sound and auditory display in recent years (for example, the *SoundInSpace* event held at UCSB in March, 2000). A comprehensive history of "sound in space" written by CREATE's Professor Curtis Roads will be appearing soon in *Computer Music Journal.*

At CREATE, we focus our efforts on a small number of special issues that will allow us to build better systems for spatial sound performance, either in the context of small-scale computer-based applications such as virtual environments, or for larger, concert performance installations.

One of our current interests is in implementing scalable spatial audio rendering on off-the-shelf hardware (i.e., desk-top PCs). The current auralizer mixing and processing system is time-domain-oriented, meaning that the reverberators and spatial filters are all implemented as time-domain tapped delay lines. These have the advantage of good scalability and easy distribution (over a multiprocessor network). The problem is that, for complex ray-tracing or spatial filters, the delay lines become impossible to compute in real-time, and that they are difficult to split across processors in a network.

For large-scale systems, we are designing a frequency-domain auralizer, where the delay lines will be carried out by frequency-domain convolution using the FFT / IFFT process. This will allow us to better scale system performance and support even higher-grade reverberators and filters

The other research area we are concentrating on is how to represent spatial audio in such a format that we can pre-process as much as possible, but still render the spatialization of the sound (be it speech, sound effect, or music) at “performance” time. This would allow us to support a variety of output formats (ranging from 1 to 32 channels) for playback over a range of configurations including headphones, home theater, automobile sound, and concert-hall situations.

Plans for Future Work

Based on our recent breakthroughs with the auralizer and the Creatophone, we plan to continue and increase our R&D efforts on spatial and surround sound. Our current project plans incorporate the following themes:

- Flexible representations of spatialized sound for reproduction over 1-, 2-, 4-, 5.1-, 7.1-, 8-, 16-, and 32-channel systems;
- Scalable distributed auralizer geometry and processing architecture;
- Synthesis of new psychoacoustical cues such as refraction and non-linear reflection;
- Software for the configuration and management of many-channel playback systems (see the HPDM project and DPE white-paper);
- Integration of low-latency spatial sound in multimedia virtual environments;
- Efficient ray-tracing for calculating early reflections in complex rooms;
- Vector processing for spatial filters and dynamic reverberators (e.g., using MMX- or AltiVec-enabled processors);
- Pure frequency-domain techniques for sound representation, reverberation, and filtering;
- Spatial sound processing for many-channel real-time performance systems;
- Integration of auralizer technology into real-time sound synthesis languages.

Contact Information

Stephen Travis Pope
Email: stp@create.ucsb.edu
Tel: (805) 967-2621

Updated 2001/3/16