SIGGRAPH 2002 Course Notes

"Sounds Good to Me!"
Computational Sound for Graphics, Virtual Reality, and Interactive Systems

Thomas Funkhouser
Princeton University

Jean-Marc Jot
Creative Technology, Ltd.

Nicolas Tsingos
REVES-INRIA

Abstract

This course covers concepts, models, techniques, and systems for simulation and rendering of sound in virtual environments. It will focus on real-time methods for spatializing sounds in interactive systems. Discussion will include both technical aspects of algorithms and practical aspects of applications. So, the course is appropriate both for researchers interested in learning about sound simulation and developers interested in including spatialized sounds in their virtual environments.

Prerequisites: some knowledge of signal processing and geometric computation.

Course Schedule

Module 1 - Geometric models

1:30 Introduction - Funkhouser

1:45 Geometric acoustic modeling methods - Funkhouser

2:10 Recent work in geometric acoustic modeling - Funkhouser

2:35 From geometric models to spatialized sound - Tsingos

2:55 3D auditory displays - Tsingos

3:15 Break

Module 2 - Perceptual models

3:30 Overview of perceptual models - Jot

3:55 Artificial reverberation algorithms - Jot

4:20 Standards, tools, & demonstrations - Jot

4:50 Conclusions and future directions - Tsingos

5:00 Questions - Funkhouser, Jot, & Tsingos

1 Introduction

The motivating problem addressed by this course is to model the propagation of sound waves traveling through a 3D environment. This research problem is important for applications ranging from concert hall design to immersive video games. Although acoustic modeling of architectural spaces have been studied for decades [74, 92], this course focuses on new applications of spatialized sound in graphics, virtual reality, and interactive systems.

In the computer graphics community, spatialized sound is most common in immersive virtual environment applications. Typically, such a system allows a user to “explore” a virtual world and/or “interact” with other users by rendering images and sounds of the environment in real-time while the user “moves” interactively. If the system renders images and sounds of the 3D environment quickly and realistically enough, the illusion of immersion in the environment can be achieved, providing the user with a sense of presence.

Example virtual environment applications include distributed training, computer-aided design, telepresence, electronic commerce, education, and entertainment. For example, simulated environments are used to train soldiers, fire fighters, and other people whose missions are too dangerous or too expensive to re-create in the real-world [15, 80]. Architects, mechanical engineers, and urban planners use modeling of proposed designs to avoid costly re-design costs (e.g., for concert halls [16, 88]). Video games provide a user the experience of inhabiting and interacting with other people in a 3D virtual world.

Spatialized sound effects are important in these applications because they combine with visual cues to aid localization of objects, separation of simultaneous sound signals, and formation of spatial impressions of an environment [12]. For instance, binaural auditory cues are essential in localizing objects outside a user’s field of view, such as when a car comes around a blind corner in a driving simulation, or when a soldier must find a sharpshooter in a military training application. They also help users separate simultaneous sounds, such as when we listen to one of many speakers at a cocktail party. Finally, qualitative changes in sound reverberation can enhance and reinforce visual comprehension of the environment, such as when a user of a video game moves from a large stone cave into a small wooden house. Experiments have shown that more accurate acoustic modeling provides a user with a stronger sense of presence in virtual environments [31].

A difficult challenge for virtual environment systems is to provide accurate (or at least plausible) spatialization of sound in non-trivial 3D environments. Sound waves traveling from a *source* (e.g., a speaker) and arriving at a *receiver* (e.g., a microphone) travel along a multitude of *propagation paths* representing different sequences of reflections, transmissions, and diffractions at surfaces of the environment (Figure 1). The effect of these propagation paths is to add *reverberation* (e.g., echoes) to the original source signal as it reaches the receiver. Auralizing a sound for a particular source, receiver, and environment can be achieved by applying filter(s) to the source signal that model the acoustical effects of sound propagation through the environment.

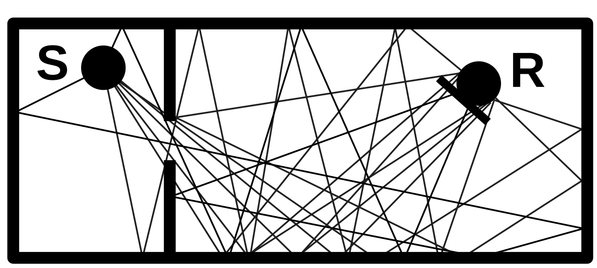


Figure 1: Sound propagation paths from a source (S) to a receiver (R).

Figure 2 shows a basic processing pipeline for auralization (rendering spatialized sound). The input to the system is a description of a virtual environment (e.g., a set of polygons), an audio source location, an audio receiver location, and an input audio signal. The auralization system computes a model for the propagation of sound waves through the environment and constructs digital filter(s) (e.g., *impulse response(s)*) that encode the delays and attenuations of sound traveling along different propagation paths. Convolution of the input audio signal with the filter(s) yields a spatialized sound signal for output with an auditory display device.

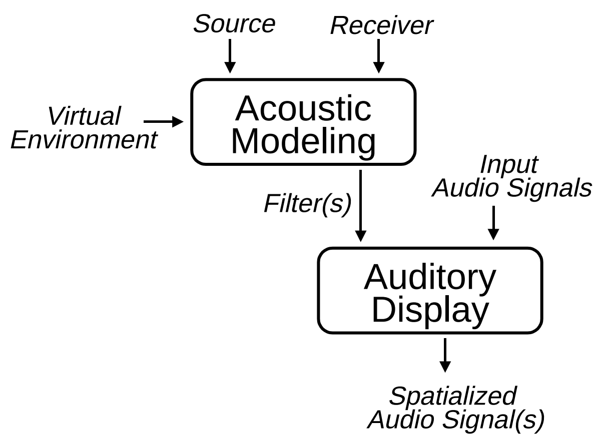


Figure 2: Basic auralization pipeline.

Impulse responses representing acoustic environments are usually considered in three parts: (1) direct sound, (2) early reflections, and (3) late reverberation (Figure 3). Direct sound represents the earliest arriving (and usually strongest) sound wave. Early reflections describe the sound waves that arrive within the first t_e milliseconds of the impulse response (e.g., $20ms \leq T_e \leq 80ms$ [9, 50]), when the density of reverberations is low enough that the human ear is able to distinguish individual paths (e.g., less than 2,000 reflections per second) [24]. These early reflections (and possibly diffractions) provide a human listener with most of the spatial information about an environment, because of their relatively high strengths, recognizable directionalities, and distinct arrival times [9, 51, 89, 128]. Consequently, they are modeled as accurately as possible with geometric techniques [74]. In the late reverberation phase, when the sound has reflected off many surfaces in the environment, the impulse response resembles an exponentially decaying noise function with overall low power [9] and with such a high density that the ear is no longer able to distinguish them independently [24]. Consequently, late reverberations are usually modeled with statistical, perceptual approximations (e.g. [1, 101]).

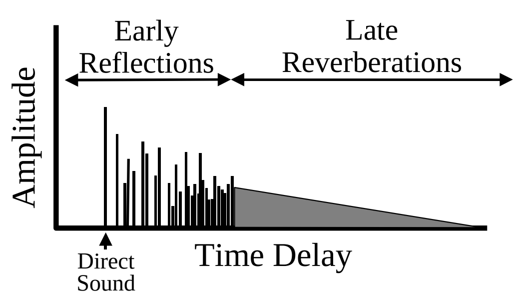


Figure 3: Direct, early, and late parts of an impulse response.

In this course, we review basic auralization methods for 3D virtual environment applications. The following section introduces the problem of modeling sound propagation more formally and discusses its similarities and differences with respect to global illumination. The next three sections describe different computational modeling approaches: finite element methods, geometrical simulations, and artificial reverberation models. They are followed by a discussion of how spatialized sound can be output with 3D auditory displays. Finally, the last section contains a brief summary and a discussion of topics for future work.

2 Overview

At a fundamental level, the problem of modeling sound propagation is to find a solution to an integral equation expressing the wave-field at some point in space in terms of the wave-field at other points (or equivalently on surrounding surfaces). Of course, this transport problem is similar to global illumination, which is described by Kajiya's rendering equation [63]. For sound simulations, the wave equation is described by the Helmoltz-Kirchoff integral theorem [14], which incorporates time and phase dependencies.

Although acoustics and graphics rendering both simulate wave propagation, sound has characteristics different from light which introduce several new and interesting challenges:

- **Wavelength:** the wavelengths of audible sound range between 0.02 and 17 meters (for 20KHz and 20Hz, respectively), which are five to seven orders of magnitude longer than visible light. Therefore, as shown in Figure 4, reflections are primarily specular for large, flat surfaces (such as walls) and diffraction of sound occurs around obstacles of the same size as the wavelength (such as tables), while small objects (like coffee mugs) have little effect on the sound field (for all but the highest wavelengths). As a result, when compared to computer graphics, acoustics simulations tend to use 3D models with far less geometric detail. But, they must find propagation paths with diffractions *and* specular reflections efficiently, and they must consider the effects for different obstacles at a range of wavelengths.

- **Speed:** at 343 meters per second, the speed of sound in air is six orders of magnitude less than light, and sound propagation delays are perceptible to humans. Thus, acoustic models must compute the exact time/frequency distribution of the propagation paths, and sound must be auralized by convolution with the corresponding *impulse response* that represents the delay and amplitude of sounds arriving along different propagation paths. In contrast, the propagation delay of light can be ignored and only the energy steady-state response must be computed.

- **Coherence:** sound is a coherent wave phenomenon, and interference between out-of-phase waves can be significant. Accordingly, acoustical simulations must consider phase when summing the cumulative contribution of many propagation paths to a receiver. More specifically, since the phase of the wave traveling along each propagation path is determined by the path length, acoustical models must compute accurate path lengths (up to a small percentage of the wavelength). In contrast, most light sources (except lasers) emit largely incoherent waves, and thus lighting simulations simply sum the power of different propagation paths.

- **Dynamic range:** the human ear is sensitive to five orders of magnitude difference in sound amplitude [12], and arrival time differences allow some high-order reflections to be audible. Therefore, as compared to computer graphics, acoustical simulations usually aim to compute several times more reflections, and the statistical time/frequency effects of late sound reverberation are much more significant than for global illumination.

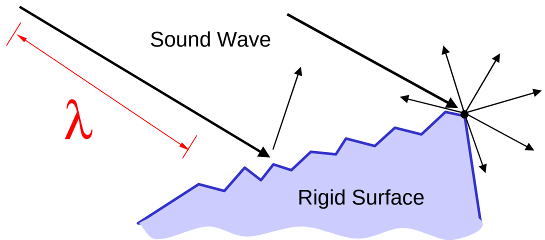


Figure 4: Sound waves impingent upon a surface usually reflect specularly and/or diffract at edges.

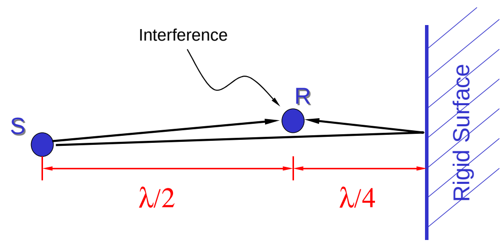


Figure 5: Interference can occur when two sound waves meet.

Despite these differences, many of the same techniques are used in acoustic modeling as are used for global illumination. In both cases, a major difficulty arises from the wave-field discontinuities caused by occlusions and specular highlights, resulting in large variations over small portions of the integration domain (i.e. surfaces and/or directions). Due to these discontinuities, no general-purpose, analytic formula can compute the wave-field at a given point, and solutions must rely upon sampling or subdivision of the integration domain into components that can be solved efficiently and accurately.

Prior computational methods for simulating the propagation of sound through an environment can be classified into three major approaches: 1) numerical solutions to the wave equation (e.g., finite and boundary element methods), 2) high frequency approximations based on geometric propagation paths (e.g., image source methods, ray tracing, and beam tracing), and 3) perceptually-based statistical models (e.g., feedback delay networks). The following three sections review these approaches. They are followed by a discussion of signal processing and auditory displays for auralization.

3 Finite and Boundary Element Methods

Finite and boundary element methods solve the wave equation (and associated boundary conditions), subdividing space (and possibly time) into *elements* [19, 32, 68, 69] (Figure 6). The wave equation is then expressed as a discrete set of linear equations for these elements. The boundary integral form of the wave equation (i.e., Green's or Helmoltz-Kirchoff's equation) can be solved by subdividing only the boundaries of the environment and assuming the pressure (or particle velocity) is a linear combination of a finite number of basis functions on the elements. One can either impose that the wave equation is satisfied at a set of discrete points (collocation method) or ensure a global convergence criteria (Galerkin method). In the limit, finite element techniques provide an accurate solution to the wave equation. However, they are mainly used at low frequencies and for simple environments since the compute time and storage space increase dramatically with frequency.

Finite element techniques have also been used to model acoustic *energy* transfer between surfaces [86, 124] (as in radiosity [47]). While they can be used to compute energy decay characteristics in a given environment, they do allow direct reconstruction of an impulse response for auralization. Instead, they require the use of an underlying statistical model and a random phase assumption [75]. Moreover, most surfaces act primarily as specular or glossy reflectors for sound. Although extensions to non-diffuse environments have been proposed in computer graphics [106, 21], they are often time and memory consuming. Accordingly, finite and boundary element methods are not generally used for interactive virtual environment applications.

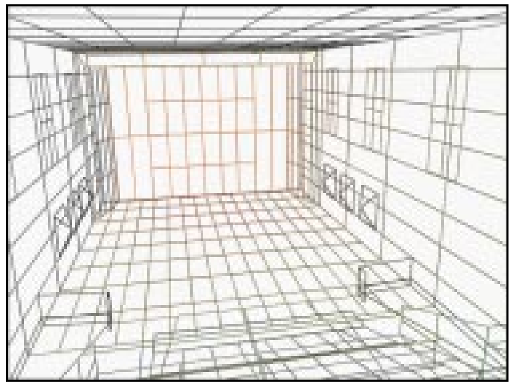
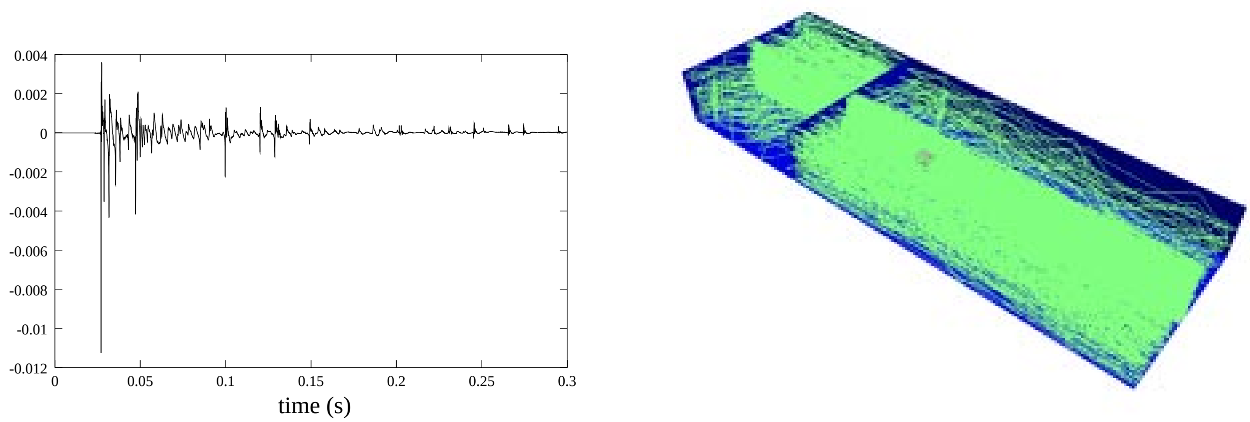


Figure 6: Boundary element mesh.

4 Geometric Methods

Geometrical acoustic simulations model the acoustical effects of an environment with computations based on ray theory. They make the assumption that sound wavelengths are significantly smaller than the size of obstacles, and thus they are valid only for high-frequency sounds.

The general approach is similar to methods used in computer graphics. A geometric algorithm is used to find significant ray paths along which sound can travel from a source to a receiver (Figure 7), and mathematical models are used to approximate the filters corresponding to source emission patterns, atmospheric scattering, surface reflectance, edge diffraction, and receiver sensitivity for sound waves traveling along each path. Finally, an impulse response is constructed by combining the filter(s) for each propagation path.



(a) Impulse response (b) Propagation paths

Figure 7: Impulse response (left) representing 353 propagation paths (right) for up to ten orders of specular reflections between a point source and a point receiver (omnidirectional) in a coupled-rooms environment (two rooms connected by an open door).

Geometric algorithms currently provide the most practical and accurate method for modeling the early part of an impulse response for high-frequency sounds. The delays and attenuations of the direct sound and early reflections/diffractions are computed explicitly, and thus simulated impulse responses contain the main perceptually significant peaks used for localization. Also, correct phase and directivity of sound waves can be obtained from the lengths and vectors of computed paths. However, geometric methods are generally practical and accurate only for the early part of the response, as the errors in geometric approximations and the computational complexity of geometric algorithms increase with larger numbers of reflections and diffractions. As a result, common practice is to use geometric methods to find early reflections and to fill in the late reverberations with statistical methods (discussed in the next section).

4.1 Enumerating Propagation Paths

The first challenges of geometric acoustic modeling is to enumerate the significant propagation paths along which sound waves travel from a source to a receiver. Since rays follow the shortest path when the propagation medium is homogeneous, the problem for sound traveling through air reduces to finding piecewise-linear paths from source to receiver with vertices on edges/surfaces of obstacles. Three approaches are most commonly used to address this problem: image sources, ray tracing, and beam tracing.

4.1.1 Image Sources

Image source methods [2, 13] compute specular reflection paths by considering *virtual sources* generated by mirroring the location of the audio source, S, over each polygonal surface of the environment (see Figure 8). For each virtual source, S_i, a specular reflection path can be constructed by iterative intersection of a line segment from the source position to the receiver position, R, with the reflecting surface planes (such a path is shown for virtual source S_c in Figure 8). Specular reflection paths are computed up to any order by recursive generation of virtual sources.

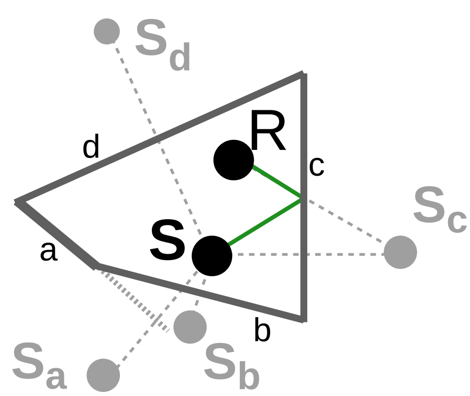


Figure 8: Image source method.

The primary advantage of image source methods is their robustness. They guarantee that all specular paths up to a given order or reverberation time are found. However, image source methods model only specular reflection, and their expected computational complexity grows exponentially. In general, $O(n^r)$ virtual sources must be generated for r reflections in environments with n surface planes. Moreover, in all but the simplest environments (e.g., a box), complex validity/visibility checks must be performed for each of the $O(n^r)$ virtual sources since not all of the virtual sources represent physically realizable specular reflection paths [13]. For instance, a virtual source generated by reflection over the non-reflective side of a surface is "invalid" [13]. Likewise, a virtual source whose reflection is blocked by another surface in the environment or intersects a point on a surface's plane which is outside the surface's boundary (e.g., S_a in Figure 8) is "invisible" [13]. During recursive generation of virtual sources, descendents of invalid virtual sources can be ignored. However, descendents of invisible virtual sources must still be considered, as higher-order reflections may generate visible virtual sources (consider mirroring S_a over surface d). Due to the computational demands of $O(n^r)$ visibility checks, image source methods are practical for modeling only a few specular reflections in simple environments [72].

In the special case of box-shaped environment, image source methods are very efficient. Due to the rectilinear symmetries of a box, image sources representing different permutations of specularly reflecting surfaces all fall on the same location. They tile space in a rectilinear grid pattern as shown in Figure 9, which makes construction of virtual sources efficient and simple to code. More importantly, the set of virtual sources lying at the same location partition potential receiver points inside of the box according to visibility. That is, for any set of specular reflections, every potential receiver point is visible for one and only one permutation, which eliminates the need for expensive visibility tests. For these reasons, some 3D audio systems approximate complex environments as a box and only modeled early specular reflections physically (e.g., [35]).

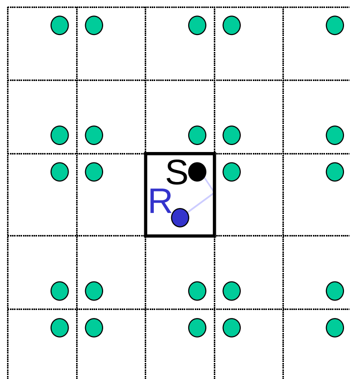


Figure 9: Construction of image sources for a 2D box-shaped environment. The audio source is labeled 'S.' Virtual sources appear as unlabeled dots. The "walls" of the box-shaped room are shown as wide lines near the middle. The thinner lines forming a rectilinear tiling pattern are included only for visualization purposes.

4.1.2 Ray Tracing

Ray tracing methods [73, 133] find reverberation paths between a source and receiver by generating rays emanating from the source position and following them through the environment until an appropriate set of rays has been found that reach a representation of the receiver position (see Figure 10).

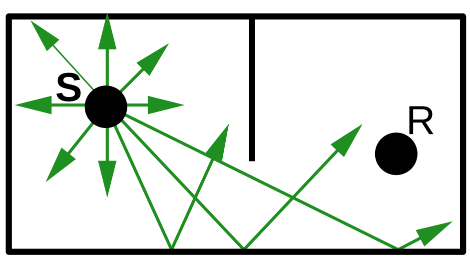


Figure 10: Ray tracing method.

Monte Carlo path tracing methods consider randomly generated paths from the source to the receiver [63]. For instance, the Metropolis Light Transport algorithm [125] generates a sequence of light transport paths by randomly mutating a single current path by adding, deleting, or replacing vertices. Mutated paths are accepted according to probabilities based on the estimated contribution they make to the solution. As contributing paths are found, they are logged and then mutated further to generate new paths in a Markov chain. Mutation strategies and acceptance probabilities are chosen to insure that the method is unbiased, stratified, and ergodic.

A primary advantage of these methods is their simplicity. They depend only on ray-surface intersection calculations, which are relatively easy to implement and have computational complexity that grows sublinearly with the number of surfaces in the model. Another advantage is generality. As each ray-surface intersection is found, paths of specular reflection, diffuse reflection, diffraction, and refraction can be sampled [22, 63], thereby modeling arbitrary types of indirect reverberation, even for models with curved surfaces.

The primary disadvantages of path tracing methods stem from the fact that the continuous 5D space of rays is sampled by a discrete set of paths, leading to aliasing and errors in predicted room responses [76]. For instance, in ray tracing, the receiver position and diffracting edges are often approximated by volumes of space (in order to admit intersections with infinitely thin rays), which can lead to false hits and paths counted multiple times [76]. Moreover, important reverberation paths may be missed by all samples. In order to minimize the likelihood of large errors, path tracing systems often generate a large number of samples, which requires a large amount of computation. Another disadvantage of path tracing is that the results are dependent on a particular receiver position, and thus these methods are not directly applicable in virtual environment applications where either the source or receiver is moving continuously.

4.1.3 Beam Tracing

Beam tracing methods [25, 52] classify propagation paths from a source by recursively tracing pyramidal beams (i.e., sets of rays) through the environment (see Figure 11). Briefly, for each beam, polygons in the environment are considered for intersection with the beam in front-to-back visibility order (i.e., such that no polygon is considered until all others that at least partially occlude it have already been considered). As intersecting polygons are detected, the original beam is clipped to remove the shadow region, a transmission beam is constructed matching the shadow region, a reflection beam is constructed by mirroring the transmission beam over the polygon's plane, and possibly other beams are formed to model other types of scattering. This method has been used in a variety of application areas, including acoustic modeling [25, 37, 38, 39, 85, 110, 129], global illumination [18, 36, 46, 48, 52, 130], radio wave propagation [34, 33], and visibility determination [40, 55, 79, 116].

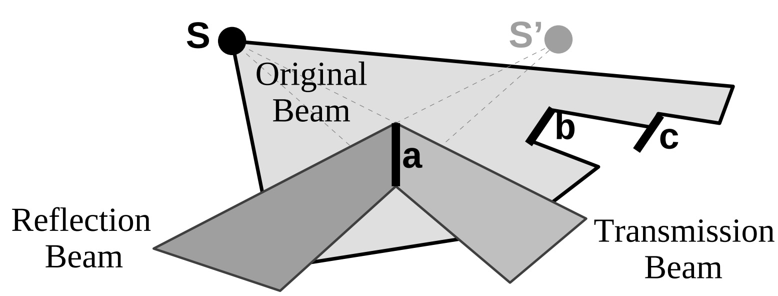


Figure 11: Beam tracing method.

As compared to image source methods, the advantage of beam tracing is that fewer virtual sources must be considered for environments with arbitrary geometry. Since each beam represents the region of space for which a corresponding virtual source (at the apex of the beam) is visible, higher-order virtual sources must be considered only for reflections of polygons intersecting the beam. For instance, in Figure 12, consider the virtual source S_a, which results from reflection of S over polygon a. The corresponding reflection beam,

R_a, contains exactly the set of receiver points for which S_a is valid and visible. Similarly, R_a intersects exactly the set of polygons (c and d) for which second-order reflections are possible after specular reflection off polygon a. Other polygons (b, e, f, and g) need not be considered for second order specular reflections after a. Beam tracing allows the recursion tree of virtual sources to be pruned significantly. On the other hand, the image source method is more efficient for a box-shaped environment for which a regular lattice of virtual sources can be constructed that are guaranteed to be visible for all receiver locations [2].

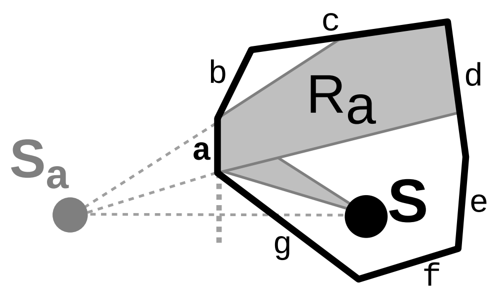


Figure 12: Beam tracing culls invisible virtual sources.

As compared to path tracing methods, the benefit of beam tracing is that it takes advantage of spatial coherence, as each beam-surface intersection represents an infinite number of ray-surface intersections. Polyhedral beam tracing does not suffer from sampling artifacts of ray tracing [76] or the overlap problems of cone tracing [3, 126], since the entire 2D space of directions leaving the source can be covered by beams exactly. As a result, beam tracing can enumerate all potential propagation paths up to some termination criteria without risk of missing any. This feature is particularly important for modeling diffraction [123]. Also, it enables bidirectional methods that find propagation paths more efficiently by combine beaming traced from both sources and receivers [39].

For interactive applications, the main advantage of beam tracing is that beams can be precomputed, during an off-line phase, and stored in a data structure (e.g., a beam tree) for later evaluation of reverberation paths at interactive rates [37]. For instance, beams emanating from a stationary source can be precomputed, enabling fast construction of reverberation paths to an arbitrarily moving receiver (Figure 13) [37]. Alternatively, beams from predicted source regions can be updated asynchronously with time-critical computing strategies to enable interactive generation of reverberation paths between simultaneously moving source and receivers [39].

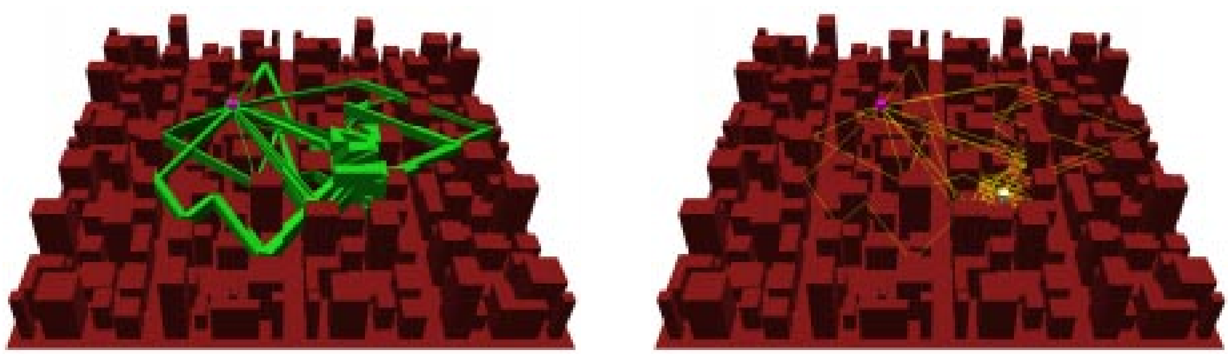


Figure 13: Beams (left) can be precomputed and then queried quickly to update propagation paths (right) at interactive rates.

The primary disadvantage of beam tracing is that the geometric operations required to trace beams through a 3D model (i.e., intersection and clipping) are relatively complex, as each beam may be reflected and/or obstructed by several surfaces. Several methods have been proposed to accelerate these geometric operations, including ones based on BSP-trees [25], cell adjacency graphs [55, 33, 37, 39, 79, 116], layers of 2D triangulations [34], and medial axis approximations [71, 78, 96]. These methods tend to work well only for simple scenes or densely-occluded environments (e.g., cities or building interiors). Beam tracing also is difficult in scenes with curved surfaces and non-linear refracting objects, although conservative beam tracing methods combined with validation of constructed paths is probably suitable for these situations [39].

4.2 Modeling Attenuation, Reflection, and Scattering

Once geometric propagation paths have been computed, they are combined to form filter(s) for spatializing a sound signal. The challenge here is to model the attenuation and scattering of sound as it travels along each path, taking into account source emission patterns, distance attenuation, atmospheric scattering, reflectance functions, diffraction models, and receiver sensitivity. These effects correspond to source models, distance falloff, fog, and bidirectional reflectance distribution functions (BRDFs), and camera response in computer graphics. As in graphics, sound propagation models are approximations, and for each model, there are usually several alternatives which provide trade-offs between computational expense and accuracy.

4.2.1 Distance Attenuation and Atmospheric Scattering

Sound intensity gets attenuated with distance. In virtual acoustics, sound sources are usually modeled as points – i.e. infinitely small points in space radiating a spherical wave-front. In such a case, the free-field *intensity* of the radiation decays with the inverse square of the distance (i.e., in free space, without interfering obstacles). Since we are usually interested in sound pressure rather than intensity, this translates into the well known inverse-distance law:

$$P(R) = P(O)/r,$$

where R is the receiving location, O is the center of radiation and r is the Euclidean distance in 3D space between R and O.

High frequencies also get attenuated with distance due to atmospheric scattering. The expression for a frequency-dependant attenuation coefficient is provided by the ANSI acoustical standard [5] (an ISO equivalent is also available).

4.2.2 Doppler Shifting

When a sound source S and/or a receiver R are moving relative to each other, sound waves undergo a compression or dilatation in the direction of the relative speed of the motion. This compression or dilatation creates a modification of the frequency of the received sound relative to the emitted sound. This effect, which was first discovered by Christian Johann Doppler in 1842, is called *Doppler shifting*.

The Doppler shift between the frequency of the emitted signal and the received signal can be expressed as (see Figure 14):

$$\Delta_{Doppler} = \frac{f_R}{f_S} = \frac{1 - \frac{\mathbf{n}\cdot\mathbf{v}_R}{c}}{1 - \frac{\mathbf{n}\cdot\mathbf{v}_S}{c}},$$

where $\mathbf{v}_S$ is the speed of the source, $\mathbf{v}_R$ the speed of the receiver and $\mathbf{n} = \frac{\vec{SR}}{\|\vec{SR}\|}$ is the source-to-receiver direction.

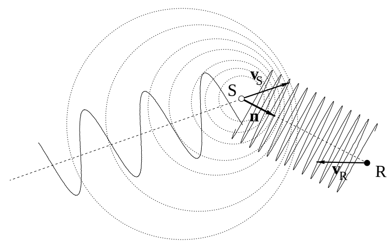


Figure 14: Notations for Doppler shifting.

Doppler shifting can also be expressed in time domain. If we note $\tau(t)$, the time-variant propagation delay between the moving source and receiver, the signal reaching the receiver at time t is expressed as:

$$r(t) = s(t - \tau(t)),$$

where $s(t)$ is the signal emitted by the source at time t. The received signal can thus be expressed by resampling the emitted signal according to the propagation delay $\tau(t)$, which can be expressed as:

$$\tau(t) = \frac{1}{c}\|\mathbf{R}(t) - \mathbf{S}(t - \tau(t))\|,$$

where $\mathbf{R}(t)$ and $\mathbf{S}(t)$ are the relative locations of the receiver and source at t (note that this expression considers a mobile source relative to a fixed receiver at time t). This equation is not linear in τ and cannot be solved directly. But, it can be approximated by a recursive process [90]. A comparison of time- and frequency-domain approaches is provided in [122].

4.2.3 Sound Reflection Models

For virtual acoustics applications, surfaces are generally assumed to be pure specular reflectors of sound waves (Figure 15). This assumption applies when the size of bumps on a surface are significantly smaller than the wavelengths of sounds, and when obstacles are significantly bigger than sound wavelengths.

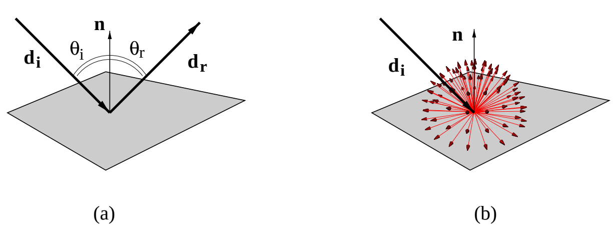


Figure 15: (a) specular reflection: $\theta_r = \theta_i$, (b) diffuse lambertian reflection.

The most common sound reflection model, valid for plane waves and infinite planar surfaces, expresses the complex pressure reflection coefficient as:

$$R(\theta, f) = \frac{\zeta(f)cos\theta - 1}{\zeta(f)cos\theta + 1},$$

where f is the frequency and $\zeta(f) = Z(f)/\rho c$ is the ration of the frequency-dependent specific impedance of the material to the characteristic impedance $\rho\, c$ of the medium[1]. Each frequency component of the original

[1] $\rho\, c = 414 kg.m^{-2}.s^{-1}$ for air in normal conditions

signal must be multiplied by the complex reflection coefficient to yield the final reflected pressure. The exact expression for the reflection of a spherical wave off an impedant surface is far more complicated [117] and, to the authors knowledge, has not made its way into interactive acoustics simulations.

For *locally reacting* surfaces, it can be assumed that ζ is independent of the angle and thus can be considered an intrinsic property of the material. Some experiments [113] have shown that using a scalar instead of a complex valued coefficient can lead to satisfying results in many cases. For more complex surfaces, such as porous materials, the impedance is itself a function of the incident direction. Several impedance models can be found in the literature [29, 8]. Complex impedances or pressure reflection coefficients can be measured on a sample of the material [67, 20], although good measurements are usually difficult to obtain.

When a significant amount of surface detail is present, a common technique in room acoustics simulation is to model the surface as a simple plane and consider it as a pure diffuse (lambertian) reflector. This is analogous to a diffuse surface in graphics. However, unlike graphics, it is difficult to model diffuse reflections with a single attenuation coefficient. Due to the possibility of interferences, diffuse reflection in sound cannot be represented by a single, independent propagation path. Hence, longer filters must be used to model the contribution of all possible diffusely reflected paths. Such filters are usually modeled using a colored noise signal, whose envelope is related to the amount of energy exchanged between surfaces. For additional details for the use of diffuse reflection in room acoustics and auralization, see [17, 122, 75].

4.2.4 Sound Diffraction Models

When the wavelength of the sound wave is similar to the geometric feature size, diffraction becomes an essential effect. While this is not a major phenomena in computer graphics (except for extreme cases, like the surface of a CD-ROM [107]), it cannot be ignored in sound simulation, especially when large obstacles are present between the source and the listener.

Geometrical Theory of Diffraction (GTD) and its extension, the Uniform Theory of Diffraction provide a way of computing a diffraction filter for a single propagation path involving diffraction over a polyhedral edge in an environment [64, 70, 82]. Because the diffraction for an entire edge can be approximated by a single (shortest) path, this model fits well with the geometrical acoustics approaches discussed in these notes, and it is practical to use in interactive virtual environment applications [123].

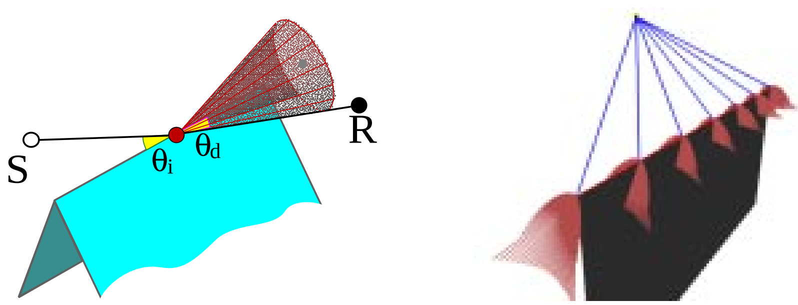


Figure 16: The Geometric Theory of Diffraction approximates diffraction of a ray incident to a wedge as a cone of diffracted rays, such that $\theta_i = \theta_d$ (left). For each diffraction, a complex diffraction coefficient can be computed. Values of the UTD coefficients are visualized on the right for point-source radiating over a half-plane.

According to the GTD, a ray incident on an edge in the environment gives rise to a cone of diffracted rays such that the angle between a diffracted ray and the edge is the same as the angle between the incident ray and the edge (Figure 16). A filter for the effect of the diffracting edge can be modeled by a complex frequency-dependent diffraction coefficient [82]. For additional details on the expression of the coefficient and how it can be used for virtual acoustics applications, we refer the reader to [123].

Other models, closer to a finite element formalism, give a more accurate time domain model of the diffraction filter [114, 118]. However, they require dense sampling of the edges in elementary point sources to construct the filter and thus are not as well suited for interactive applications.

4.2.5 Sound Occlusion and Transmission Models

Several occlusion models are also available as a simpler alternative to diffraction models in the case of obstruction by a large obstacle. In this case, the occlusion results in a "muffling" effect which can be modeled by a frequency-dependent scalar attenuation. The attenuation is usually modeled as a simple re-equalization where the input signal is decomposed in several frequency bands that are scaled independently. A way to estimate approximate attenuation factors is to use Fresnel ellipsoids to estimate a visibility factor that depends on frequency [124]. An even simpler model, used in most audio rendering APIs, globally attenuates the signal and then filters it using a low-pass filter of variable cut-off frequency [23]. Direct transmission through a wall is also often modeled using such a technique.

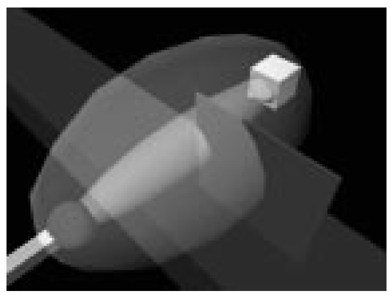


Figure 17: A 3D view of a virtual source (right), microphone (left), obstacles and first Fresnel ellipsoids computed at 400 and 1000 Hz. Occlusion ratio of the ellipsoids can be used to derive a frequency-dependent visibility factor [124].

4.3 Signal Processing for Geometric Acoustics

Once the acoustical effect of each sound propagation path has been modeled, we can construct signal processing filters that take an (anechoic) input audio signal and produce an output audio signal spatialized according to the simulated environment (Figure 18). In this section, we present a basic signal processing pipeline for auralizing sound from geometric propagation paths. Good general overviews of this process appear in [77, 66]. A recent description of auralization for a virtual environment can be found in [100].

The signal processing for each geometric path generally consists of 3 phases: 1) a resampling phase, 2) a "filtering" phase, and 3) a spatial output phase. We will present each phase separately. But, they can be grouped together for computational efficiency. As in most issues addressed in this course, every phase of the pipeline can be implemented using algorithms of various complexity depending on the desired trade-off between accuracy and computational cost.

A widely-used signal processing pipeline is shown in Figure 19 (e.g. [124, 132]). Each sound path is processed independently as follows. The input signal is first resampled according to the length of the path (and thus the propagation delay). This stage is usually implemented in time domain using a variable delay line (note that variable delay lines account for the proper Doppler shift). To implement the integer part of

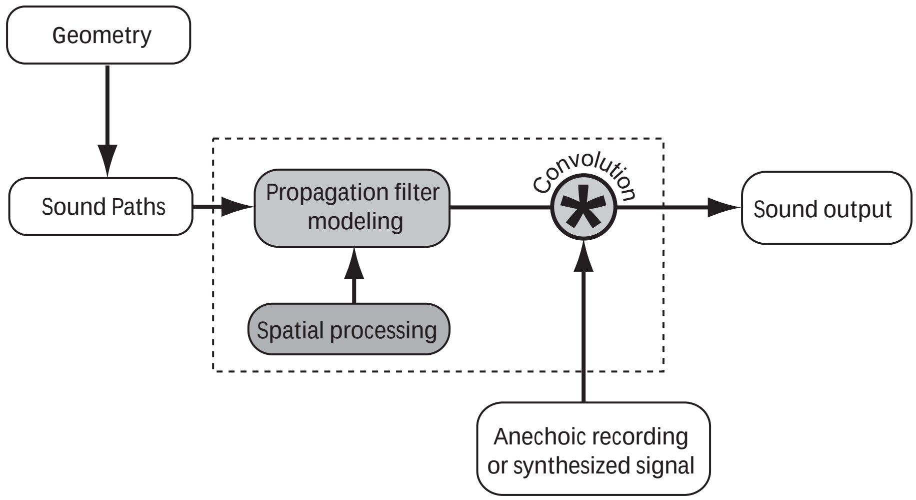


Figure 18: Auralization pipeline: From the geometry, propagation paths are constructed between each sound source and the listener. Then, for each propagation path, a digital filter is created and is convolved with the source signal. Spatial processing can be implemented in the pipeline to reproduce 3D positional audio effects.

the delay, the signal is simply delayed by the corresponding number of samples. The fractional part can be implemented by interpolating between the two closest samples. Linear interpolation is used most often and generally gives good results. See [134, 112] for more details on interpolation algorithms. Next, the signal is convolved with a sequence of digital filters representing the effects of reflectance, diffraction, and other propagation effects along a propagation path. Finally, spatial filters are applied to process the signal for output with a 3D auditory display.

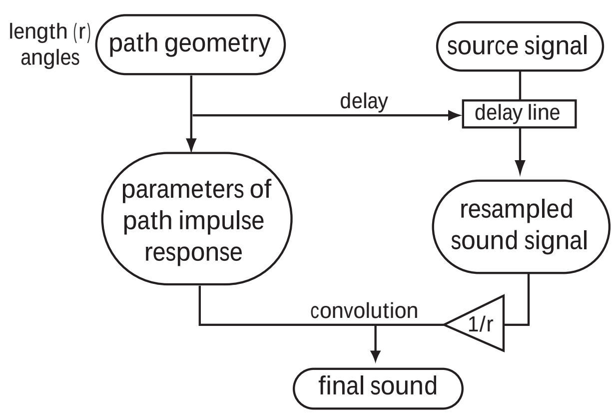


Figure 19: Signal processing pipeline for a sound path.

An alternative signal processing pipeline constructs the complete impulse response of the environment by superimposing all the filters for each propagation path. Convolution with the input audio signal is then delayed to the final stage of the process at the expense of having to convolve a longer filter. Although specific hardware is available to achieve such long convolutions in real-time [97] (e.g. Lake DSP's Huron workstation), this method is not well adapted to dynamic environments since long filters cannot easily be interpolated.

When the source and the receiver can move around during an interactive simulation, the attributes of geometric paths (length in particular) and the corresponding DSP parameters (delays, etc.) are usually modified at a rate slower than the audio sampling rate, which can cause clicking or popping artifacts to appear in the output audio. Hence, signal processing parameters (e.g. delay) are usually computed for blocks of samples and then linearly interpolated for every sample (from the previous block to the next). This introduces a latency in the pipeline corresponding to the size of the audio processing block. Another option is to use an extrapolation mechanism which allows to run the auralization process and the DSP-parameters update process at two different rates [119]. For instance, geometric calculations are performed at 10-20 Hz, while audio rendering is performed at 20-100 Hz. The extrapolation mechanism always provides the audio rendering pipeline with smooth parameters to be used for DSP operations and is well suited to approaches similar to [39]. This can also be useful if the updates of the parameters are irregular or slow (e.g. heavy processor load, update through a lossy network transmission, etc.).

5 Artificial Reverberation Algorithms

Another approach to providing reverberation in a real-time system is based on parametric models. For example, an artificial reverberator is shown in Figure 20. In this case, two input signals are delayed and passed to the early reflection and late reverberation blocks. The early reflections are created by tapping the input delays and passing the summed signals through all-pass filters. While this approach does not provide an accurate model of a specific acoustic environment, it does provide plausible models for late reverberation, and it provides a simple and efficient parameterization of synthetic reverberation effects. Thus, it is commonly used for providing late reverberations in video games.

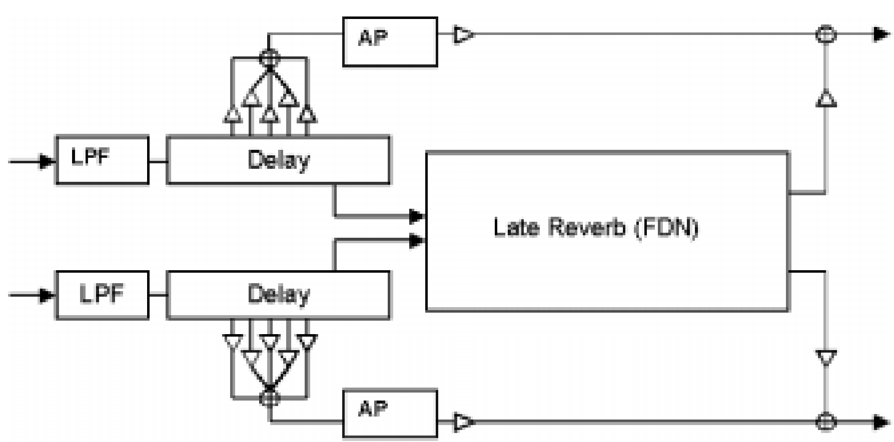


Figure 20: Example artificial reverberator.

The historical and theoretical background of artificial reverberation algorithms is reviewed in [42] and [59]. The use of feedback delay networks (FDNs) for modeling late reverberation is justified in the framework of the statistical theory of reverberation [74] and relies on the condition that sufficient overlap (or "density") of acoustic modes in the frequency domain and of reflections in the time domain are achieved [102, 59, 57]. Under this assumption, late reflections (both in a real room or in the reverberator's response) can be modeled as a Gaussian exponentially decaying random signal, characterized by a spectral energy envelope, denoted $E(\omega)$ and the reverberation time (or decay time) vs. frequency, denoted

$T_r(\omega)$ [93, 59, 57]. Hence, such techniques can complement a geometrical simulation in order to efficiently reproduce late reverberation effects while maintaining an accurate modeling of the early part of the reverberation.

5.1 Feedback Delay Network (FDN)

In a feedback delay network reverberator, the resonating behavior of an environment is characterized by a feedback matrix **A** which connects the outputs and inputs of the delay units in the network, according to the model introduced by Stautner and Puckette [108] (Figure 21).

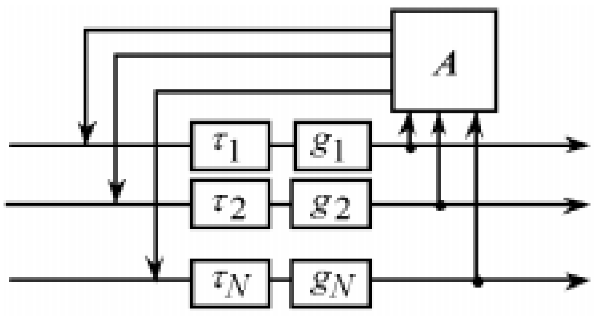


Figure 21: Basic feedback delay network.

A general framework for optimizing the topology of the FDN and the control of reverberation decay characteristics independently was proposed in [58, 59]. In this framework, the modal density and the echo density of the reverberation are controlled by adjusting the delay lengths, while the decay characteristics are controlled by associating a frequency-dependent attenuation to each delay unit. A "prototype network" is defined as any network having only non-decaying and non-increasing eigenmodes (which implies that all system poles have unit magnitude, and corresponds to an infinite reverberation time). Associating an attenuation $g_i = \alpha^{m_i}$ to each delay unit (where m_i is the delay length expressed in samples) then has the effect of multiplying all poles by α, i.e. multiplying the reference impulse response by a decaying exponential envelope [59, 26]. Frequency-dependent decay characteristics, specified by the reverberation time vs. frequency $Tr(\omega)$, are obtained by use of "absorptive filters" making each attenuation g_i frequency-dependent:

$$20log_{10}|g_i(\omega)| = -60\tau_i/Tr(\omega), \ \ (i = 1, ..N), \tag{1}$$

where $\tau_i = m_i T$ is the delay length expressed in seconds (T is the sample period).

An equivalent framework for reverberator design is given by digital waveguide networks (DWNs) [98]. A DWN is defined as a set of bi-directional delay lines connected by "scattering junctions" (or nodes), modeling a set of interconnected acoustic tubes. In this approach, a reverberator is designed by building a prototype DWN having lossless scattering junctions and then introducing frequency-dependent losses in this network. The practical implementation involves splitting each bi-directional delay line into a pair of (mono-directional) delay units.

The total delay length $\sum_i \tau_i$ in the network's feedback loop equals the modal density of the artificial reverberator (i.e., the average number of eigenmodes per Hz). Perceptually adequate modal overlap can be achieved by making $\sum_i \tau_i$ at least equal to one fourth of the decay time [102, 59]. With an appropriate choice of the feedback matrix and care to avoid degenerated distributions of delay lengths, the impulse response can be made indistinguishable from an exponentially decaying random Gaussian noise with a frequency-dependent decay rate. Designing the absorptive filters according to Eq. 1 maintains a constant frequency density along the response by imposing a uniform decay of all neighboring modes at any frequency, and thus avoiding isolated “ringing modes” in the reverberation tail.

5.2 Unitary-Feedback Delay Networks (UFDN)

The use of a unitary (energy-preserving) feedback matrix, i.e., verifying $\mathbf{A}^* \mathbf{A} = \mathbf{I}$, where $\mathbf{A}^*$ denotes the (Hermitian) transpose of $\mathbf{A}$ and $\mathbf{I}$ denotes the identity matrix, was proposed in [108]. It can be shown that this choice yields a prototype network, i.e., forces all system poles to lie on the unit circle.

The unitary character can be defined not only for mixing matrices, but, more generally, for N-input, N-output delay networks: a network is said to be unitary if its matrix transfer function $H(z)$ is unitary for any complex variable z on the unit circle, or, equivalently, if signal energy is preserved through the system [44]. Unitary networks are thus defined as the multichannel equivalent of all-pass filters.

It can be shown that any FDN whose open loop forms a unitary (or all-pass) network has all of its poles on the unit circle [59, 26]. This means that a general class of arbitrarily complex topologies for artificial reverberation is defined by cascaded or embedded unitary networks and all-pass filters in a multichannel lossless feedback loop.

5.3 Practical FDN design

The topology of Figure 21 can be regarded as only one particular way of designing a multichannel unitary feedback loop. However, whatever the topology of the prototype FDN, it can always be represented as a bank of delay units whose outputs and inputs are connected by a “canonic feedback matrix” having only scalar entries (for generality, additionnal input and output matrices are also necessary, connecting the bank of delay units to the input and output channels).

When the canonic feedback matrix has a low enough crest factor (i.e., all matrix coefficient have similar magnitudes), 8 to 16 delay units adding up to about 1 second are sufficient in practice to provide sufficient density in both the time and frequency domains, even for very long or infinite decay times. For reverberators with multiple input and/or output channels, the prototype network should be made to behave as a multichannel noise generator where the set of impulse responses associated to the different input/output channel combinations are mutually uncorrelated white Gaussian noises with equal variance. This can be readily obtained with a unitary feedback matrix in the canonic topology of Figure 21, which yields a reverberator appropriate for simulating a diffuse sound field as shown in Figure 22.

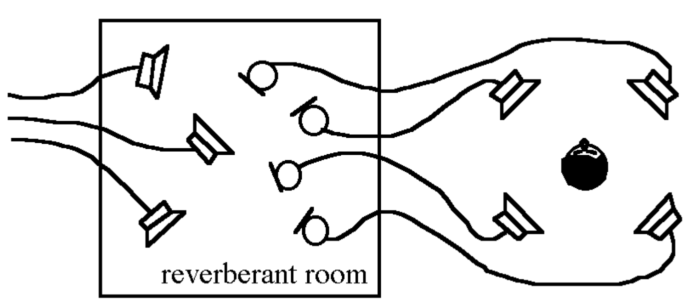


Figure 22: Conceptual analog of multi-channel reverberator simulating diffuse-field reverberation.

Examples of practical UFDN topologies can be found in [59, 57, 42, 26]. As an illustration, Figure 23 shows a reverberator comprising two chains of six absorbent all-pass filters and one delay line, each of which are fed back through an energy preserving matrix $\mathbf{M}$ as shown in Figure 4. Two independent output signals are obtained by tapping the chains after each absorbent all-pass filter.

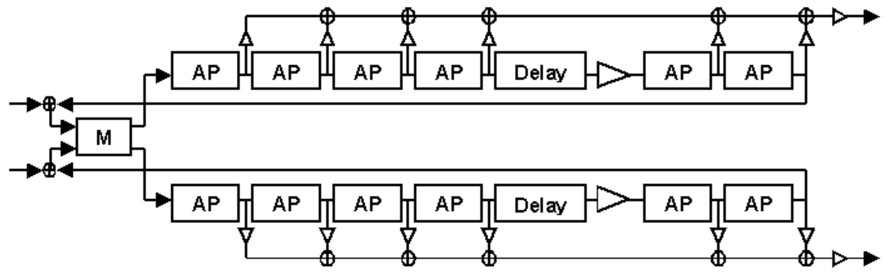


Figure 23: Example FDN reverberator based on cascaded all-pass filter chains.

5.4 Control of time-frequency decay characteristics

In order to control the reverberation spectrum $E(\omega)$ and the decay time $Tr(\omega)$ independently, it is necessary to predict how the power gain of the FDN is affected by the attenuation introduced by the absorptive filters. Assuming a unitary feedback matrix, the total power gain of the loop is $k = \sum_i {g_i}^2/N$, where g_i is given by Eq. 1, and the power gain of the FDN is given by $k + k^2 + k^3 + ... = k/(1-k)$. Therefore, the spectrum $E(\omega)$, the reverberation time $Tr(\omega)$ and the delay lengths τ_i can be controlled independently by inserting a correcting filter $c(\omega)$ in cascade with the FDN:

$$|c(\omega)|^2 = E(\omega)(1/k - 1), \text{ where } k = \sum_{i=1}^{N} (10^{-6\tau_i/Tr(\omega)})/N \tag{2}$$

Equations 1 and 2 provide explicit control over the reverberation time $Tr(\omega)$ and the spectrum $E(\omega)$ with an error smaller than a few percents or a fraction of a dB, respectively. The design of the filters $g_i(\omega)$ and $c(\omega)$ can be optimized by a dedicated analysis-synthesis procedure to simulate the diffuse reverberation decay of an existing room, with arbitrary accuracy and frequency resolution [60, 57].

With this design methodology, an inexpensive FDN can simulate the late diffuse reverberation of a room with a degree of accuracy and naturalness comparable to that of a convolution technique, initially reported in [87], using an exponentially decaying noise signal to model the impulse response. However, even when compared to fast zero-delay convolution algorithms, FDNs yield more efficient implementations and offer the advantage of providing several independent input or output channels for no additional processing cost. Perhaps more importantly, FDNs also allow unlimited reverberations times and control of decay characteristics through a small set of filter coefficients. Parametric control of the decay time vs. frequency can be implemented with simple 1st- or 2nd-order absorptive filters, for a wide range of applications.

6 3D Auditory Display

The final stage of the auralization pipeline is to reproduce a three-dimensional sound field for the ears of the listener. Here we use 3D auditory display devices to deliver the sound to the user. The goals are similar to those of stereo or holographic displays for imagery. For virtual environments, it is especially important that the auditory display device at least produce directional sound waves that provide 3D localization cues.

In this section, we provide an overview of several common 3D auditory display techniques, comparing them in terms of setup, directional range, directional accuracy, size of sweet spot, robustness of imaging, sensitivity to loudspeaker placement, availability of a corresponding recording setup, and complexity. Finally, we conclude with some remarks related to the use of such techniques in immersive virtual reality systems.

6.1 Binaural and Transaural Techniques

One class of techniques focuses on recreating the wave field at both ears of the listener using either headphones (binaural techniques) or loudspeakers (transaural techniques) [84, 83, 61].

Human auditory 3D localization is based on the effects of resonances inside the ear and scattering in the vicinity of the head and upper body. The dominant cues are *interaural time delay* (ITD) and *interaural level difference* (ILD). These cues depend on the incident direction of the sound wave, and curvature of the incident wavefronts, which in turn depend on head shadow, pinna and ear canal filtering, and shoulder echoes. For additional details on human spatial hearing, we refer the reader to [12, 9, 31]. See [104] for additional information on perception in reverberant environments.

Modeling and measuring HRTFs

Head Related Transfer Functions (HRTFs) provide filters that model the overall effects of head, ear and torso on sound propagation. For binaural and transaural auditory displays, an HRTF filter should be applied for every geometric propagation path according to the direction of sound waves traveling along the path as they reach the listener.

HRTFs can either be simulated or modeled. A simple approximation is to model the head of the listener as a simple sphere and derive corresponding diffraction filters. Other more sophisticated techniques use a 3D model of the head and run boundary element simulations to derive the filters [62]. HRTFs can also be measured directly for a real listener by placing small microphones either directly at the entrance or inside the ear canal. Dummy-heads and manikins are also available which can be used for binaural recordings or HRTFs measurements. The MIT media lab made available HRTFs measurements of a KEMAR dummy head (see Appendix B for details). Techniques have also been proposed to smoothly interpolate HRTFs, thus reducing the amount of data to measure and allow for artifact-free rendering of mobile sources (see [43] (part 5) and [127]). However, HRTF data is not usually publicly available and hence, every implementation of a binaural rendering system usually relies on an ad-hoc HRTF set. Also, although it is recognized that the curvature of the wave-fronts (i.e. distance of the source to the listener) impacts the head related effects, most HRTF sets are measured at a single reference distance from the user (e.g. 1 meter). Rendering of distance cues usually relies only on attenuation and atmospheric scattering Although such techniques can lead to satisfying results when combined with reverberation effects, it does not produce convincing results when a non-reverberant soundfield is simulated. Measuring a set of HRTFs which also depends on the distance, as suggested by [103], might be a solution to externalization problems encountered in binaural rendering. Another problem might arise from the headphone itself. Indeed when a listener wears headset, an acoustic cavity forms between the headphone and the eardrum. The resonances of this cavity and the response of the headset itself from an additional transfer function that can significantly impact the quality of the reproduction. For best results, it is thus advised that this additional transfer function is taken into account.

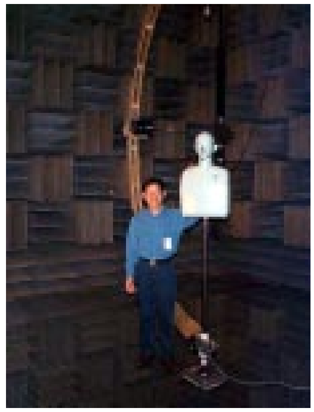


Figure 24: HRTF measurement setup in Bell Labs' anechoic chamber.

Adapting HRTFs

HRTFs vary upon individuals due to the specificity of one's head/ear/torso morphology [115]. The use of non-individual HRTF for binaural synthesis entails several perceptual artifacts such as increased inside-the-head localization and increased front-back confusion. Adding early reflections and reverberation provides enhanced distance cues and improved outside-the-head localization (externalization). However, to maximize the quality of HRTF filtering it may be necessary to adapt the filters to the particular morphology and hearing of each individual. This is usually achieved by warping some sailent frequency-space features of HRTFs filters [53, 54]. Currently, no system provides a simple, user-controlled way to calibrate a set of HRTFs to a particular user. However, some related work has appeared in the context of transaural rendering [56].

Efficient implementation of binaural filtering

Binaural filtering can be efficiently implemented using a principal component analysis of the filters (see [43] (part 5) and [127]). This allows, for instance, for efficient rendering of multiple sound paths in the context of a simulation based on geometrical acoustics. In such a model, HRTF filters for all possible incoming directions are expressed as weighted contribution of a set of eigenfilters. The cost of rendering multiple sources is not linear in the number of sources anymore, but instead also depends on the number of eigenvectors chosen to represent the HRTFs. If the number of source is greater than the number of eigenvectors, the number of instructions is reduced: $MN + MK$ operations per sample required instead of KN for K sources and M eigenfilters of length N. For a description of the corresponding processing pipeline we refer the reader to [43] (part 5).

Cross-talk cancellation and transaural stereo

Binaural reproduction can also be achieved over a pair of loudspeakers, a process referred to as transaural restitution. In this case, the signal emitted by the left speaker will reach the right ear of the listener and vice versa (see Figure 25). Hence, the cross-talk needs to be cancelled to reproduce the correct restitution. For more information on how to implement transaural filtering we refer the reader to the extensive literature on the subject [41, 83, 43].

Transaural stereo techniques suffer from a limited sweet-spot and are usually limited to desktop use. They also suffer from frequent front back reversal problems, although recent *double transaural* approaches

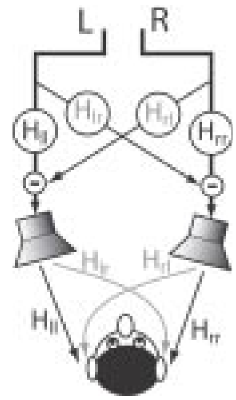


Figure 25: Transaural audio rendering: a binaural signal is playedback through loudspeakers requiring the cross-talk between the speakers and the ears of the listener to be cancelled.

improve the restitution by rendering frontal sources on a frontal stereo pair of speakers and back sources on an additional pair of speakers located behind the listener.

6.2 Multi-Channel Auditory Displays

Another class of techniques constructs a 3D sound field using an array of loudspeakers. In this case, speakers are placed around the listening area to reproduce directional sound waves. This method can reproduce correct localization cues without the expense of HRTF filtering. Such techniques are usually used for large audiences, but they tend to suffer from sweet spot problems.

Multi-channel amplitude/intensity panning

A simple and efficient technique for multi-speaker output is Vector-Based Amplitude Panning (VBAP) [95]. The idea is to locate the triplet of speakers S_i (corresponding to directions D_i) closer to the desired sound incoming direction D and use gains g_i that satisfy:

$$D = \sum_{i=1}^{3} g_i D_i$$

Other solutions are based on an optimization process [30].

Amplitude panning is fundamentally based on our auditory perception, especially the way we perceive the combination of two sources [12] as “phantom images”. Assuming the sources are equidistant to the listening position, increasing or decreasing the gain for one source will respectively shift the perceived source location towards or away from the real sources [2]. Although this is mostly true for stereo listening and sources in front of the listener, it is possible to generalize this principle to an arbitrary 3D array of loudspeakers.

For all amplitude-based panning techniques, a good compromise must be found between the number of active speakers for a given direction (thus the sharpness of sound imaging) and the smoothness of the generated sound-field when a source is moving around a listener. This problem is especially evident when a virtual source enters the reproduction region (see Figure 26). Amplitude panning does not support such a case and it is impossible to render virtual sources close to the listener.

[2] this is also true if one source is delayed relative to the other but produces much less stable results

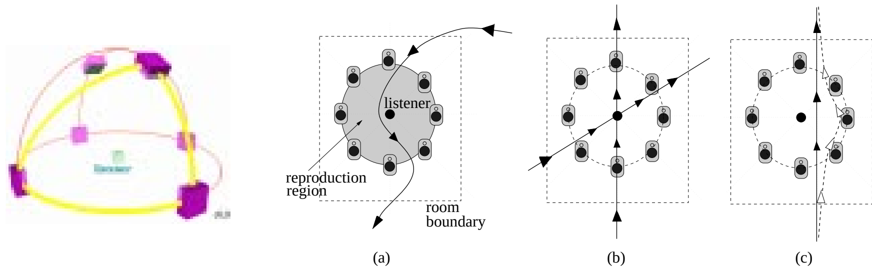


Figure 26: Lefthandside image: Triplets of speakers can be used to achieve 3D panning over arbitrary loudspeaker arrays [95]. Righthandside images: (a) A mapping needs to be established between the real reproduction space and the virtual world (b) the direction change gets very sharp when the virtual source moves close to the virtual listener. (c) Real loudspeaker position tends to affect the perceived location of the virtual sound source. Amplitude panning techniques have difficulties reproducing trajectories crossing the reproduction region.

For "surround-sound" (e.g. 5.1/7.1) like setups, computing the gains is even easier since all speakers lie in a plane and the problem becomes two-dimensional. Of course such systems do not allow for reproduction of elevation of the virtual sound source [3]. Recently multi-channel recording setups have appeared, which combine the inputs of several microphones to give a 5.1 soundtrack (see Appendix B)

6.2.1 Ambisonics

Ambisonics [45, 81] is a case of amplitude panning that uses a spherical harmonics decomposition of the directional pressure-field incident to the listener location.[4] Classic ambisonic encoding uses four channels (i.e. 1st order spherical harmonics) to model the directional soundfield. From these four channels, corresponding to an omnidirectional (pressure) and three "figure-of-8" (pressure gradient) coincident recordings, the spatial soundfield can be reproduced over multiple loudspeakers using several decoding techniques [27].

Ambisonics can be considered a sound field representation format since it encodes in four channels all the spatial information. Ambisonics encoded soundfields support straightforward transformations such as rotations, "perspective" corrections, etc.

Moreover, a recording device, the *Soundfield* microphone, is available that can record the four coincident channels required for the spatial reproduction. The four channels can also be encoded into two channels for recording onto traditional stereo media or converted to standard 5.1 surround using a dedicated processor.

Ambisonics can also be extended to higher orders although in that case no recording device is currently available on the market. For further details on Ambisonics we also refer the reader to the links in Appendix B and the numerous papers of Michael Gerzon.

6.2.2 Wave-field synthesis

Wave-Field Synthesis (WFS) aims at reproducing the exact wave fronts inside the listening space [11]. WFS is based on the Kirchoff integral theorem stating that any wave front can be reconstructed inside a close volume from a distribution of monopole and dipole sources located on the surface enclosing the volume[5] (see Figure 27):

[3] such systems are also called *pantophonic* as opposed to *periphonic* systems that can reproduce full 3D effects
[4] see http://mathworld.wolfram.com/SphericalHarmonic.html for more info on spherical harmonics
[5] no source must be present inside the volume

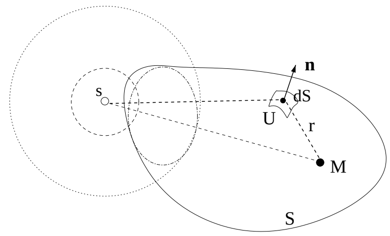


Figure 27: Notations for the Kirchoff integral theorem.

$$\hat{P}(M) = \iint_S \hat{P}(U)\nabla\left(-\frac{e^{ikr}}{4\pi r}\right)\cdot\mathbf{dS} - \iint_S -\frac{e^{ikr}}{4\pi r}\nabla\hat{\mathbf{P}}(U)\cdot\mathbf{dS}, \quad (3)$$

où $\mathbf{dS} = \mathbf{n}dS$ ($\mathbf{n}$ is a unit vector), $-\frac{e^{ikr}}{4\pi r}$ is the Green function representing the propagation of a spherical wave in free field from an arbitrary location.

WFS usually requires a large array of loudspeakers and heavy processing resources to feed all the channels. Current implementations usually limit the speakers to a line array at listener's height all around the reproduction room and use about 120-160 output channels. In this case, only monopole sources are required and correction terms can be included to account for the fact that loudspeakers cannot be considered as true monopole sources [11].

6.3 A Comparison of 3D Auditory Displays

Table 6.3 gives an overview of several aspects of the previously described techniques. Due to the countless variations in implementations and optimizations of the various techniques, it is difficult to make a simple comparison. Most techniques, except for wave field synthesis reproduce the sound field for a specific listening point. Binaural and transaural techniques directly attempt to model the sound field at both ears, while techniques based on loudpeaker arrays reconstruct the sound field at the center of the reproduction setup (and usually degrade quickly as the listener moves off-center). Multi-channel panning techniques are simple and efficient, but are more limited in imaging quality than Ambisonic techniques. Wave-field synthesis is uniquely able to reconstruct the correct wavefronts everywhere inside the listening region and is thus a true multi-user reproduction system. However, the inherent complexity of a WFS setup has, to date, prevented its use in virtual environment systems.

Choosing a spatial audio reproduction system is a very difficult task and strongly depends on the configuration and constraints of the listening space and the application (constraints relative to the display, number of simultaneous users, importance of accuracy in the restitution). A thorough investigation has yet to be performed in order to determine which setup best supports which application or virtual reality setup. A excellent overview of spatial sound reproduction techniques can be found in [28] (available only in french).

Restitution technique	Listening point	sweet spot	quality of imaging	limitations	# channels
Binaural	unique	n/a	excellent	mono-user	2
Transaural	unique	very small	excellent	mono-user	2 (4 for extended transaural)
Amplitude panning	unique	very small	average	fails for virtual sources inside reproduction region	4 and more
Ambisonics	unique	small	good	fails for virtual sources inside reproduction region	$(N+1)^2$ for periphonic N^{th} order ambisonics $2N+1$ for pantophonic N^{th} order ambisonics usually requires M speakers, $M > N$
WFS	global	n/a	excellent	heavy setup	100+

6.4 Additional Considerations for Immersive VR Systems

Auralization in immersive VR applications requires addressing additional issues:

System Latency and Update Rates

System latency and update rates can have significant impact on the perceived quality of any virtual acoustics simulation. For instance, [99] shows that binaural virtual source localization is degraded when overall system latency is larger than 96ms. Similarly, an update rate of 10Hz degrades the speed at which the user is able to localize virtual sources and produces azimuth errors. When updates rates become higher than 20 Hz, almost no difference is noticed. Refer to [131] for more details.

Listening-point and viewpoint consistency

When reproducing audio for immersive (first-person) virtual environments, one should pay attention to the relationships existing between the viewpoint in the virtual space and the listening point in the real world reproduction space. Relative position of the speakers to the listener in the reproduction environment should be taken into account to achieve consistent results. The virtual viewpoint (camera position) used for display should also be consistent with the virtual listening location. We note here that although techniques such as wave field synthesis reproduce the soundfield at any location in the listening area, the point of view is usually unique in virtual environment.

User tracking and adaptive systems

Auditory displays often utilize tracking of the listener's position. For binaural and transaural systems, tracking allows a consistent presentation of the auditory scene relative to the motion of the listener's head. Moreover, it enables adapting the cross-talk cancellation system to reduce sweet spot problems in the case of transaural recording. For speaker-based systems, tracking can be used to correct speaker gains relative to the listener's location in a reproduction space.

Loudspeakers arrays and wall reflections

Adding an audio reproduction system to an immersive virtual reality environment can be a challenging problem. Sound reflection off the screen surfaces in CAVE-like setups is a well know acoustical problem, which to date has no satisfying solution. A similar problem is the focusing of sound by cylindrical screens in Reality-Center like installations. In this case, the use of a tilted conical screen and diffusers on the ceiling can help avoiding formation of sound caustics in the room.

7 Validation of Acoustical Models

One important and difficult task when implementing a virtual acoustics rendering pipeline is validation. Relevant questions include: "Is the acoustics modeling method accurate enough?" and "Is the signal processing pipeline delivering appropriately spatialized sound to the listener?" Unfortunately, no extensive study has been conducted to date that provides definitive guidelines for answering these questions.

Validation of geometrical acoustic techniques has been studied for decades in the architectural acoustics community. There, several criteria have been designed to quantify acoustic quality, which are then used to compare the result of simulations with measurements (e.g., [10]). In the case of concert halls, a dedicated vocabulary has been designed so that acousticians and musicians can exchange their opinion about the acoustic quality of a space [10, 6, 7, 4]. Based on these subjective observations, a related set of objective criteria have been designed that describe how energy is distributed in the impulse response [10]. Three main families exist:

- *Energy decay* criteria such as *reverberation time*, *early decay time*, etc.
- *Clarity* criteria that measure a ration of early to late energy in the impulse response.
- *Binaural* criteria[6] linked to our stereophonic perception, which measure the "sensation of space" or "envelopment" perceived by the listener. Such criteria include *Lateral efficiency* and IACC, the Interaural Cross-Correlation coefficient.

Other criteria, such as *Bass ratio* and *Initial Time-Delay Gap* are also commonly used and related to the "warmth" or "intimacy" of an acoustic space. Such criteria provide an interesting framework to evaluate the quality of an acoustic simulation. But, they provide only gross comparisons for simulated and measured impulse responses.

Numerous references are available which compare the results of simulations against measurements in real spaces or using scale models [65]. The results of these experiments, show that geometric acoustics provide an efficient and accurate tool to model high-frequency acoustics within the limits of their assumptions.[7] As an example, Figure 28 shows a comparison between the early part of a simulated and computed impulse response. In this case, the simulation was carried out using a fine-band frequency domain approach using measured impulse responses of the source and wall materials. Both specular reflection and diffraction effects were included in the simulation (diffraction was modeled using the UTD). It can be seen that the model manages to captures the fine details of the impulse response. For more details, please refer to [120].

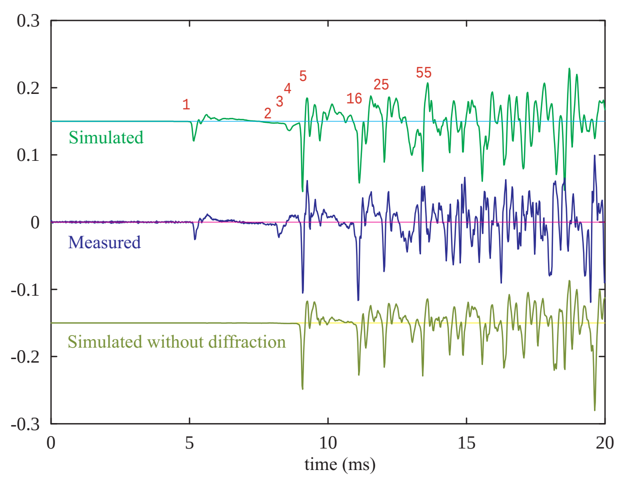


Figure 28: Comparison of the early part of a simulated and measured impulse response in the "Bell Labs Box", a simple test environment. Some of the first 55 propagation paths are labeled on the top curve out of the 1300 simulated.

[6] usually computed on a pair of impulse responses

[7] Contrary to the common belief, geometrical acoustics *are* do account for phase and other wave-related phenomena. Correct rendering of interferences is a signal processing issue and is not related to the technique used for computing propagation paths.

8 Conclusion and Future Directions

In this course, we review basic methods for auralization in interactive virtual environment applications. To summarize, common practice is to use geometric acoustic modeling algorithms to compute early reflections, feedback delay networks to fill late reverberations, and headphones or speakers to deliver spatialized sound. Current geometric algorithms are able to compute several orders of specular reflections for non-trivial environments at interactive rates, while artificial reverberators can model responses of arbitrary length. Audio technology has reached a point where advanced algorithms, hardware, and auditory display technology are becoming standard components of personal computers and home entertainment systems. Based on recent advances in both hardware and software algorithms, it seems that we have reached a time when every virtual environment application should be including spatialized sounds.

However, there is still a large amount of research still to be done on auralization for interactive applications. First of all, validation of simulations by comparison to measured data is an important topic for further study. Validation studies traditionally done for concert halls tend to compare only gross qualitative measures (e.g., reverberation time), which may have little relevance for localization and other tasks of importance in interactive virtual environment applications. Further work is required on developing new measures for comparison of impulse responses and incorporating interactive criteria into validation studies.

Utilizing human perception of sound to improve auralization is another interesting topic for further study. We believe it is possible to guide computational methods based on perceptual relevance, and it should be possible to produce better auditory displays with better psychoacoustic data. Recent studies have shown that visuals have an impact on sound perception and vice-versa [105, 111]. In particular, it has recently been shown [111] that 1) medium or high quality auditory displays coupled with high-quality displays increases the perceived quality of the visual displays compared to the evaluation of the visual displays alone and 2) that low-quality auditory displays coupled with high-quality visual displays decrease the perceived quality of the auditory displays compared to the evaluation of the auditory displays alone. However, too few results are currently available in the field of cross-modal perception. This topic deserves considerable further study.

Finally, an interesting direction of future work is to investigate the possible applications of *interactive* acoustic modeling. It is only recent that acoustic modeling has been possible in interactive applications. What can we do with interactive manipulation of acoustic model parameters that would be difficult to do otherwise? As a first application, we can imagine building a system that uses interactive acoustic simulations to investigate the psychoacoustic effects of varying different acoustic modeling parameters. Such a system could allow a user to interactively change various acoustic parameters with real-time auralization and visualization feedback. With this interactive simulation system, it may be possible to address psychoacoustic questions, such as "how many reflections are psychoacoustically important to model?," or "which surface reflection model provides a psychoacoustically better approximation?."

Throughout, researchers should investigate the synergies between sound and light and apply the lessons learned from one wave phenomenon to the other. As a historical example of this type of cross-fertilization, Turner Whitted [133] patterned his seminal ray tracing algorithm after similar methods described for acoustics in the 1960s [73]. More recently, hierarchical radiosity methods [49] developed in computer graphics have been used for modeling sound propagation [121]. Along these lines, future research could consider whether recent trends in graphics, such as image-based rendering and non-photorealistic rendering, can/should be applied in acoustics. We believe that understanding the interactions between sound and light is of critical importance to future design and engineering of interactive virtual environment systems.

References

[1] Wolfgang Ahnert. EARS auralization software. *J. Audio Eng. Soc.*, 11(41):894–904, November 1993.

[2] J.B. Allen and D.A. Berkley. Image method for efficiently simulating small room acoustics. *J. of the Acoustical Society of America*, 65(4), 1979.

[3] John Amanatides. Ray tracing with cones. *ACM Computer Graphics, SIGGRAPH'84 Proceedings*, 18(3):129–135, July 1984.

[4] Y. Ando. *Concert Hall Acoustics*. Springer-Verlag, 1985.

[5] American national standard method for the calculation of the absorption of sound by the atmosphere. *ANSI S1.26-1978, American Institute of Physics (for Acoustical Society of America), New York*, 1978.

[6] M. Barron. The subjective effects of first reflections in concert halls - the need for lateral reflections. *J. of Sound and Vibration*, 15(4):475–494, 1971.

[7] M. Barron and A.H. Marshall. Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *J. of Sound and Vibration*, 77(2):211–232, 1981.

[8] B.Brouard, D.Lafarge, J.-F. Allard, and M. Tamura. Measurements and prediction of the reflection coefficient of porous layers at oblique incidence and for inhomogeneous waves. *J. of the Acoustical Society of America*, 106(3):1493–1509, 1999.

[9] Durand R. Begault. *3D Sound for Virtual Reality and Multimedia*. Academic Press Professional, 1994.

[10] Leo L. Beranek. *Concert and opera halls: how they sound*. Published for the Acoustical Society of America through the American Institute of Physics, 1996.

[11] A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *J. of the Acoustical Society of America*, 93(5):2764–2778, may 1993.

[12] J. Blauert. *Spatial Hearing : The Psychophysics of Human Sound Localization*. M.I.T. Press, Cambridge, MA, 1997.

[13] J. Borish. Extension of the image model to arbitrary polyhedra. *J. of the Acoustical Society of America*, 75(6), 1984.

[14] Max Born and Emil Wolf. *Principles of Optics*. 7th edition, Pergamon Press, 1999.

[15] James Calvin, Alan Dickens, Bob Gaines, Paul Metzger, Dale Miller, and Dan Owen. The SIMNET virtual world architecture. In *Proceedings of the IEEE Virtual Reality Annual International Symposium*, pages 450–455, September 1993.

[16] CATT-Acoustic, Gothenburg, Sweden, http://www.netg.se/catt.

[17] C.F. Chien and M.M. Carroll. Sound source above a rough absorbent plane. *J. of the Acoustical Society of America*, 67(3):827–829, March 1980.

[18] J.H. Chuang and S.A. Cheng. Computing caustic effects by backward beam tracing. *The Visual Computer*, 11(3):156–166, 1995.

[19] R.D. Ciskowski and C.A. Brebbia, editors. *Boundary Element Methods in Acoustics*. Elseiver Applied Science, 1991.

[20] C.Nocke. In-situ acoustic impedance measurement using a free-field transfer function method. *Applied Acoustics*, 59:253–264, 2000.

[21] Michael F. Cohen and John R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, 1993.

[22] R.L. Cook, T. Porter, and L. Carpenter. Distributed ray-tracing. *ACM Computer Graphics, SIGGRAPH'84 Proceedings*, 18(3):137–146, July 1984.

[23] Creative. Environmental audio extensions: EAX 2.0 Creative©, 1999. http://www.soundblaster.com/eaudio.

[24] Lothar Cremer and Helmut A. Müller. *Principles and applications of room acoustics*, volume 1. Applied Science, 1978.

[25] N. Dadoun, D.G. Kirkpatrick, and J.P. Walsh. The geometry of beam tracing. *Proceedings of the Symposium on Computational Geometry*, pages 55–71, June 1985.

[26] L. Dahl and J.-M. Jot. A reverberator based on absorbent all-pass filters. *Conference on Digital Audio Effects (DAFX-00), Verona, Italy*, December 2000.

[27] J. Daniel, J.-B. Rault, and J.-D. Polack. Ambisonic encoding of other audio formats for multiple listening conditions. *105th AES convention, preprint 4795*, september 1998.

[28] Jerôme Daniel. *Repr'esentation de champs acoustiqies, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université de Paris VI, July 2000.

[29] M.E Delany and E.N. Bazley. Acoustical characteristics of fibrous absorbent materials. Technical Report NPL AERO REPORT Ac37, National Physical Laboratory, aerodynamics division, March 1969.

[30] G. Dickins, M. Flax, A. McKeag, and D. McGrath. Optimal 3D-speaker panning. *Proceedings of the AES 16th international conference, Spatial sound reproduction, Rovaniemi, Finland*, pages 421–426, april 1999.

[31] N.I. Durlach and A.S. Mavor. *Virtual Reality Scientific and Technological Challenges*. National Research Council Report, National Academy Press, 1995.

[32] P. Filippi, D. Habault, J.P. Lefevre, and A. Bergassoli. *Acoustics, Basic Physics, Theory and Methods*. Academic Press, 1999.

[33] S.J. Fortune. Topological beam tracing. In *Proc. 15th ACM Symposium on Computational Geometry*, pages 59–68, 1999.

[34] Steve Fortune. Algorithms for prediction of indoor radio propagation. Technical Report Document 11274-960117-03TM, Bell Laboratories, 1996.

[35] S.H. Foster, E.M. Wenzel, and R.M. Taylor. Real-time synthesis of complex environments. *Proc. of the ASSP (IEEE) Workshop on Application of Signal Processing to Audio and Acoustics*, 1991.

[36] Akira Fujomoto. Turbo beam tracing - A physically accurate lighting simulation environment. *Knowledge Based Image Computing Systems*, pages 1–5, May 1988.

[37] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West. A beam tracing approach to acoustic modeling for interactive virtual environments. *ACM Computer Graphics, SIGGRAPH'98 Proceedings*, pages 21–32, July 1998.

[38] Thomas Funkhouser, Ingrid Carlbom, Gary Elko, Gopal Pingali, Mohan Sondhi, and Jim West. Interactive acoustic modeling of complex environments. *in proceedings of The Joint Meeting of the 137th Regular Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association: Forum Acusticum '99, Journal of the Acoustical Society of America*, 105(2), March 1999.

[39] Thomas Funkhouser, Patrick Min, and Ingrid Carlbom. Real-time acoustic modeling for distributed virtual environments. *ACM Computer Graphics, SIGGRAPH'99 Proceedings*, pages 365–374, August 1999.

[40] Thomas Funkhouser, Carlo H. Sequin, and Seth J. Teller. Management of large amounts of data in interactive building walkthroughs. *ACM Computer Graphics (1992 SIGGRAPH Symposium on Interactive 3D Graphics)*, pages 11–20, March 1992.

[41] Bill Gardner. Transaural 3D audio. Technical Report 342, M.I.T. Media Lab Perceptual Computing, July 1995.

[42] W. G. Gardner. Reverberation algorithms. *in Applications of Signal Processing to Audio and Acoustics, ed. M. Kahrs, K. Brandenburg*, 1997.

[43] Steven Gay and Jacob Benesty, editors. *Acoustic Signal Processing for Telecommunications*. Kluwer Academic Publishers, 2000.

[44] M.A. Gerzon. Unitary (energy preserving) multichannel networks with feedback. *Electronics Letters*, 12(11), 1976.

[45] M.A. Gerzon. Ambisonics in multichannel broadcasting and video. *J. of the Audio Engineering Society*, 33(11):859–871, 1985.

[46] G. Ghazanfarpour and J. Marc Hasenfratz. A beam tracing with precise antialiasing for polyhedral scenes. *Computers and Graphics*, 22(1), 1998.

[47] C.M. Goral, K.E. Torrance, D.P. Greenberg, and B. Battaile. Modeling the interaction of light between diffuse surfaces. *ACM Computer Graphics, SIGGRAPH'84 Proceedings*, 18(3), 1984.

[48] Eric Haines. Beams O' Light: Confessions of a hacker. *Frontiers in Rendering, Course Notes, SIGGRAPH'91*, 1991.

[49] P. Hanrahan, D. Saltzman, and L. Aupperle. A rapid hierarchical radiosity algorithm. *ACM Computer Graphics, SIGGRAPH'91 Proceedings*, 25(4), July 1991.

[50] William Morris Hartmann. *Binaural and Spatial Hearing in Real and Virtual Environments*, chapter 10, Listening in a Room and the Precedence Effect, pages 191–210. Lawrence Erlbaum Associates, 1997.

[51] W.M. Hartmann. Localization of sound in rooms. *J. Acoust. Soc. Am.*, 74(5):1380–1391, November 1983.

[52] P. Heckbert and P. Hanrahan. Beam tracing polygonal objects. *ACM Computer Graphics, SIGGRAPH'84 Proceedings*, 18(3):119–127, July 1984.

[53] J.Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *J. of the Acoustical Society of America*, 106(3):1480–1492, 1999.

[54] J.Middlebrooks. Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. *J. of the Acoustical Society of America*, 106(3):1493–1509, 1999.

[55] C. B. Jones. A new approach to the 'hidden line' problem. *Computer Journal*, 14(3):232–237, August 1971.

[56] A. Jost and J.-M. Jot. Transaural 3-d audio with user-controlled calibration. *Proceedings of COST-G6 Conference on Digital Audio Effects, DAFX2000, Verona, Italy*, december 2000.

[57] J.-M. Jot, L. Cerveau, and O. Warusfel. Analysis and synthesis of room reverberation based on a statistical time-frequency model. *Proc. 103rd Conv. of the Audio Eng. Soc. (preprint no. 4629)*, 1997.

[58] J.-M. Jot and A. Chaigne. Digital delay networks for designing artificial reverberators. *Proc. 90th Conv. Audio Eng. Soc. (preprint no. 3030)*, 1991.

[59] Jean-Marc Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physique et perceptifs*. PhD thesis, Ecole Normale Supérieure des Télécommunications, Paris, 1992.

[60] J.M. Jot. An analysis/synthesis approach to real-time artificial reverberation. *Proc. of ICASSP*, 1992.

[61] J.M. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. *Proc. 98th Audio Engineering Society Convention*, 1995.

[62] Y. Kahana, P.A. Nelson, M. Petyt, and S. Choi. Numerical modelling of the transfer functions of a dummy-head and of the external ear. *Proceedings of the AES 16th international conference, Spatial sound reproduction, Rovaniemi, Finland*, pages 330–345, april 1999.

[63] J.T. Kajiya. The rendering equation. *ACM Computer Graphics, SIGGRAPH'86 Proceedings*, 20(4), 1986.

[64] J.B. Keller. Geometrical theory of diffraction. *J. of the Optical Society of America*, 52(2):116–130, 1962.

[65] M. Kleiner, R. Orlowski, and J. Kirszenstein. A comparison between results from a physical scale model and a computer image source model for architectural acoustics. *Applied Acoustics*, 38:245–265, 1993.

[66] Mendel Kleiner, Bengt-Inge Dalenback, and Peter Svensson. Auralization – An overview. *J. Audio Eng. Soc.*, 41(11):861–875, November 1993.

[67] Mendel Kleiner, Hans Gustafsson, and Joakim Backman. Measurement of directional scattering coefficients using near-field acoustic holography and spacial transformation of sound field. *J. of the Audio Engineering Society*, 45(5):331–346, May 1997.

[68] A. Kludszuweit. Time iterative boundary element method (TIBEM) - a new numerical method of four-dimensional system analysis for the calculation of the spatial impulse response. *Acustica*, 75:17–27, 1991. In German.

[69] S. Kopuz and N. Lalor. Analysis of interior acoustic fields using the finite element method and the boundary element method. *Applied Acoustics*, 45:193–210, 1995.

[70] Robert G. Kouyoumjian and Prabhakar H. Pathak. A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *Proc. of IEEE*, 62:1448–1461, November 1974.

[71] P. Kreuzgruber, P. Unterberger, and R. Gahleitner. A ray splitting model for indoor radio propagation associated with complex geometries. *Proceedings of the 1993 43rd IEEE Vehicular Technology Conference*, pages 227–230, 1993.

[72] U.R. Kristiansen, A. Krokstad, and T. Follestad. Extending the image method to higher-order reflections. *J. Applied Acoustics*, 38(2–4):195–206, 1993.

[73] U.R. Krockstadt. Calculating the acoustical room response by the use of a ray tracing technique. *J. Sound and Vibrations*, 8(18), 1968.

[74] Heinrich Kuttruff. *Room Acoustics (3rd edition)*. Elseiver Applied Science, 1991.

[75] K.H. Kuttruff. Auralization of impulse responses modeled on the basis of ray-tracing results. *J. of the Audio Engineering Society*, 41(11):876–880, November 1993.

[76] H. Lehnert. Systematic errors of the ray-tracing algorithm. *Applied Acoustics*, 38, 1993.

[77] H. Lehnert and J. Blauert. Principles of binaural room simulation. *Applied Acoustics*, 36:259–291, 1992.

[78] T. Lewers. A combined beam tracing and radiant exchange computer model of room acoustics. *Applied Acoustics*, 38, 1993.

[79] David Luebke and Chris Georges. Portals and mirrors: Simple, fast evaluation of potentially visible sets. In Pat Hanrahan and Jim Winget, editors, *1995 Symposium on Interactive 3D Graphics*, pages 105–106. ACM SIGGRAPH, April 1995.

[80] Michael R. Macedonia, Michael J. Zyda, David R. Pratt, Donald P. Brutzman, and Paul T. Barham. Exploiting reality with multicast groups. *IEEE Computer Graphics and Applications*, 15(5):38–45, September 1995.

[81] D.G. Malham and A. Myatt. 3D sound spatialization using ambisonic techniques. *Computer Music Journal*, 19(4):58–70, 1995.

[82] D.A. McNamara, C.W.I. Pistorius, and J.A.G. Malherbe. *Introduction to the Uniform Geometrical Theory of Diffraction*. Artech House, 1990.

[83] Henrik Møller. Reproduction of artificial-head recordings through loudspeakers. *J. of the Audio Engineering Society*, 37(1/2):30–33, 1989.

[84] Henrik Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218, 1992.

[85] M. Monks, B.M. Oh, and J. Dorsey. Acoustic simulation and visualisation using a new unified beam tracing and image source approach. *Proc. Audio Engineering Society Convention*, 1996.

[86] G.R. Moore. *An Approach to the Analysis of Sound in Auditoria*. PhD thesis, Cambridge, UK, 1984.

[87] J.A. Moorer. About this reverberation business. *Computer Music Journal*, 23(2), 1979.

[88] J.M. Naylor. Odeon - Another hybrid room acoustical model. *Applied Acoustics*, 38(1):131–143, 1993.

[89] Soren H. Nielsen. Auditory distance perception in different rooms. *J. Audio Eng. Soc.*, 41(10):755–770, October 1993.

[90] H. Noser and D. Thalmann. Synthetic vision and audition for digital actors. *EUROGRAPHICS'95*, 14(3):326–336, 1995.

[91] Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing*. Prentice-Hall Inc., 1975.

[92] A.D. Pierce. *Acoustics. An introduction to its physical principles and applications*. 3rd edition, American Institute of Physics, 1984.

[93] J.-D. Polack. *La transmission de l'energie sonore dans les salles*. PhD thesis, Universite du Maine, Le Mans, France, 1988.

[94] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. *Numerical Recipes in C, 2nd edition*. Cambridge University Press, New York, USA, 1992.

[95] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. of the Audio Engineering Society*, 45(6):456–466, june 1997.

[96] A. Rajkumar, B.F. Naylor, F. Feisullin, and L. Rogers. Predicting RF coverage in large environments using ray-beam tracing and partitioning tree represented geometry. *Wireless Networks*, 2(2):143–154, 1996.

[97] A. Reilly and D. McGrath. Convolution processing for realistic reverberation. *Proc. 98th Audio Engineering Society Convention*, February 1995.

[98] D. Rochesso and J.O. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *IEEE trans. Speech & Audio Processing*, 5(1), 1997.

[99] J. Sandvad. Dynamic aspects of auditory virtual environments. *Proc. 100th Audio Engineering Society Convention, preprint 4226*, 1996.

[100] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *J. of the Audio Engineering Society*, 47(9):675–705, September 1999.

[101] Lauri Savioja, Jyri Huopaniemi, Tommi Huotilainen, and Tapio Takala. Real-time virtual audio reality. In *Proc. ICMC 1996*, pages 107–110, Hong Kong, August 1996.

[102] M.R. Schroeder. Natural sounding artificial reverberation. *J. of the Audio Engineering Society*, 10(3):219–223, 1962.

[103] Sensaura. ZoomFX, MacroFX, Sensaura©, 1999. http://www.sensaura.co.uk.

[104] B. Shinn-Cunningham. Learning reverberation: Implications for spatial auditory displays. *Proceedings of ICAD 2000, Atlanta, USA*, april 2000.

[105] B. Shinn-Cunningham and J. Park. Changes in auditory localization responses are mediated by visual input. *J. of the Acoustical Society of America*, 101(5):3193, 1997.

[106] François X. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann Publishers inc., 1994.

[107] Jos Stam. Diffraction shaders. *ACM Computer Graphics, Proc. SIGGRAPH'99*, pages 101–110, August 1999.

[108] J. Stautner and M. Puckette. Designing multi-channel reverberators. *Computer Music J.*, 6(1), 1982.

[109] Ken Steiglitz. *A DSP Primer with applications to digital audio and computer music*. Addison Wesley, 1996.

[110] U. Stephenson and U. Kristiansen. Pyramidal beam tracing and time dependent radiosity. *Fifteenth International Congress on Acoustics*, pages 657–660, June 1995.

[111] Russell L. Storms. *Auditory-Visual Cross-Modal Perception Phenomena*. PhD thesis, Naval Postgraduate School, Monterey, California, September 1998.

[112] Holger Strauss. Implementing doppler shifts for virtual auditory environments. *Proc. 104th Audio Engineering Society Convention*, May 1998.

[113] J.S. Suh and P.A. Nelson. Measurment of transcient responses of rooms and comparison with geometrical acoustic models. *J. of the Acoustical Society of America*, 105(4):2304–2317, April 1999.

[114] U. P. Svensson, R. I. Fred, and J. Vanderkooy. Analytic secondary source model of edge diffraction impulse responses. *J. Acoust. Soc. Am.*, 106:2331–2344, 1999.

[115] Takashi Takeushi, P.A. Nelson, O. Kirkeby, and H. Hamanda. Influence of individual head related transfer function on the performance of virtual acoustic imaging systems. *Proc. 104th Audio Engineering Society Convention*, May 1998.

[116] Seth Teller. Computing the antiumbra cast by an area light source. *ACM Computer Graphics, SIGGRAPH'92 Proceedings*, 26(2):139–148, July 1992.

[117] Sven-Ingvar Thomasson. Reflection of waves from a point source by an impedance boundary. *J. of the Acoustical Society of America*, 59(4):780–785, April 1976.

[118] Rendell R. Torres. Computation of edge diffraction for more accurate room acoustics auralization. *J. Acoust. Soc. Am.*, to appear, November 2000.

[119] N. Tsingos. Artifact-free asynchronous geometry-based audio rendering. *ICASSP'2001, Salt Lake City, USA*, may 2001.

[120] N. Tsingos, I. Carlbom, G. Elko, T. Funkhouser, and R. Kubli. Validation of acoustical simulations in the "Bell Labs Box". *to appear in Computer Graphics and Applications, special issue on* Virtual World, Real Sound, July-August 2002.

[121] N. Tsingos and J.D. Gascuel. A general model for the simulation of room acoustics based on hierarchical radiosity. *technical sketch, in visual proceedings of SIGGRAPH'97, Los Angeles, USA*, August 1997.

[122] Nicolas Tsingos. *Simulating High Quality Virtual Sound Fields for Interactive Graphics Applications*. PhD thesis, Universite J. Fourier, Grenoble I, December 1998.

[123] Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, and Ingrid Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. *to appear in ACM Computer Graphics, SIGGRAPH 2001 Proceedings*, August 2001.

[124] Nicolas Tsingos and Jean-Dominique Gascuel. Soundtracks for computer animation: sound rendering in dynamic environments with occlusions. *Graphics Interface'97*, pages 9–16, May 1997.

[125] Eric Veach and Leonidas J. Guibas. Metropolis light transport. *ACM Computer Graphics, SIGGRAPH'97 Proceedings*, pages 65–76, August 1997.

[126] J.P. Vian and D. van Maercke. Calculation of the room response using a ray tracing method. *Proceedings of the ICA Symposium on Acoustics and Theater Planning for the Performing Arts*, pages 74–78, 1986.

[127] V.Larcher, J.M. Jot, G. Guyard, and O. Warusfel. Study and comparison of efficient methods for 3d audio spatialization based on linear decomposition of HRTF data. *Proc. 108th Audio Engineering Society Convention*, 2000.

[128] W.M. Wagenaars. Localization of sound in a room with reflecting walls. *J. Audio Eng. Soc.*, 38(3), March 1990.

[129] John P. Walsh and Norm Dadoun. What are we waiting for? The development of Godot, II. *103rd Meeting of the Acoustical Society of America*, April 1982.

[130] Mark Watt. Light-water interaction using backward beam tracing. *ACM Computer Graphics, SIGGRAPH'90 Proceedings*, pages 377–385, August 1990.

[131] E. Wenzel. Effect of increasing system latency on localization of virtual sounds with short and long duration. *Proceeding of ICAD 2001, Espoo, Finland*, august 2001.

[132] E. Wenzel, J. Miller, and J. Abel. A software-based system for interactive spatial sound synthesis. *Proceeding of ICAD 2000, Atlanta, USA*, april 2000.

[133] Turner Whitted. An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349, June 1980.

[134] Udo Zölzer and Thomas Bolze. Interpolation algorithms: theory and applications. *Proc. 97th Audio Engineering Society Convention, preprint 3898*, November 1994.

A Signal Processing Basics

In this section we review basic terminology of signal processing for audio signals. For more information, we refer the reader to [91, 109].

- **Digital signal**: a function considered in its discretely sampled form. In the case of an audio signal, the function is a one-dimensional function of time and represents sound pressure variations.

- **Harmonic wave:** a wave which space-time dependence has the form:

  $$p(t,x) = P\ sin(\omega t - kx), x \in \mathbb{R}^3$$

  also noted in complex exponential form:

  $$p(t,x) = P\ Re(e^{i(\omega t - kx)}), i^2 = -1$$

  where $Re(x)$ denotes the real part of the complex number x. ω and k are linked by the relation $k = \omega/c$, where c is the speed of sound. $f = \omega/2\pi$ is the frequency of the vibration and $\lambda = 2\pi c/\omega = c/f$ is the wavelength.

  Harmonic waves are basic solutions of the *wave equation* governing the acoustic state of the medium (here the air) [92]. Although no realistic sound signal is usually harmonic, harmonic waves are very important since they are the basic functions for the Fourier domain decomposition of a signal, a widely used tool in signal processing.

- **Fourier transform and frequency-domain representation:**

  In the general case, sound waves are not harmonic. However, any signal can be represented as a sum of harmonic signals. For a real signal $s(t)$, such decomposition into harmonic signals of frequency f is calculated using *Fourier analysis*:

  $$s(t) = \int_{-\infty}^{+\infty} \hat{S}(f) e^{i2\pi ft} df,$$

  where $\hat{S}(f)$ (valued in $\mathbb{C}$) is the *Fourier transform* of $s(t)$.

  Note : Similarly we have : $\hat{S}(f) = \int_{-\infty}^{+\infty} s(t) e^{-i2\pi ft} dt$, where $s(t)$ is the *inverse Fourier transform* of $\hat{S}(f)$.

  The interest of such representation relies on the hypothesis that acoustical phenomena are linear. Within this hypothesis, the solution of the wave equation for a general (non-harmonic) wave is the sum of the solutions for the harmonic waves resulting from its Fourier analysis.
  Two important functions, equally defined using the Fourier transform of a signal are:

  - the *spectrum*: the function $P(f) = |\hat{S}(f) e^{i2\pi ft}|$, which describes the frequential contents of the signal[8].
  - the *phase*: the function $\Phi(f) = \arctan\left(-\frac{\mathrm{Im}\left(\hat{S}(f)e^{i2\pi ft}\right)}{\mathrm{Re}\left(\hat{S}(f)e^{i2\pi ft}\right)}\right)$ (where $\mathrm{Re}(\hat{x})$ and $\mathrm{Im}(\hat{x})$ denote respectively the real and imaginary parts of the complex number $\hat{x}$.).

[8] here, $|\hat{x}|$ denotes the modulus of the complex number $\hat{x}$.

These two functions are sufficient to reconstruct the function $\hat{S}(f)$, and therefore, using the inverse Fourier transform, the signal $s(t)$. Hence, knowing these two functions is equivalent to knowing the signal $s(t)$ itself. For more detail, please refer to numerous signal processing textbooks, including [91, 109, 94].

- **Digital filter:** A particular type of signal. Described by its *impulse response*, i.e. the mathematical function that describes the output waveform that results when the filter is convolved with a unit impulse (Dirac signal).

- **Finite Impulse Response (FIR) filter:** a filter h which impulse response is directly given by a finite number m of sampled values (also called *taps*) through time.

  The process of applying a FIR filter to an input signal corresponds to a *convolution* of the two signals. For discrete systems, this can be expressed by:

  $$y(n) = \sum_m h(m)x(n-m), (n, m) \in \mathbb{N}^2,$$

  where x is the original discrete signal, h is the discrete filter of length m and y is the filtered signal.

  This operation can also be written in Fourier frequency domain as the product of the Fourier transforms of the two signals:

  $$Y(f) = H(f)X(f),$$

  where $Y(f)$ et $X(f)$ are the Fourier transforms of the signals $y(t)$ and $x(t)$, and $H(f)$ is the *transfer function* of the filter, i.e the Fourier transform of its impulse response $h(t)$.

- **Infinite Impulse Response (IIR) filter:** the filter has an impulse response of infinite length and the filtering operation can be described recursively in time domain by the following equation (called *direct form II transpose*):

  $$y(n) = \sum_{l=0}^{L} b_l x(n-l) - \sum_{k=1}^{K} a_k y(n-k).$$

  The value L is called the *order* of the IIR filter.

  In this case, the coefficients of the filter b_l's and a_k's cannot be determined directly. However, in some case, applying a IIR filter requires much less computational power than an FIR filter.

- **Filtering**: Applying a FIR or IIR filter to an input signal.

- **Low-pass filtering:** A specific filter preserving only the frequency components below a specified frequency. Several IIR models exist for low-pass filters.

- **High-pass filtering:** A specific filter preserving only the frequency components above a specified frequency. Several IIR models exist for high-pass filters.

- **Band-pass filtering:** A specific filter that preserves the original signal only within a specified range of frequency. Components of the signal below or above the filter's cut-off frequencies are discarded. A band-pass filter can be implemented as the combination of a high-pass and a low-pass filters.

- **Resampling:** the process of reconstructing a signal at a different sampling rate, i.e. generating sample values not originally present in the sampled signal. This is achieved by either decimating or interpolating the original samples.

- **Delay-line:** an operator that delays an audio signal by a given amount of time.
- **Re-equalization:** a term derived from the music industry. The signal is split in several frequency bands (usually octave-bands) using a bank of band-pass filters. Each subsignal is scaled by a scalar value. The output signal is then reconstructed by adding up all the subsignals.

B On-line resources

This section collects web pointers that the reader can browse for additional information.

Organizations

- The Acoustical Society of America (ASA):
  http://www.asa.aip.org
- The Audio Engineering Society (AES):
  http://www.aes.org
- The Association for Computing Machinery (ACM):
  http://www.acm.org
- The Association for Computing Machinery (IEEE):
  http://www.ieee.org
- ICAD International Community for Auditory Display :
  http://www.icad.org

Conferences

- ACM Special Interest Group on Graphics (SIGGRAPH):
  http://www.siggraph.org (On-line proceedings available to ACM members)
- International Conference on Auditory Display (ICAD):
  http://www.icad.org. (On-line proceedings available free of charge)
- COST-G6 Workshop on Digital Audio Effects (DAFX):
  - Dafx 1998 :
    http://www.iua.upf.es/dafx98/ (On-line proceedings available free of charge)
  - Dafx 1999 :
    http://www.notam02.no/dafx99/ (On-line proceedings available free of charge)
  - Dafx 2000 :
    http://profs.sci.univr.it/ dafx/ (On-line proceedings available free of charge)
- IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP):
  http://www.icassp.org
- ACM Siggraph/EUROGRAPHICS Campfire on Acoustics:
  http://www.siggraph.org and www.bell-labs.com/topic/conferences/campfire/

Labs, Research pages, Publications

- IRCAM:
  http://www.ircam.fr
- Center for New Music and Audio Technologies:
  http://www.cnmat.berkley.edu
- Thomas Funkhouser's interactive acoustic modeling page:
  http://www.cs.princeton.edu/ funk/acoustics.html
- Patrick Min's interactive acoustic modeling page:
  http://www.cs.princeton.edu/ min/audio
- Nicolas Tsingos' homepage:
  http://www-sop.inria.fr/reves/personnel/Nicolas.Tsingos/
- Perry Cook's homepage:
  http://www.cs.princeton.edu/ prc/
- Machine Listening Group at the MIT Media Lab :
  http://sound.media.mit.edu
- York University Ambisonics homepage:
  http://www.york.ac.uk/inst/mustech/3d_audio/ambison.htm
- Optimization of hall acoustics at MIT:
  http://graphics.lcs.mit.edu/ mcm/papers/audiopt.html
- Interactive Audio special interest group:
  http://www.iasig.org/
- David Griesinger's home page:
  http://world.std.com/ griesngr/
- DIVA project:
  http://www.tml.hut.fi/Research/DIVA/
- Angelo Farina's homepage:
  http://www.ramsete.com/home.htm
- Jerome Daniel's page on (higher-order) ambisonics:
  http://gyronymo.free.fr/
- HRTF Measurements of a KEMAR Dummy-Head Microphone:
  http://sound.media.mit.edu/KEMAR.html
- Dinesh Pai's homepage (sound modeling, contact sound synthesis and measurements):
  http://www.cs.ubc.ca/spider/pai/home.html
- Kees van den Doel's homepage
  (sound modeling, contact sound synthesis and measurements, sound in JAVA) :
  http://www.cs.ubc.ca/spider/kvdoel/home.html

Hardware

Processing

- Bose Auditioner:
  http://www.bose.com/technologies/prediction_simulation/html/AuditionerNfFS.html
- Tucker Davis:
  http://www.tdt.com

3D Recording

- Bruel and Kjaer Head and Torso Simulator (HATS) - Type 4128D:
  http://www.bkhome.com/
- Soundfield microphone and processors:
  http://www.soundfield.com
- Cascade audio surround microphones:
  http://www.cascade-audio.com
- Holophone surround microphone:
  http://www.theholophone.com

Audio and DSP APIs, Acoustic software

- Creative Lab's EAX extensions to DirectSound:
  http://www.eax.creative.com
- Creative Labs Developer Central:
  http://http://developer.creative.com/sitemap.asp
- OSS Open Sound System for Linux:
  http://www.opensound.com
- ALSA for Linux:
  http://www.alsa-project.org
- ESound for Linux (the Enlightened Sound Daemon) :
  http://www.tux.org/ ricdude/dbdocs/book1.html
- Intel 3D RSX:
  http://www.intel.com/ial/archive/rsx.htm or www.radgametools.com/rsxmain.htm
- Miles Sound System:
  http://www.radgametools.com/mssnew.htm
- FMOD music and sound effects system:
  http://www.fmod.org
- CATT Acoustics room simulation system:
  http://www.netg.se/ catt/

- OpenAL website:
  http://www.openal.org/home/
- Ramsete room acoustics modeling tool:
  http://www.ramsete.com/
- Some Interesting Sound and MIDI Software For Linux:
  http://www.clug.in-chemnitz.de/vortraege/multimedia/SOFTWARE/linux_soundapps.html
- Sensaura:
  http://www.sensaura.co.uk
- Odeon room acoustics software:
  http://www.dat.dtu.dk/ odeon/
- Kees Van den Doel JAVA Real-time audio synthesis system:
  http://www.cs.ubc.ca/spider/kvdoel/jass/jass.html

Tutorials, FAQs, Introductions

- DSP Guru :
  http://www.dspguru.com for FAQs on filtering
- Ambisonics:
  http://www.ambisonics.net for and intro and links to the ambisonics world
- Perry Cook's Siggraph 2000 course notes:
  http://www.cs.princeton.edu/ prc/CookSig00.pdf
- Acoustics FAQ:
  http://www.faqs.org/faqs/physics-faq/acoustics/
- Designing interactive audio content to-picture:
  http://www.gamasutra.com/features/19991220/clark_01.htm

Bibliography pages

- Virtual audio bibliography :
  http://www.hitl.washington.edu/kb/audio-bib.html
- ICAD Annotated bibliography:
  http://www.icad.org/
- Film Sound bibliography:
  http://www.filmsound.org/bibliography/littlist.htm

Additional link collections

- Ultimate Spatial Audio Index:
  http://www.dform.com/inquiry/spataudio.html

- ACM SIGSOUND links:
  http://www.acm.org/sigsound/sigsoundlinks.html
- DSP links:
  http://users.iafrica.com/k/ku/kurient/dsp/links.html
- Audio and 3D Sound links:
  http://www.users.dircon.co.uk/ wareing/3daudio.htm
- R.O. Duda's Links to the World of Spatial Sound:
  http://www-engr.sjsu.edu/ duda/Duda.R.LWSS.html
- Harmony Central Computer Music Resources:
  http://www.harmony-central.com/Computer/
- AUDITORY list:
  http://www.auditory.org discusses organizational aspects of auditory perception