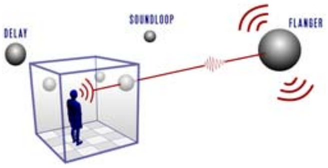
User-specific audio rendering and steerable sound for distributed virtual environments

Mike Wozniewski & Zack Settel
La Société Des Arts Technologiques
Montréal, Québec, Canada
mikewoz@sat.qc.ca zack@sat.qc.ca

Jeremy R. Cooperstock
McGill University Centre for Intelligent Machines
Montréal, Québec, Canada
jer@cim.mcgill.ca

Overview

- **Multiple users immersed in a virtual environment**
  (each having a subjective visual and auditory rendering of the scene)
- **Spatialization using a virtual microphone technique**
  (the # of virtual mics corresponds to the # of loudspeakers being used)
- **Realtime physical modelling of virtual audio**
  (used to provide the necessary signals to the virtual microphones)
- **The scene also contains other sonic objects**
  (which perform localized DSP at specific 3-D locations)
- **Users can 'steer' sound through 3-D space**
  (allows for spatially organized signal processing)



The Spatial Audio Framework

What is it?

A 3-D engine built using OpenSceneGraph (to manage arrangement of objects in space), and PureData (to manage DSP, input devices, and audio hardware). **It allows for dynamic creation, modification and control of virtual audio scenes, which can function as musical instruments or interactive sonic applications.**

What is a "Virtual Audio Scene"?

- Composed of sound processors called ***soundNodes*** and ***soundSpaces*** which perform DSP computation at specific 3-D locations.
- Nodes are organized in a ***scene graph*** (a tree-like data structure where geometric operations applied on a node are automatically propagated to all children).
- **Physical modelling** between nodes and spaces must be explicitly instantiated using ***soundConnections***.
- Sound is steered through 3-D space according to ***rolloff functions*** and ***steering vectors***.

Elements of a Virtual Audio Scene

The '*soundNode*':

The *soundNode* **is the fundamental building block** of a virtual audio scene. It can behave as **either a source, or sink, or both** at the same time. The case where it represents "both" is interesting since this is how spatially localized DSP is realized. The node will collect audio at a specific 3-D location, apply some processing, and radiate the result back into the scene.

The '*soundConnection*':

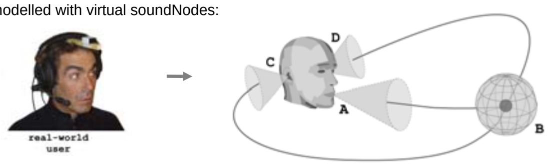
The *soundConnection* **specifies the propagation model** between soundNodes. That is, it defines *how* audio travels from a source node to a sink, based on physically-modelled audio propagation. When constructing DSP applications, soundConnections perform a similar function to traditional patch cords that used in sound studios or even those found in patcher-based audio software like Max/MSP.

The '*soundSpace*':

The soundSpace **provides volumetric processing** rather than the localized processing of a soundNode. These nodes are typically defined by some 3-D model (exported from Maya, 3D Studio Max, Blender, etc.), and capture sound from nodes within the bounding contour. These nodes are useful for simulated acoustic effects such as reverberation, but can also be used to define spatial regions of processing.

Representation of the User

Below we see how a user can be modelled with virtual soundNodes:



Source node A represents the user's microphone and emits sound into the virtual world. Node B absorbs some of this sound, performs DSP, and emits the resulting signal back into the scene, thus acting as both a sink and a source. Nodes C and D are sink nodes that can be thought of as virtual microphones. They are positioned in virtual space corresponding to the location of real world loudspeakers (in this case, the two loudspeakers of the headset). If the user was using a 5.1 channel speaker arrangement, then there would be 6 virtual mics in the scene (the centre mic would have 0º rotation, the right mic would have 30º, etc.).

Applications

3-D Audio Mixing

- A multi-track recording can be distributed throughout a virtual scene using soundNodes to radiate each track from a specific 3-D location.
- Users (re)mix the tracks by changing their relative proximity and orientation to the soundNodes.
- The particular format of the mix (5.1, stereo, etc.) is determined by the number of virtual microphones, lending to the possibility of mixdowns.
- Effects boxes can also be integrated, and sound can be 'sent' or 'bussed' to those processing nodes before being captured.

Next-generation Video Conferencing

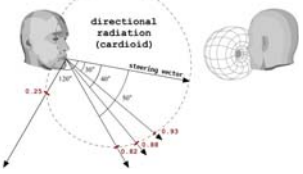
- Multiple simultaneous conversations can occur.
- Users can steer their listening or talking to participate in any of the ongoing discussions
- The image below also shows integrated video from a webcam above the avatar.

Virtual Musical Performance

- A performer's instrument can be modelled as a source node surrounded by many effects nodes such as reverb, delay, flanging, or ring modulation.
- We equip the instrument (e.g. saxophone) with an orientation sensor and a wireless microphone
- The performer points the instrument in a certain direction, radiating sound to one or more effects nodes.
- "Rolling" the horn tightens/spreads the radiation.
- Musicians can share effects, leading to potentially collaborative instruments.

Steerable Sound

One powerful feature of our framework is the **control of directivity** using rolloff functions.



NOTE: this figure only shows one rolloff. Each soundNode in fact has **two** steering vectors and rolloff functions: one for the souce direction, and one for the sink direction.

A **rolloff function** provides gain values given an angle of incidence (α) to the steering vector, with a **spread parameter** (γ) to tighten/widen the rolloff.

eg, a cardioid function:

$$\text{cardiod}(\alpha) = \left[\frac{(1+\cos(\alpha))}{2}\right]^{\gamma}$$

Physical Simulation ... and bending the rules

Sound that travels through the virtual scene is physically modelled to simulate phenomena such as:

- exponential decay of energy during travel
- travel time according to the speed of sound
- diffraction of low frequencies around volumes
- absorption of high frequencies as a function of distance
- etc.

One important feature however, is the ability to ***bend* the rules of physics** that govern the propagation of sound.

For artistic or musical purposes, this allows for interesting results that are impossible with traditional audio simulation technology.

For example:

- decay & delay of sound can be diminished in order to 'teleport' sound from one place to another
- Doppler shift can be eliminated to preserve the tonal aspects of a musical piece
- sound can be 'steered' in a precise direction instead of propagating as a spherical wavefront

References

1. M. Wozniewski, Z. Settel, and J. R. Cooperstock, "A framework for immersive spatial audio performance," in New Interfaces for Musical Expression (NIME), pp. 144-149, 2006.
2. M. Wozniewski, Z. Settel, and J. R. Cooperstock, "A paradigm for physical interaction with sound in 3-D audio space," in International Computer Music Conference (ICMC), pp. 456-463, 2006.
3. D. R. Begault, 3-D sound for virtual reality and multimedia, Academic Press Professional Inc., 1994.
4. W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," J. Acoustics Soc. of America, vol. 97, no. 6, pp. 3907–3908, 1995.
5. M. A. Gerzon, "Periphony: With-height sound reproduction," J. Audio Eng. Soc., vol. 21, no. 1, pp. 2–10, 1973.
6. V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," J. Audio Eng. Soc., vol. 45, no. 6, pp. 456–466, 1997.
7. A. Mouchtaris, S. S. Narayanan, and C. Kyriakakis, "Virtual microphones for multichannel audio resynthesis," EURASIP Journal on Applied Signal Processing, vol. 2003, pp. 968–979, 2003.
8. Jonas Braasch, "A loudspeaker-based 3D sound projection using Virtual Microphone Control (ViMiC)," Convention of the Audio Eng. Soc. 118, pp. 968–979, 2005.
9. AccessGrid," www.accessgrid.org.

McGill Centre for Intelligent Machines

Société des arts technologiques
Society for Arts and Technology