Interactive 3D Audio Rendering in Flexible Playback Configurations

Jean-Marc Jot
DTS, Inc. Los Gatos, CA, USA
E-mail: jean-marc.jot@dts.com Tel: +1-818-436-1385

Abstract— **Interactive object-based 3D audio spatialization technology has become commonplace in personal computers and game consoles. While its primary current application is 3D game sound track rendering, it is ultimately necessary in the implementation of any personal or shared immersive virtual world (including multi-user communication and telepresence). 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 facilitate the development of applications targeting multiple platforms. This paper reviews a general computationally efficient 3D positional audio and environmental reverberation rendering engine applicable to a wide range of loudspeaker or headphone playback configurations.**

I. INTRODUCTION

The applications of interactive object-based 3D audio rendering technology 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 to feed a headphone or loudspeaker system. The rendering engine is driven by an acoustic scene description model which provides positional and environmental audio parameters for all sound sources (or "audio objects"). 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. The standardization of scene description software interfaces is necessary to enable platform-independent playback and re-usability of scene elements by application developers and sonic artists. Standard object-based scene description models include high-level scripting languages such as the MPEG-4 Advanced Audio Binary Format for Scene description (AABIFS) [9] [10] and low-level application programming interfaces used in the creation of video games, such as OpenAL, OpenSL ES and JSR 234 [11]–[13]. In this paper, we shall consider a generic, low-level scene description model based on OpenAL [11] and its environmental extensions, I3DL2 and Creative EAX [14] [15]. For applications that require higher-level virtual world representations, a real-time translation software layer can be implemented above the rendering engine interface to convert the high-level representation to low-level description parameters [16].

In the next section of this paper, we review digital signal processing methods for the spatialization of multiple sound sources over headphones or loudspeakers, including discrete amplitude panning, Ambisonic and binaural or transaural techniques [17]–[25]. We then review recently developed computationally efficient multichannel binaural synthesis methods based on the decoupling of spatial and spectral filtering functions, previously introduced in [26] [27]. The description model and rendering methods are then extended to include the acoustic effects of the listener's immediate environment, including the effects of acoustic obstacles and room boundaries or partitions on the perception of audio objects. Acoustic reflections and room reverberation are rendered by use of feedback delay networks [28]–[31]. A statistical reverberation model, previously introduced in [31], is included for modeling 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. A general object-based rendering engine architecture, previously introduced in [27], is described. It realizes the spatialization of multiple audio objects around a virtual listener navigating across multiple connected virtual rooms, and includes an efficient "divergence panning" method for representing and rendering spatially extended audio objects.

The models and methods reviewed in this paper enable the realization of comprehensive, computationally efficient, flexible and scalable 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 communication and telepresence).

II. POSITIONAL AUDIO RENDERING TECHNIQUES

It is advantageous to adopt a scene description model that is independent from the spatial rendering technique or the playback setup used, so that the audio scene may be authored once and adequately reproduced on any end-user platform. Source and listener positions are therefore typically described via 3D coordinates in a Cartesian or polar system (rather than in terms of format-specific per-channel gain scaling factors, for instance). In this section, we provide an overview of signal processing methods for the real-time spatialization of sound sources over headphones or loudspeakers. More detailed reviews and comparisons can be found in [4] [25] [32].

Referring to Figure 1, a general problem statement may be formulated as follows [25]: design a directional encoding/decoding method that takes a monophonic source signal input an produces a multichannel signal to feed a set of loudspeakers, such that an ideal omnidirectional microphone placed at the position of the listener would capture the same monophonic signal, but the sound appears to emanate from a point source located in a specified direction. The default idealized assumption is that the sound source is located in far field (plane wave reconstruction). However, the controlled simulation of near-field sources may be useful in some applications (video games, virtual or augmented reality). A vector-based formulation was proposed by Gerzon [18] for predicting the perceived direction on the reproduced sound, and shown in [33] to coincide with the active sound intensity vector of the reconstructed sound field.

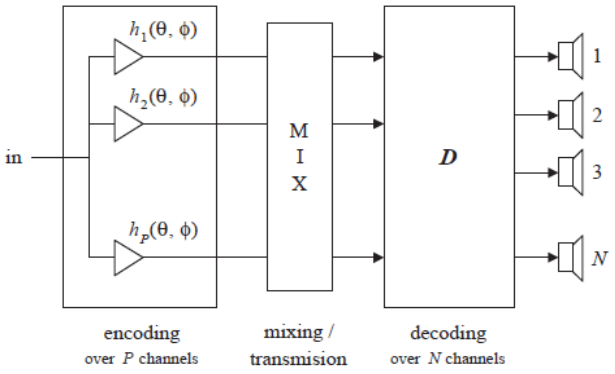

Fig. 1 General 3D positional audio encoding/decoding framework. In the directional encoder (or "panner"), the blocks h_i may be simple scaling coefficients or, more generally, direction-dependent linear filters which can emulate the directivity characteristics of an array of transducers, and include delays in the case of non-coincident microphone recording techniques. The decoding matrix is optional, necessary when the mixing or transmission format is different from the reproduction format.

The pair-wise amplitude panning principle originally proposed by Chowning over four loudspeakers [7] may be generalized to arbitrary 2D or 3D loudspeaker array configurations by use of the Vector-Based Amplitude Panning (VBAP) method [17]. This method is recast in [25] as reconstructing Gerzon's direction vector (and, consequently, the active intensity vector) by local interpolation over the nearest loudspeakers. In contrast, the Ambisonic method [19] [20] aims at reconstructing this direction vector by combining contributions from all the loudspeakers. Both methods require at most four gain multipliers per source (or audio object). The Ambisonic technique may also be viewed as a sound-field reconstruction in the immediate vicinity of the reference listening point. As the number of loudspeaker channels is increased, this approach may be extended to higher-order sound-field reconstruction covering a larger listening area. In this respect, it then becomes comparable to the Wave-Field Synthesis (WFS) method [34]–[37]. Both WFS and Ambisonics enable the reproduction of audio objects located inside the perimeter delimited by the loudspeakers [38] [39].

3D audio reproduction over headphones calls for a binaural synthesis process where head-related transfer functions (HRTF) are modeled as digital filters to reproduce the due pressure signals at each ear. When this process is implemented directly in the form of a pair of HRTF filters for each audio object, the incremental cost per object is typically 10-20 times that of a discrete or Ambisonic panning technique [22]. However, as discussed in the next section, this discrepancy can be resolved via reformulations employing a common bank of HRTF filters for all sources [26]. Convincing near-field effects may be simulated by appropriately adjusting the inter-aural level difference for laterally localized sound sources [27]. By use of "transaural" cross-talk cancellation techniques, a binaural 3D audio scene may be delivered to a single listener over two frontal loudspeakers [21] [22]. For a more robust reproduction of rear or elevated sources, this technique can be extended to systems of four or more loudspeakers using multiple cross-talk cancellers. The stringent constraint on the listener's position may be relaxed by employing band-limited inter-aural transfer function models in the cross-talk canceller design [23]. This approach can be extended in order to achieve a behavior equivalent, at high frequencies, to that of a pair-wise amplitude panner [24].

III. MULTICHANNEL BINAURAL ENCODING

Malham [40] and Travis [41] proposed a binaural encoding method that drastically reduces the incremental computational cost per sound source, while circumventing HRTF filter interpolation and dynamic positioning issues. It consists of combining a multichannel panning technique for N-channel loudspeaker playback (in their case, Ambisonics) and a bank of static binaural synthesis filter pairs each simulating a fixed direction (or "virtual loudspeaker") over headphones, as illustrated in Figure 2. In terms of reproduction fidelity, this approach suffers from the inherent limitations of the

multichannel directional encoding techniques used. It requires a large number of encoding channels to faithfully reproduce the localization and timbre of sound signals in any direction. A more accurate variant of this approach, shown in Figure 3, consists of explicitly reproducing the position-dependent inter-aural time difference (ITD) and reconstructing the set of position-indexed HRTF filters by decomposition over a set of spatial functions $\{g_i\}$ and a set of reconstruction filters $\{(L_i, R_i)\}$.

$$L(\theta, \varphi, f) \cong \sum_{\{i = 0,\ldots N-1\}} g_i(\theta, \varphi)\, L_i(f),$$
$$R(\theta, \varphi, f) \cong \sum_{\{i = 0,\ldots N-1\}} g_i(\theta, \varphi)\, R_i(f). \qquad (1)$$

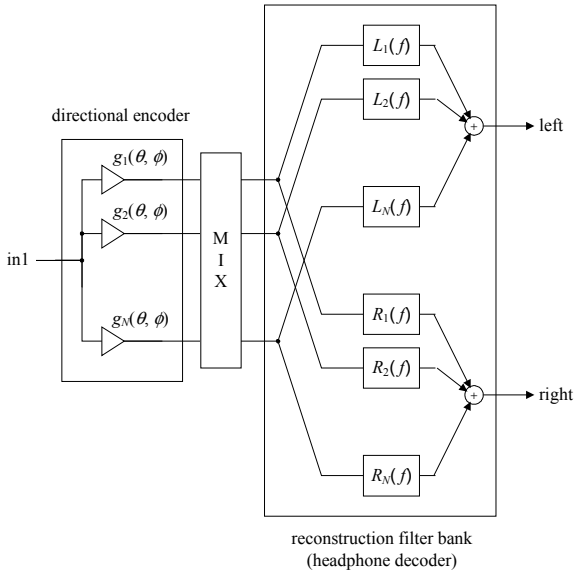

Fig. 2 Binaural encoding by multichannel panning and post-filtering.

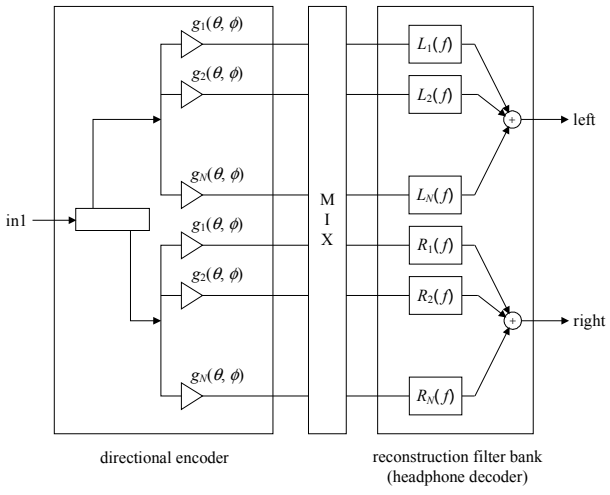

Fig. 3 Multichannel binaural encoding with per-source ITD encoding.

This HRTF data decomposition model, originally proposed by Chen et al. [42], is reviewed and discussed extensively in [26] [27], where several mathematical approaches to the joint optimization of the spectral and spatial functions are reviewed. The use of predetermined spatial functions in this context was proposed in [43] and further investigated in [25]–[27]. In this variant, the set of spatial functions $\{g_i\}$ is chosen a priori and the spectral functions (reconstruction filters) are derived by projection of the HRTF data set onto the spatial functions. In [43], spherical harmonics were proposed as the basis of predetermined spatial functions, resulting in an encoding format termed 'Binaural B Format'. This format is a superset of the standard (first-order) Ambisonic B Format and enables recordings using available microphone technology that preserve ITD cues when decoded over headphones or loudspeakers.

In [27], a multichannel binaural encoding scheme based on predetermined discrete panning functions is introduced, which has the following advantages:

- exact HRTF reproduction of selected principal directions and controlled accuracy vs. complexity trade-off
- computational efficiency (minimizing the number of non-zero panning weights for each audio object)
- advantageous decoding scheme for robust directional reproduction over loudspeakers.

Figure 4 shows the set of discrete panning functions derived by applying the VBAP method of [17] for a horizontal-only encoding system.

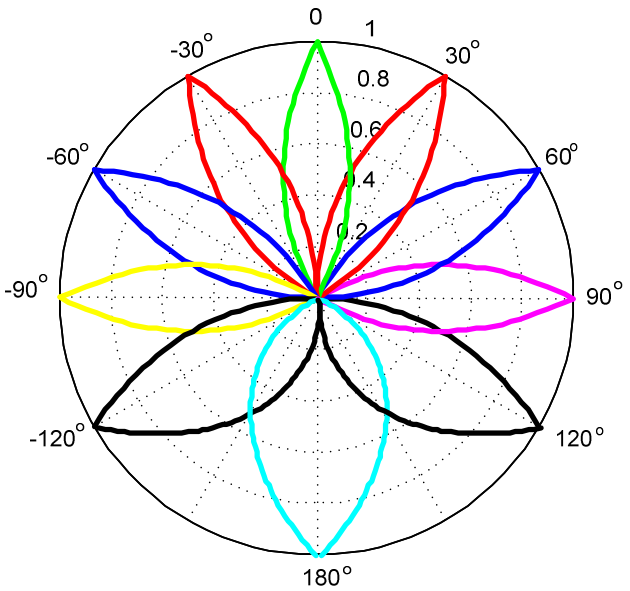

Fig. 4 Discrete multichannel horizontal-only panning functions obtained by the VBAP method for the principal direction azimuths {0, ±30, ±60, ±90, ±120, 180 degrees}.

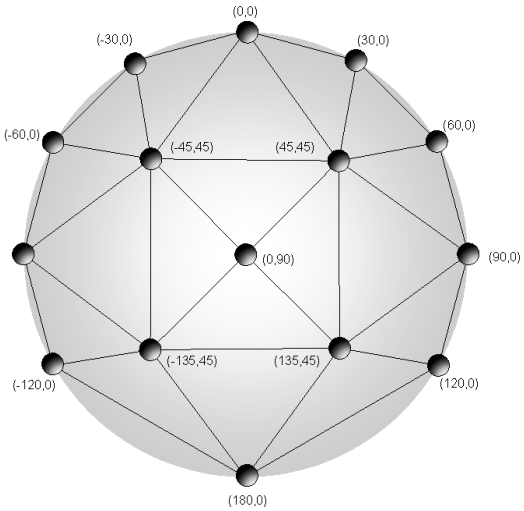

Fig. 5 Example principal direction set for 3D multichannel binaural rendering.

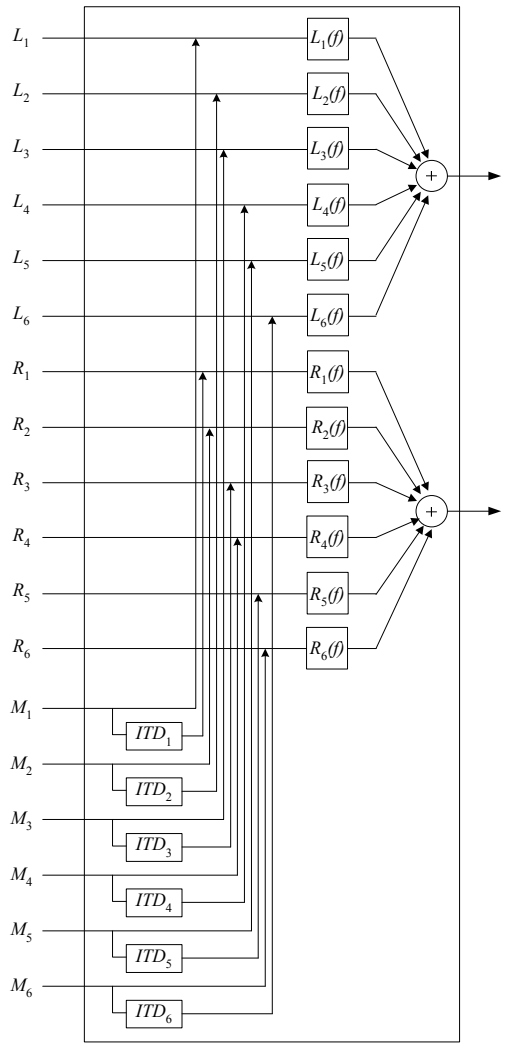

Fig. 6 Hybrid multichannel binaural decoder including additional input bus in standard multichannel format, assuming 6 principal directions.

Figure 5 shows an example of selected principal directions for 3D encoding, designed to emphasize the accuracy of reproduction in the horizontal plane and to include the conventional channel positions for reproduction of multichannel audio sources provided in the standard 5.1, 6.1 or 7.1 formats [32]. This is advantageous in order to implement, at minimum computational expense, a hybrid multichannel binaural decoder including a standard multichannel input bus (denoted $\{M_i\}$ in Figure 6) suitable for the binaural reproduction of multichannel recordings. Decoder designs leveraging this property for 3D audio reproduction over two or four loudspeakers using cross-talk cancelation are described in [27].

IV. Environmental Audio Rendering

In this section, we review a low-level environment reverberation rendering model including the sound muffling effects associated with acoustic diffraction and transmission by room partitions. As mentioned in introduction, we assume a low-level scene description as proposed in the OpenAL or OpenSL ES APIs and in the I3DL2/EAX extensions [11] [12] [14] [15]. The task of mapping the virtual world description to this low-level scene description is left to the application developer or to a higher-level physical modeling engine. This approach is adequate in program-driven virtual reality, simulation or 3D gaming applications, where it offers more flexibility to the sound designer or programmer than would a rigid physically-based scene description model, for instance. In a data-driven interactive audio application where a user interacts with a virtual physical world constructed offline or remotely, a high-level geometrical/physical scene description of this world is necessary to enable the rendering of acoustic obstacle effects consistent with a concurrent visual presentation of the virtual world.

The I3DL2 rendering guideline [14] includes a parametric reverberation model shared by all audio objects in the scene, along with a set of parameters that describe per-object effects such as the muffling effects of obstruction or occlusion by obstacles or partitions standing between source and listener. The reverberation impulse response (Figure 7) is divided into three temporal sections: the direct path ("Direct"), the early reflections ("Reflections"), and the exponentially decaying late reverberation ("Reverb").

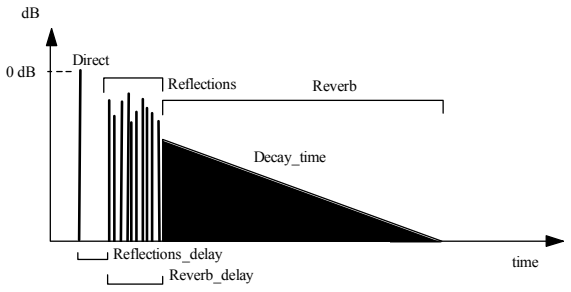

Fig. 7 I3DL2 / EAX generic reverberation model.

The I3DL2 reverberation response model is characterized by the following parameters:

- the energy and time divisions of the early and late reverberation sections
- the reverberation decay time at mid an high frequencies
- the diffusion (echo density) and the modal density of the late reverberation
- a low-pass filter applied to the Reflections and Reverb components ("Room" filter).

In EAX, this reverberation model is extended as follows:

- low-frequency control for the Reverb decay time and for the "Room" filter
- Reflections and Reverb panning vectors (providing adjustable direction and spatial focus)
- Reverb echo and pitch modulation parameters.

Figure 8 shows a mixing architecture for rendering multiple sound sources located in the same virtual room, where early reflections are rendered separately for each source while reverberation is rendered by an artificial reverberator shared among all sources [4] [8]. Rendering the reverberation decay tail as a superposition of discrete reflections for each source would be impractical from a computational standpoint.

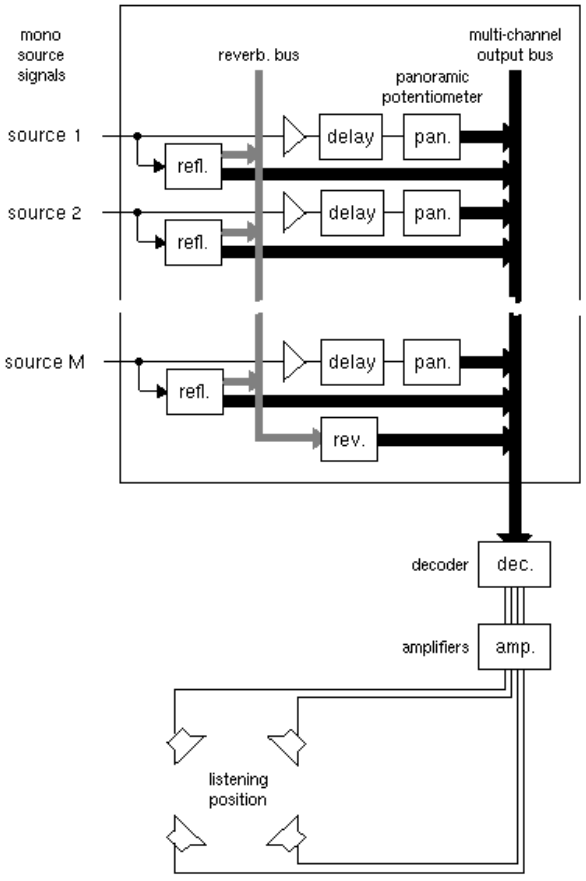

Fig. 8 Mixing architecture for reproducing multiple virtual sound sources located in the same virtual room, with per-source early reflection rendering.

Sharing a common reverberator among all sources is justified by a statistical model of room reverberation decay as a global exponentially decaying stochastic process, characterized by a reverberation decay time and an initial energy level, both frequency-dependent [31]. This model also predicts the energy contribution of each source to the reverberation according to source directivity and distance, as illustrated in Figure 9. It is based solely on general physical laws of room acoustics and thus perceptually plausible, yet it avoids direct reference to the geometry and acoustical properties of walls or obstacles.

In addition to position, orientation and directivity parameters, the low-level per-source parameters in I3DL2/EAX include low-pass filter properties for modeling muffling effects (Figure 10):

- an *Obstruction* filter applied to the direct path only, for rendering the muffling effect of an obstacle standing between the source and the listener located in the same room;
- an *Occlusion* filter applied to both the direct path and the reverberation path, for rendering the muffling effect of a partition between the source and the listener located in different rooms;
- an *Exclusion* filter applied to the reverberation path only, to simulate a source heard through an opening.

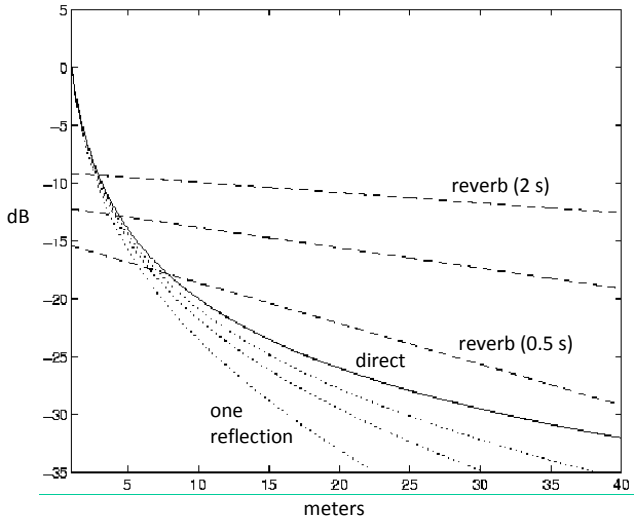

Fig. 9 Decay of intensity vs. distance according to the statistical reverberation model [31], for an omnidirectional source, a room volume of 5000 m3 and a reverberation time of 2.0, 1.0 or 0.5 s. The intensity of the diffuse reverberation or of an individual reflexion decays faster for a shorter reverberation time.

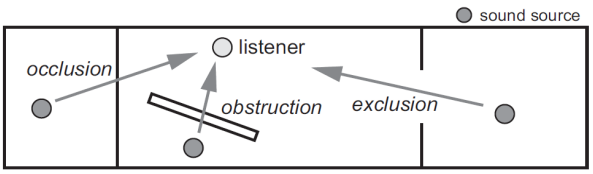

Fig. 10 I3DL2 / EAX per-source environmental muffling effects.

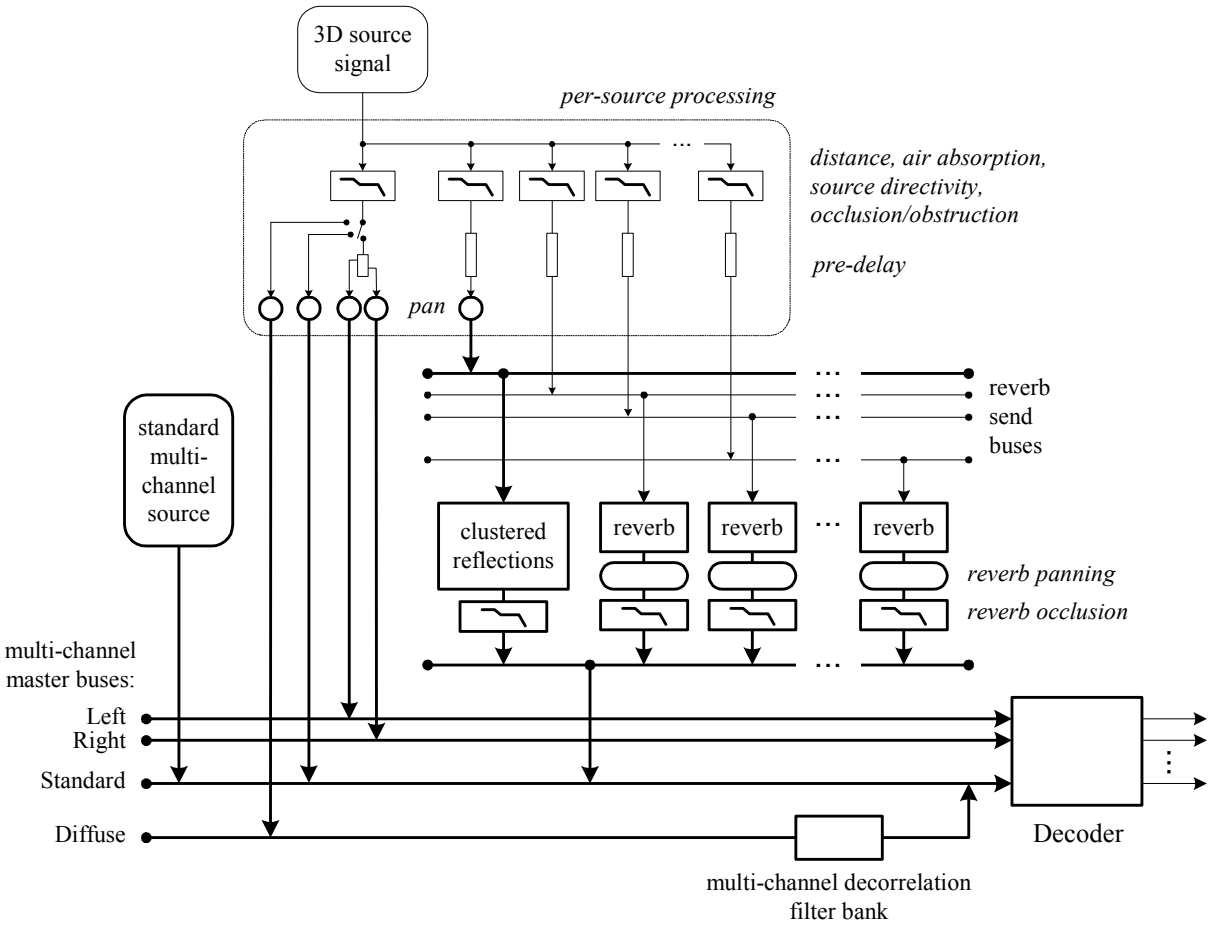

Fig. 11 Overview of complete multi-environment rendering engine. Thicker lines indicate multichannel signal paths.

V. General object-based interactive 3D audio renderer architecture

In this section, we combine and extend the models described previously to realize a general multi-environment rendering engine for reproducing complex interactive virtual acoustic scenes including the effects of environmental reflections and reverberation, acoustic obstacles, and source directivity, orientation and size. Figure 11 shows the complete binaural reverberation and positional audio engine, previously introduced in [27]

Each audio object waveform feeds a per-source processing module including a direct-path processing section and a reflected-path processing section. As shown on the left-hand side on Figure 11, the direct path is directionally encoded and mixed into one of three multichannel master output buses:

- *Left*, *Right*: binaural multi-channel encoding with per-source ITD (as described in Figure 3)
- *Standard*: multichannel panning (per Figure 2)
- *Diffuse*: "divergence panning" for a spatially extended source (see section VI).

In order to achieve a convincing simulation of a natural environment, a 3D gaming or virtual reality system may need to render more than a hundred different sound sources simultaneously. If one or more discrete acoustic reflections are included in the rendering of each audio object, the total number of virtual sound sources can reach several hundreds. In a resource-constrained binaural renderer implementation, it is advantageous to use a hybrid multichannel binaural decoder such as shown in Figure 6, so that only a subset of all audio objects are encoded in the high-accuracy binaural multichannel format (directional encoder of Figure 3 including per-source ITD synthesis). The other sources are encoded in the standard multichannel format using the directional encoder of Figure 2, which requires only half as many multipliers and does not include an additional delay unit for ITD synthesis. Furthermore, as shown in Figure 11, the decoder's standard multichannel input bus can be used to incorporate a pre-recorded multichannel background or ambience signal in order to produce a richer virtual sound scene.

Each audio object can feed one or more of a plurality of reverberators, with an adjustable pre-delay and frequency-

dependent attenuation for each reverberator feed. Each reverberator simulates the reverberation response of a room or acoustic enclosure in the virtual environment. One of these rooms may be set as the 'primary room' where the listener is located, while the others are 'secondary rooms' audible through openings or partitions. The construction of artificial reverberation algorithms suitable for this purpose is discussed e.g. in [28]–[31]. A recursive Feedback Delay Network (FDN) can efficiently model a multichannel exponentially decaying reverberation process with any desired degree of perceptual verisimilitude [31].

The per-source processing module includes a parametric filter on each of the direct and reflected paths. This filter can be efficiently designed to lump the effects of acoustic obstacles or partitions along with attenuations that depend on source distance (including air and wall absorption) and on source directivity and orientation relative to the listener. A similar parametric filter is applied on each reverberator output to account for occlusion through partitions intervening between a secondary room and the listener.

Each reverberator produces a multichannel output signal feeding the 'Standard' multichannel master bus. A 'reverb panning' module is inserted in order to enable the simulation of reverberation from a secondary room heard through a localized opening. The perceived location and size of such an opening can be controlled by the 'divergence panning' algorithm described below in section VI. Acoustical coupling between rooms (such as energy transfer through openings) can be simulated by incorporating a reverberation coupling path from each reverberator output to the inputs of the other reverberators (not shown on Figure 11).

One of the reverberators (labeled 'clustered reflections' in Figure 11) is configured as a parallel bank of early reflection generators each having single-channel input and output. This provides an efficient method for controlling the perceived effect of acoustic reflections from the listener's immediate environment. Rather than synthesizing and controlling a set of individual reflections for each primary virtual source, this method enables the synthesis of a group (cluster) of reflections with per-source control of the pre-delay, intensity, and directional distribution of these reflections, while sharing a common early reflection processing unit among all sources. The directional distribution of the clustered reflections is controlled for each source by use of the 'divergence panning' algorithm described below.

VI. Divergence Panning and Spatial Extension

A particular type of directional panning algorithm is introduced in [27] to control the spatial distribution of reverberation components and clustered reflections. In addition to reproducing a direction, this type of algorithm, referred to herein as 'divergence panning', controls the angular extent of a radiating arc centered on this direction. This is illustrated in Figure 12 for the 2D case.

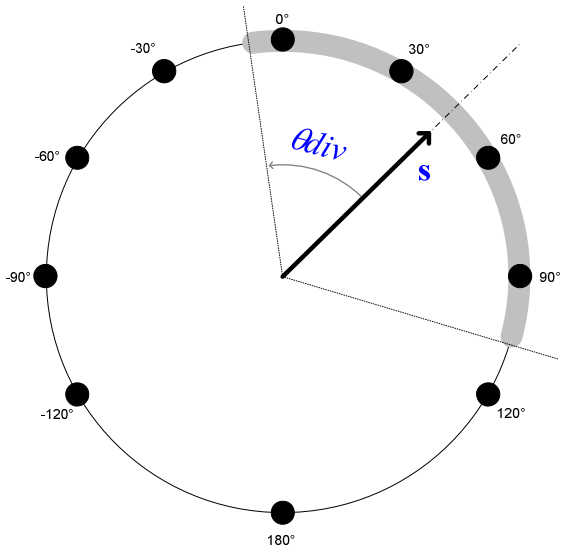

Fig. 12 Definition of divergence angle parameter θdiv and divergence panning vector **s** in the horizontal plane.

The value of the divergence angle θdiv can vary from 0 (pinpoint localization) to π (diffuse localization). A convenient alternative parameterization consists of representing the direction angle and the divergence angle jointly in the form of a panning vector whose magnitude is 1.0 for pinpoint localization and 0.0 for diffuse localization. This property is obtained in [27] by defining the panning vector, here denoted **s**, as the normalized integrated Gerzon energy vector for a continuous distribution of uncorrelated sound sources on the radiating arc shown in Figure 12. This yields a simple relation between the divergence panning vector magnitude and the divergence angle θdiv:

$$\begin{aligned} \|\mathbf{s}\| &= \left[\int_{[-\theta div,\, \theta div]} \cos\theta \, d\theta \right] / \left[\int_{[-\theta div,\, \theta div]} d\theta \right] \\ &= \sin(\theta div) / \theta div. \end{aligned} \qquad (2)$$

The implementation of the divergence panning algorithm requires a method for deriving an energy scaling factor associated to each of the output channels. This can be achieved by modeling the radiating arc as a uniform distribution of notional sources with a total energy of 1.0, assigning discrete energy panning weights to each of these notional sources and summing for each output channel the panning weight contributions of all these sources to derive the desired energy scaling factor for this channel. This method can be readily extended to three dimensions (e.g. by considering an axis-symmetric distribution of sources around the point located at direction (θ, φ) on the 3D sphere).

The notion of spatial extent of a multichannel reverberation signal or cluster of reflections, specified through the divergence panning representation proposed above, is also relevant to the rendering of spatially extended audio objects. This problem was studied extensively by Potard, who proposed a perceptually-based rendering method consisting of

generating a collection of closely positioned notional point-sources emitting mutually uncorrelated replicas of the original object waveform signal [10] [44] [45]. This approach incurs substantial per-source computation costs, since multiple decorrelated signals must be generated and each must then be spatialized individually.

A new computationally efficient method for simulating spatially extended sound sources was proposed in [27], whereby divergence panning weights are derived for each spatially extended audio object, and rendering an audio object having the specified spatial properties (direction and divergence angle) is achieved by applying these weights to a multichannel signal whose channels are mutually uncorrelated. For this purpose, the renderer architecture of Figure 11 includes the multichannel *Diffuse* master bus, which feeds a multichannel decorrelation filter bank (each channel of the bus feeds a different filter from the bank). This technique offers several advantages over the multiple emitter approach described previously:

- The per-source processing cost for a spatially extended source is significantly reduced, and comparable to that of a point source spatialized in standard multichannel mode.
- Since the decorrelation filter bank is common to all sources, its complexity is not critical and it can be designed without compromises. Ideally, it consists of mutually orthogonal all-pass filters. A review of decorrelation filter design methods for this purpose is given in [10] [45].

VII. Conclusion

The object-based rendering engine architecture described in this paper is applicable to binaural reproduction over headphones and to loudspeaker-based spatial audio reproduction techniques, including those reviewed in section II of this paper. For instance, transaural rendering using one or several pairs of loudspeakers is realized simply by replacing the output decoder module in the renderer architecture of Figure 11. For multichannel discrete amplitude panning, this spatial decoder and the *Left* and *Right* binaural multichannel buses are omitted, since the *Standard* multichannel bus can carry all point-source direct path contributions.

In the case of binaural or transaural reproduction, the methods described in this paper enable a significant reduction of the computational complexity overhead commonly associated to HRTF-based spatialization technology, without compromising its benefits in terms of positional audio rendering fidelity. The proposed multi-channel binaural encoding solution can be viewed as a hybrid processing architecture combining a discrete multichannel panning technique (VBAP) and a HRTF-based virtualization back-end (spatial decoder), providing the ability to separately optimize the reproduction of inter-aural time difference (ITD) cues and the reconstruction of HRTF spectral cues for all audio object positions around the listener.

The resulting object-based 3D audio and multi-reverberation rendering engine is suitable for the interactive audio reproduction of complex virtual scenes, driven via a low-level generic audio scene representation compatible with the Open AL and OpenSL ES APIs, for instance. This, along with its computational efficiency, makes this engine applicable to the deployment of immersive interactive 3D audio rendering systems in a wide range of consumer appliances ranging from personal computers to home theater and mobile entertainment or communication devices.

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. Potard, 3D-audio object oriented coding, Ph.D. thesis, Univ. of Wollonlong (2006).

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

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

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

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

[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] 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).

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

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

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

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

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

[22] 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).
[23] W. G. Gardner, *3-D Audio Using Loudspeakers*, Ph.D. Thesis, Massachussets Institute of Technology (1997),
[24] A. Jost and J.-M. Jot "Transaural 3-D Audio with User-controlled Calibration," *Proc. Int. Conf on Digital Audio Effects* (DAFX 2000).
[25] 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).
[26] 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.).
[27] 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.).
[28] J.-M. Jot, "Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality", *Proc. International Computer Music Conference* (1997).
[29] W. G. Gardner, "Reverberation algorithms," *Applications of Signal Processing to Audio and Acoustics* (ed. M. Kahrs, K. Brandenburg), Kluwer Academic (1998).
[30] L. Dahl and J.-M. Jot, "A Reverberator Based on Absorbent All-pass Filters", *Proc. Int. Conf on Digital Audio Effects* (DAFX 2000).
[31] 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.).
[32] F. Rumsey, *Spatial Audio* (Focal Press, 2001).
[33] J. Merimaa, "Energetic Sound Field Analysis of Stereo and Multichannel Loudspeaker Reproduction," Proc. 123rd Conv. Audio Eng. Soc., preprint 7257 (2007 Oct.).
[34] A.J. Berkhout, "A holographic approach to acoustic control," *J. Audio Eng. Soc.* 36:977–995 (1988 Dec.).
[35] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," Proc. 124th Conv. Audio Eng. Soc. (2008 May).
[36] J. Daniel, S. Moreau and R. Nicol "Further Investigations of High-Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging," Proc. 114th Conv. Audio Eng. Soc., preprint 5788 (2003 Feb.).
[37] S. Spors and J. Ahrens, "A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling," Proc. 125th Conv. Audio Eng. Soc. (2008 Oct).
[38] J. Daniel and S. Moreau, "Further Study of Sound Field Coding with Higher Order Ambisonics," Proc. 116th Conv. Audio Eng. Soc., preprint 6017 (2004 May).
[39] S. Spors, H. Wierstorf, M. Geier, and J. Ahrens, "Physical and perceptual properties of focused sources in wave field synthesis," Proc. 127th Conv. Audio Eng. Soc. (2009 Oct.).
[40] D. G. Malham, "3-D Sound for Virtual Reality Using Ambisonic Techniques," *3rd Annual Conf. on Virtual Reality* (1993) [www.york.ac.uk/inst/mustech/3d_audio/vr93papr.htm].
[41] C. Travis, "Virtual Reality Perspective on Headphone Audio," Proc. 101st Conv. Audio Eng. Soc., preprint 4354 (1996).
[42] J. Chen, B. D. Van Veen and K. E. Hecox, "A Spatial Feature Extraction and Regularization Model for the Head-Related Transfer Function," *J. Acoust. Soc. Am.* 97(1):439-52 (1995 Jan.).
[43] J.-M. Jot, S. Wardle, and V. Larcher, "Approaches to Binaural Synthesis," Proc. 105th Conv. Audio Eng. Soc. (1998 Aug.).
[44] G. Potard, "Study of Sound Source Shape and Wideness in Virtual and Real Auditory Displays," Proc. 114th Conv. Audio Eng. Soc. (2003 March).
[45] G. Potard, "Decorrelation techniques for the rendering of apparrent sound source width in 3D audio displays," *Proc. Int. Conf on Digital Audio Effects* (DAFX 2004).