EFFICIENT DESCRIPTION AND RENDERING OF COMPLEX INTERACTIVE ACOUSTIC SCENES

Jean-Marc Jot

Creative Advanced Technology center
Scotts Valley, USA
jmj@atc.creative.com

ABSTRACT

Interactive environmental audio spatialization technology has become commonplace in personal computers and is migrating into portable entertainment platforms (including cell phones) and multi-player game servers (virtual online worlds). While the primary current application of this technology is 3D game sound track rendering, it is ultimately necessary in the implementation of any personal or shared immersive virtual world ("virtual reality"). The successful development and deployment of such applications in new mobile or online platforms involves maximizing the plausibility of the synthetic 3D audio scene while minimizing the computational and memory footprint of the audio rendering engine. It also requires a flexible, standardized scene description model to facilate the development of applications targeting multiple platforms. This paper reviews a computationally efficient 3-D positional audio and spatial reverberation processing architecture for real-time virtual acoustics over headphones or loudspeakers, compatible with current interactive audio standards (including MPEG-4, OpenAL, JSR 234 and OpenSL ES).

1. INTRODUCTION AND OVERVIEW

The applications of interactive 3D audio technologies include simulation and training, telecommunications, video games, multimedia installations, movie or video soundtracks, and computer music [1]-[5].

Virtual acoustics technology has its origins in research carried out in the 1970's, which targeted two distinct applications:
- *Architectural acoustics:* Schroeder et al. developed simulation methods based on geometrical acoustics to derive a computed echogram from a physical model of room boundaries and the source and listener positions [6];
- *Computer music:* Chowning developed a 4-channel spatialization system for simulating dynamic movements of sounds, which provided direct control of two perceptual control parameters for each source: apparent direction of sound arrival and apparent distance to the listener, along with a derived Doppler shift [7]. Artificial reverberation was included to enhance the robustness of distance effects. Later, Moore proposed an extension of this approach where early reflections were controlled indirectly via a geometrical acoustic model [8].

Interactive virtual acoustics systems require real-time rendering and mixing of multiple audio streams (sound sources) to feed a set of loudspeakers or headphones. This rendering system is driven by an acoustic scene description model which provides positional and environmental audio parameters for all sound sources.

The scene description represents a virtual world including sound sources and one or more listeners within an acoustical environment which may incorporate one or more rooms and acoustic obstacles.

Standardization is essential for enabling platform-independent playback and re-usability of scene elements by application authors and sound designers. Current standard interactive audio scene description models include high-level scripting languages such as the MPEG-4 Advanced Audio Binary Format for Scene description (AABIFS) [9] and low-level application programming interfaces used in the creation of video games, such as OpenAL, JSR 234 and OpenSL ES [10]-[12]. In this paper, we will consider a generic, low-level scene description model based on OpenAL [10] and its environmental extensions, I3DL2 [13] and EAX [14]-[15]. For applications that require higher-level world representations, a real-time translation software layer can be implemented above the rendering engine to convert the high-level representation to low-level description parameters [14].

In the first section of this paper, we discuss and compare digital signal processing methods for computationally efficient real-time spatialization of multiple sound sources over headphones or loudspeakers. This includes discrete amplitude panning, Ambisonic and binaural or transaural techniques [16]-[25] and introduces a recently developed multi-channel binaural synthesis method based on discrete spatial functions, previously introduced in [26]. The description model and rendering methods are then extended to include the acoustic effects of the listener's immediate environment. This includes the effects of acoustic obstacles and room boundaries or partitions on the perception of each sound source. Acoustic reflections and room reverberation are rendered by use of feedback delay networks [27]-[30]. A statistical reverberation model, previously introduced in [30], is included for modelling per-source distance and directivity effects. We further extend the model to account for the presence of acoustic environments or rooms adjacent to the listener's environment. An efficient spatial reverberation and mixing architecture, previously introduced in [26], is described for the spatialization of multiple sound sources around a virtual listener navigating across multiple connected virtual rooms. This processing architecture includes a novel cost-efficient method for simulating multiple spatially extended sound sources or sound events.

The models and methods reviewed in this paper enable the realization of comprehensive, computationally efficient, flexible and scalable high-quality interactive 3D audio rendering systems for deployment in a variety of consumer appliances (ranging from personal computers to home theater and mobile entertainment systems) and services (including multi-user comunication and telepresence).

2. REFERENCES

[1] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic Press, New York, 1994).

[2] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization - an Overview," *J. Audio Eng. Soc.* 41(11): 861-875 (1993 Nov.).

[3] M. Cohen and E. Wenzel, E, "The Design of Multidimensional Sound Interfaces," Tech. Rep. 95-1-004, Human Interface Laboratory, Univ. of Aizu (1995).

[4] J.-M. Jot, "Real-time Spatial processing of sounds for music, multimedia and interactive human-computer interfaces," *ACM Multimedia Systems J.* 7(1) (1999 Jan.).

[5] A. Harma & al., "Augmented Reality Audio for Mobile and Wearable Appliances," *J. Audio Eng. Soc.* 52(6): 618-639 (2004 June).

[6] M. R. Schroeder, "Computer Models for Concert Hall Acoustics," *American J. Physics* 41:461-471 (1973).

[7] J. Chowning, "The Simulation of Moving Sound Sources.," *J. Audio Eng. Soc.* 19(1) (1971).

[8] F. R. Moore, "A General Model For Spatial Processing of Sounds," *Computer Music J.* 7(6) (1983)

[9] R. Väänänen and J. Huopaniemi, "Advanced AudioBIFS: Virtual Acoustics Modeling in MPEG-4 Scene Description," *IEEE Trans. Multimedia* 6(5): 661-675 (2004 Oct.)

[10] G. Hiebert & al., "OpenAL 1.1 Specification and Reference," www.openal.org (1995 June)

[11] M. Paavola & al., "JSR 234: Advanced Multimedia Supplements," Java Community Process spec. www.jcp.org (2005 June).

[12] Khronos Group, "OpenSL ES – Open Standard Audio API for Embedded Systems", www.khronos.org/opensles (2007).

[13] J.-M. Jot & al., "IA-SIG 3D Audio Rendering Guideline, Level 2" (I3DL2) www.iasig.org (1999).

[14] J.-M. Trivi and J.-M. Jot, "Rendering MPEG-4 AABIFS Content Through a Low-level Cross-platform API," Proc Int. Conf. Multimedia (ICME 2002).

[15] J.-M. Jot and J-.M. Trivi, "Scene Description Model and Rendering Engine for Interactive Virtual Acoustics," Proc. 120th Conv. Audio Eng. Soc., preprint 6660 (2006 May).

[16] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.* 45(6): 456-466 (1997 June).

[17] M. A. Gerzon, "General Metatheory of Auditory Localization," Proc. 92nd Conv. Audio Eng. Soc., preprint 3306 (1992).

[18] M. A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video," *J. Audio Eng. Soc.* 33(11) (1985).

[19] D. G. Malham and A. Myatt, "3-D Sound Spatialization Using Ambisonic Techniques," *Computer Music J.* 19(4) (1995).

[20] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Audio Eng. Soc.* 37(1/2) (1989).

[21] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony," Proc. 98th Conv. Audio Eng. Soc., preprint 3980 (1995).

[22] W. G. Gardner, *3-D Audio Using Loudspeakers*, Ph. D. Thesis, Massachussets Institute of Technology (1997),

[23] A. Jost and J.-M. Jot "Transaural 3-D Audio with User-controlled Calibration," *Proc. Int. Conf on Digital Audio Effects* (DAFX 2000).

[24] J.-M. Jot, V. Larcher, and J.-M. Pernaux, "A Comparative Study of 3-D Audio Encoding and Rendering Techniques," *Proc. 16th Int. Conf. Audio Eng. Soc.* (1999 March).

[25] V. Larcher, J.-M. Jot, G. Guyard, and O. Warusfel, "Study and Comparison of Efficient Methods for 3-D Audio Spatialization Based on Linear Decomposition of HRTF Data", Proc. 108th Conv. Audio Eng. Soc., preprint 5097 (2000 Jan.).

[26] J.-M. Jot, M. Walsh and A. Philp, "Binaural Simulation of Complex Acoustic Scenes for Interactive Audio," Proc. 121st Conv. Audio Eng. Soc., preprint 6950 (2006 Oct.).

[27] J.-M. Jot, "Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality", *Proc. International Computer Music Conference* (1997).

[28] W. G. Gardner, "Reverberation algorithms," *Applications of Signal Processing to Audio and Acoustics* (ed. M. Kahrs, K. Brandenburg), Kluwer Academic (1998).

[29] L. Dahl and J.-M. Jot, "A Reverberator Based on Absorbent All-pass Filters", *Proc. Int. Conf on Digital Audio Effects* (DAFX 2000).

[30] J.-M. Jot, L. Cerveau, and O. Warusfel, "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model," Proc. 103rd Conv. Audio Eng. Soc., preprint 4629 (1997 Aug.).