Comparison of Spatial Audio Simulation Systems

Vladimír Arnošt*
arnost@fit.vutbr.cz

Filip Orság*
orsag@fit.vutbr.cz

Abstract: Modelling and simulation of spatial (3D) sound propagation in real-time applications is a challenging task. Its complexity, both algorithmic and computational, is a limiting factor in most current auralization systems. Each system is designed with different goals and its designers had to make different trade-offs to be able to implement it using current software and hardware technologies. This paper tries to analyse some of these systems, describe their operation, features and limitations in order to evaluate author's own proposed spatial audio simulation system. Attention is focused mainly to the employed algorithms, their computational complexity versus their subjective quality. Both commercial and academic auralization systems are considered, although it is not possible to consider all existing systems in the limited scope of this paper.

Keywords: sound propagation, simulation, auralization, comparison

1 Introduction

The problem of spatial sound propagation simulation has been tackled by many individuals and companies. Their solutions vary greatly in the used methods, simplifying assumptions, required computational power and subjective quality. This paper is trying to present some of these systems and compare their most important characteristics: complexity and quality. No attempt is made to create an absolute scale from the 'worst' to the 'best' though, because each system has its particular positive and negative aspects and the systems were not designed with the same purpose in mind.

2 Overview of Spatial Audio Systems

This section is not trying to provide an exhaustive analysis of *all* existing auralization[1] systems. Instead, the most common commercially available systems are compared to selected non-commercial systems, including the one developed by the author.

2.1 Microsoft DirectSound 3D

DirectSound 3D [5] is a component of Microsoft DirectX API. It allows the developer to define the position, orientation and speed of several sound sources and listeners. The system calculates how the listener would hear the sources in real time, including the Doppler

* Department of Computer Science and Engineering, Faculty of Information Technology, Brno University of Technology, Božetěchova 2, CZ-612 66 Brno, Czech Republic

[1] *auralization:* the process of rendering audible, by physical or mathematical modelling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modelled space. [Kleiner et al, 1993].

shift caused by mutual movement. On the other hand, there are no means for taking room geometry, obstruction, occlusion and sound reflections into account. The simulation environment is equivalent to infinite open space filled with air (or approximately to an anechoic room). The resulting sound is very unnatural.

DirectSound 3D is a basic building block for many other auralization systems and its hardware-independent architecture greatly influences these systems. Thanks to its design, it is possible to accelerate some calculations in hardware if available.

2.2 Creative EAX

Environmental Audio Extensions [6] were created by Creative Labs. to add some of the missing features of DirectSound 3D: better reverberation, reflections, obstruction[2] and occlusion[3]. These features are not calculated from the room geometry, but are approximated by a set of predefined reverberation algorithms. The software developer can manually modify some parameters of these algorithms, like the apparent room size (timing of early reflections) and the speed of reverberation decay (attenuated room versus a large hall). There is however no guarantee that the *pre-set* reverb type really matches the room geometry the listener is in. Furthermore, the reverb is the same in the whole room and reflections from nearby walls are not taken into account. A sound source located in the corner of the room will therefore sound the same as if it were in the centre. The same applies to listener location.

Note: Creative Labs. have released EAX version 4.0 in 2004, but starting with EAX 3.0, the specifications are available only to developers who sign a five-year NDA[4] and are restricted from publishing any information about EAX. So even if I signed the NDA, I would not be legally permitted to publish my findings here. Secondly, the sound card I am using for my tests (Sound Blaster Live) does not support these new standards either. This paper will therefore deal only with the public specification of EAX 2.0.

2.3 OpenAL

The Open Audio Library [8] is a public standard defined by Loki Software to provide a uniform, hardware-independent and extensible interface for controlling sound cards. It was originally invented to create an abstract layer for porting games to Linux, but the library is also available for Windows. OpenAL is similar in functionality to DirectSound. It supports both 2D (stereo) and 3D (spatial) sound. It can even co-exist with EAX 2.0–4.0 and use their reverberation algorithms to enhance the spatial sounds.

2.4 Aureal A3D

Aureal A3D [3] tried a different approach to create spatial sound than EAX. Instead of using pre-set reverberation algorithms, the Aureal chip calculated the direct ray path and first-order reflections based on source and listener location and the real room geometry. The perception of the surrounding space was very good. Unfortunately, A3D required a

[2] *obstruction:* when a sound source and the listener are in the same environment but an obstacle blocks the direct sound path from the source to the listener

[3] *occlusion:* when a sound source and the listener are in different environments, e.g. on the opposite sides of a wall

[4] Non-Disclosure Agreement

rather expensive chip, was never implemented in hardware by other companies and finally Aureal Inc. went out of business. A3D 1.0 can be partially emulated in software and is supported by third parties now, but A3D 2.0 is too complex to be efficiently emulated by common desktop processors while leaving enough CPU power to other tasks.

2.5 Non-commercial Systems

There is a large variety of non-commercial auralization systems, but most of them exist only as research prototypes and quite surprisingly their sources are usually not available. Some of them also require custom-built hardware to run, like special DSP boards, etc. Since it is impossible to test these systems personally, they will be included only for general comparison.

2.5.1 Burgess — Real-Time Audio Spatialization

This system [4] uses Head-Related Transfer Functions (HRTF) to model the frequency characteristics of the human ear with FIR filters. The HRTF were obtained by measuring the frequency response of a dummy head at various sound source azimuths and elevations. The HRTF are accompanied by reverberation, several non-linear distortion functions and simple FIR filters (low pass, comb filter) to simulate the room response. All calculations are performed twice, once for the left ear and once for the right ear.

The audio spatialization system runs on a Motorola DSP56001 processor at a sampling rate of 32kHz. The HRTF functions are modeled by 55-point FIR filters (in each channel).

2.5.2 Funkhouser — Real-Time Acoustic Modeling

The authors have developed a powerful system [7] for distributed calculation of sound reflections in large virtual-reality worlds. Multiple users can enter the system and communicate with each other and hear realistic spatialized sound. The users can move and the system calculates the sound reverberation path in real time.

The system uses *beam tracing* algorithm to find all reverberation paths between the sound source and the listener. The algorithm complexity grows exponentially with the number of calculated reflections and can be too slow in complex scenes. Several optimization methods are involved in the system to achieve considerable speed-ups by calculating only the most significant sound beams.

3 Proposed System

The proposed spatial audio simulation system [1, 2] will calculate the impulse response corresponding to the reverberation path using *cone tracing* method. Using wide cones instead of thin rays should accelerate the algorithm significantly, because less cones should be necessary to cover all sound source directions. Thousands or more rays are typically required to get the required space coverage, but mere tens to hundreds of cones might do the same job. This is because the majority of rays miss the listener and do not contribute to the resulting impulse response. Cones occupy varying sections of space, can reflect from walls, break into smaller cones on edges and even “bend” around corners like real sound waves.

When a sound cone passes over a listener, exact propagation delay and attenuation level are calculated from the travelled distance and the number of reflections. These two values can be converted into a point in the impulse response. The impulse responses obtained by simulating all the sound sources can be simply added together to get an overall impulse response (assuming that all sources emit the same signal). Every listener will have its own distinct impulse response. Typically, a human head would consist of two listeners (ears) with opposite orientation and about 20cm apart.

The calculated impulse responses may be split into two principal parts:

- **early reflections**, which characterize the room shape
- **reverberation tail**, which produces the impression of space

The early reflections can be modelled by a FIR filter, because there is usually only a limited number of discrete impulses at this stage. The reverberation tail usually consists of an exponentially decaying dense series of reflections, which are best modelled by feedback IIR filter(s). There is no exactly defined boundary dividing the early reflections from the reverberation tail. Article [7] states that the typical early reflections last between 20ms and 80ms.

4 Comparison of Spatial Audio Systems

Comparing such a diverse set of auralization systems is very difficult. It is close to impossible to find a common set of objectively measurable parameters which could be used to assess *all* the systems. To make it more difficult, most of the analysed systems are available only on paper and can not be tested directly. Their performance must be estimated from the available data.

4.1 Auralization System Features

Table 1 summarizes the most important features supported by the compared auralization systems. It is important to realize that even if two systems support the same feature, the actual implementation may vary and the results might be very different.

System	**Room geometry**	**HRTF**	**Occlusion and obstruction**	**Reverbe-ration**	**Doppler effects**
DirectSound 3D	no	yes	no	yes	yes
EAX	no	no	yes	yes	yes
OpenAL	no	no	no	no	yes
A3D	yes	yes	yes	yes	yes
Burgess	no	yes	no	yes	no
Funkhouser	yes	no	yes	no	no
Cone Tracing	yes	no	yes	yes	no

Tab. 1: Supported features of auralization systems

System	Type	CPU Usage
DirectSound 3D	SW with HW acceleration	low
EAX	SW + special HW	low
OpenAL	SW with HW acceleration	low
A3D	SW + special HW	medium
Burgess	SW + special HW	low
Funkhouser	SW	high
Cone Tracing	SW	high

Tab. 2: Computational complexity of auralization systems

System	Source localization	Sensation of space
DirectSound 3D	good	fair
EAX	good	good
OpenAL	fair	fair
A3D	very good	good
Burgess	very good	fair
Funkhouser	good	very good
Cone Tracing	good	very good

Tab. 3: Subjective quality of auralization systems

4.2 Computational Complexity

Auralization systems (table 2) can be divided into several groups by their use of dedicated hardware to accelerate the calculations:

– software-only systems
– software systems, which can optionally use hardware acceleration
– software systems, which require special hardware to run

It is clear that hardware-accelerated systems will provide better performance than software-only ones. Their disadvantage is the need for a special device to be present in the computer system, which narrows the range of potential users of the system.

The proposed cone tracing system has similar computational complexity as the existing Funkhouser's system [7]. Both use a similar algorithm and perform both the reverberation path calculations and the convolutions of the input signals with the calculated impulse responses in software.

4.3 Subjective Quality

Testing the subjective quality of the auralization systems is hard, because the very nature of the systems prohibits using the same model of the scene and the same sound sources to maintain identical testing conditions. Worse, some hardware-accelerated systems are not physically available and thus no first-hand experience can be considered in the comparison and only the authors' published evaluations can be used. Table 3 hence contains only *estimated* quality indicators, which indicate only the relative quality differences between various systems.

4.3.1 Source Localization

The ability to locate the relative position of the sound source to the listener's head is crucial for the orientation in the virtual world. Best localization results are achieved by HRTF functions, combined with filters to provide realistic distance cues.

4.3.2 Sensation of Space

To make the listener "sense" the surrounding space, it is necessary to model the early reflections from the nearby surfaces as well as the reverberation tail to underline the impression of huge closed spaces. Most systems do not calculate the early reflections dynamically depending on the listener position and thus the person cannot detect an obstacle by walking around it and e.g. hearing the echoes of his own steps. Reverberation tail is easier to implement because it is already an approximation of a large number of periodically bouncing sound waves with exponentially decreasing energy.

5 Conclusions

This paper presented a selection of auralization systems, described their basic features and tried to compare them to each other by their subjective quality.

The systems using HRTF functions usually offer the best sound localization and reverberation path tracing systems excel at producing the correct impression of the surrounding space. In the ideal case, these systems should be combined to get the best results.

The computing power requirements are roughly proportional to the quality of spatial sound processing. The better and more detailed simulation, the higher the cost.

Bibliography

1. Arnošt, V.: *Discrete Simulation of Sound Wave Propagation.* MOSIS 2000 Proceedings, roč. I. (2000), MARQ, Rožnov pod Radhoštěm, Czech Republic, 2000, pp. 241–246, ISBN 80-85988-44-5.
2. Arnošt, V.: *Počítačová simulace šíření zvuku.* Proceedings of the 1st Conference of Czech Student AES Section on Audio Technologies and Processing, roč. 1. (2000), FEI VUT Brno, Brno, Czech Republic, 2000, pp. 76–79, ISBN 80-214-1602-5.
3. *A3D 3.0 Software Developer Kit Reference Guide.* Document Number DO3010–030100, Aureal Inc., U.S.A., 2000, pp. 19–20.
4. Burgess, D. A.: *Real-Time Audio Spatialization with Inexpensive Hardware.* Georgia Institute of Technology, Atlanta, U.S.A., 1992.
   `http://www.cc.gatech.edu/gvu/multimedia/spatsound/spatsound.html`
5. *Direct Sound 3D Software Developer Kit.* Microsoft Inc., U.S.A., 2000.
   `http://msdn.microsoft.com/directx/`
6. *EAX 2.0 Software Developer Kit.* Creative Labs Inc., U.S.A., 2001.
   `http://developer.creative.com`
7. Funkhouser, T.—Min, P.—Carlbom, I.: *Real-Time Acoustic Modeling for Distributed Virtual Environments.* Computer Graphics (SIGGRAPH 99 Proceedings), Addison-Wesley, 1999, pp. 365–374, ISBN 0-201-48560-5.
8. *OpenAL 1.1 Specification and Reference.* Loki Software et al, U.S.A., 2005.
   `http://www.openal.org`