Creating Interactive Virtual Auditory Environments

Tapio Lokki, Lauri Savioja, and Riitta Väänänen
Helsinki University of Technology, Finland

Jyri Huopaniemi
Nokia Research Center, Finland

Tapio Takala
Nokia Ventures Organization, Finland

We survey sound rendering techniques by comparing them to visual image rendering and describe several approaches for performing sound rendering in virtual auditory environments.

Sound rendering[1] is analogous to graphics rendering when creating virtual auditory environments. In graphics, we can create images by calculating the distribution of light within a modeled environment. Illumination methods such as ray tracing and radiosity are based on the physics of light propagation and reflection. Similarly, sound rendering is based on physical laws of sound propagation and reflection.

In this article, we aim to clarify real-time sound rendering techniques by comparing them to visual image rendering. We also describe how to perform sound rendering, based on the knowledge of sound source(s) and listener locations, radiation characteristics of sound sources, geometry of 3D models, and material absorption data—in other words, the congruent data used for graphics rendering. In several instances, we use the Digital Interactive Virtual Acoustics (DIVA) auralization system,[2] which we've been developing since 1994 at the Helsinki University of Technology, as a practical example to illustrate a concept. (The sidebar "Practical Applications of the DIVA System" [next page] briefly describes two applications of our system.)

In the context of sound rendering, the term *auralization*[3]—making audible—corresponds to visualization. Applications of sound rendering vary from film effects, computer games, and other multimedia content to enhancing experiences in virtual reality (VR).

Approaches to sound rendering

The sound rendering methods we present in this article enable dynamic rendering, in which the position of the listener or sound sources can change. Equally, the virtual world's geometry can change during the rendering process. Although a static setting would be technically much simpler, it has only limited applicability (virtual concert recordings, for example). In many VR applications, user interaction is an essential feature, and thus we must perform dynamic rendering in real time.

In sound rendering, we can add spatial cues such as sound source location or reverberation to the sound signal. These cues can be based on either human perception or the physics of sound. In many applications, the original sound signal is the essential content, and rendering only adds spatial features as effects. Then, we can obtain accurate enough results by applying perceptual sound rendering. In this approach, a detailed 3D model isn't necessary, but we can use descriptive parameters (such as reverberance or brilliance) to control a signal-processing algorithm to produce a desired auditory sensation. The perceptual approach is well suited for postprocessing music performances to create additional room acoustic effects.[4]

The physics-based approach is more appropriate when we're interested in the acoustic environment rather than the sound itself. For navigating in a virtual environment (VE), locations of objects and sound sources are necessary. For an architect or acoustics designer, the essential elements are the audible effects caused by a constructed environment's objects and properties.

Interactive audio–visual virtual environments

Figure 1 (next page) illustrates the basic architecture of a virtual auditory environment. Typically, the observer can fly around in a 3D model where light and sound sources are enclosed. Graphics rendering is based on the 3D geometry, light sources, and materials. Sound rendering shares the same (although often simplified) geometry and requires audio signal sources located in the 3D space, as well as acoustic properties of materials.

Due to differences in physics and sense organs of sound and light, the rendering system requires different signal resolutions and update rates. Human vision processes many light signals from different directions (high spatial resolution) in parallel, whereas these signals are integrated over time such that a frame rate of about 30 Hz usually suffices. Sound arriving from all directions to the ear is integrated into one signal, where

0272-1716/02/$17.00 © 2002 IEEE

Practical Applications of the DIVA System

The DIVA virtual orchestra[1] is a showcase designed for the Siggraph 97 Electric Garden. The main goal was to demonstrate several important audio–visual simulation themes (computer graphics and human animation, sound synthesis, auralization, and user interaction) in one interactive virtual concert experience.

The virtual orchestra consists of four modules:

- tracking of tempo and other musical characteristics from a live human conductor;
- visualizing animated virtual musicians and virtual spaces, as Figure A depicts;
- generating sound based on MIDI-encoded scores, especially applying physical modeling of musical instruments; and
- auralizing sound sources in acoustical spaces using binaural or multichannel reproduction techniques.

During one week at Siggraph 97, more than 700 people conducted the virtual orchestra. Since then, it has been a popular demonstration and also used in artistic installations. In one concert, the virtual players were accompanied by a real string quintet. The virtual orchestra was also part of the Second International Sibelius Conductors' Competition, where the audience was offered a chance to conduct the competition pieces in the concert hall's lobby.

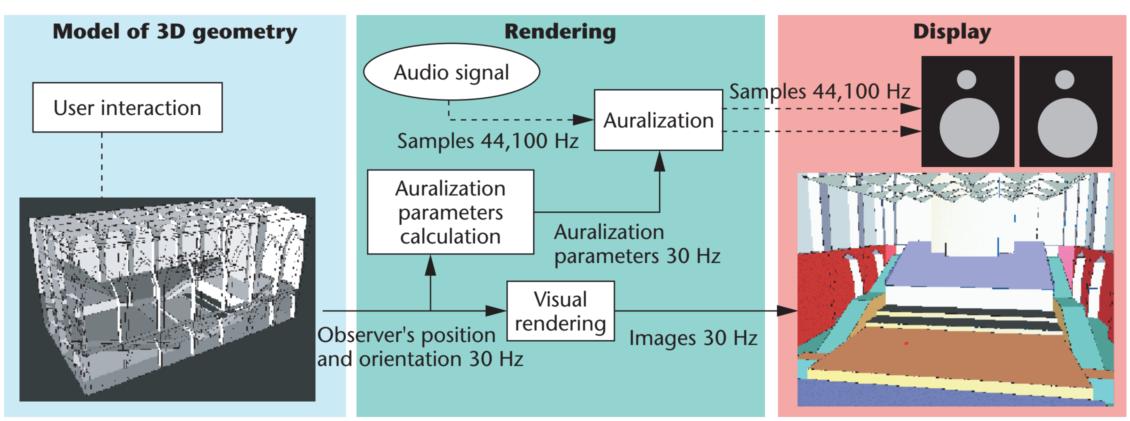
We've also applied the DIVA system in the field of room acoustic design. The Marienkirche film,[2] shown in Siggraph 98 Electronic Theater, is an example of the first high-quality audio–visual demonstration where both sound and images were automatically rendered, based on room geometry and surface data. Figure B shows one example frame.

References

1. T. Lokki et al., "Real-time Audiovisual Rendering and Contemporary Audiovisual Art," *Organised Sound*, vol. 3, no. 3, 1998, pp. 219-233.
2. T. Takala et al., "Marienkirche - A Visual and Aural Demonstration Film," *Electronic Art and Animation Catalog* (Proc. Siggraph 98), ACM Press, New York, 1998, p. 149.

A **The cartoonish animated players of the virtual orchestra, playing at the Electric Garden subway station during Siggraph 97.**

B **A snapshot of the Marienkirche film, presented in the Electronic Theater in Siggraph 98.**



1 **Processes of graphics and sound rendering. Using the 3D model and user interactions as input, they produce an audio–visual scene with a rendered image and spatialized sound.**

spectral and temporal features give information about the sound paths through the environment. Although a sample rate of more than 40 kHz is necessary for the signal to cover the whole audible spectrum, most dynamic perceptional cues are blurred over a sensory fusion interval of about 50 ms. Thus, a similar update rate (as with the visual display) is sufficient for simulating changes in an auditory environment. However, to avoid abrupt localization changes (which might cause audible clicks), we smooth these changes using interpolation.

Virtual auditory environments

The basic defining components of a visual VE are light sources, 3D geometry, and the light transmittance properties of surfaces and materials. Respectively, for virtual acoustics, we need to model sound sources, room acoustics, and in some cases (depending on the sound reproduction method), human spatial hearing properties.

Sound-source modeling

Light sources emit light at different colors, usually modeled with the three-component RGB spectrum. Lights can have different radiation characteristics, such as omnidirectional or spot-like, even with color textures. With sound sources, the emitted signal can be variable over the whole audio spectrum, and the radiation characteristics are usually frequency dependent.

Sound-source modeling deals with methods to produce sound in an auditory scene. The most straightforward way is to take prerecorded digital audio as a source signal. In such a case, however, we have little opportunity to change it interactively. Another choice is to use generative techniques such as speech synthesis, musical instrument models, or algorithms producing everyday sounds. This article doesn't concentrate on sound synthesis methods, but they are an important part of virtual auditory environments.

In most auralization systems, sound sources have traditionally been treated as omnidirectional point sources. This approximation is valid for many cases. For example, most musical instruments have frequency-dependent radiation patterns. Typical sound sources, such as the human voice, loudspeakers, or many musical instruments radiate more energy to the frontal hemisphere, whereas sound radiation gets attenuated and low-pass filtered when the angular distance from the on-axis direction increases.

Modeling room acoustics

Because sound and light are both waves—one electromagnetic and the other mechanical motion—many similarities appear in their propagation. In most cases, we can handle both as rays emanating from the sources, traveling through the air, bouncing from surfaces, and finally ending at a receiver.

With respect to traveling speed and wavelength, however, light and sound differ. In human scales, light is transmitted instantaneously, whereas the speed of sound (340 meters per second in air) causes noticeable propagation delays perceived as echoes and reverberation. The wavelength of light (400 to 740 nanometers) is negligibly small and must be considered only in special cases.

To experience VR, we need both visual and auditory displays with 3D capability.

The range of audible wavelengths covers three orders of magnitude (0.017 to 17 m) comparable in size to objects in human environments. Therefore, diffraction is a phenomenon we must consider as well. Occlusion becomes frequency dependent such that while higher frequencies get attenuated by an occluding object, lower frequencies can pass it without noticeable effect. For example, a speaking person can be heard out-of-sight around a corner, although the sound is muffled.

Another interesting phenomenon is diffuse reflections. In graphics, the topic is a popular research area, and current radiosity algorithms solve it nicely. With real-time physics-based acoustic simulations, this still is an open issue. The radiosity method solves an energy equilibrium, which we can apply to stationary noise, but it isn't suitable for auralization of time-varying sounds.

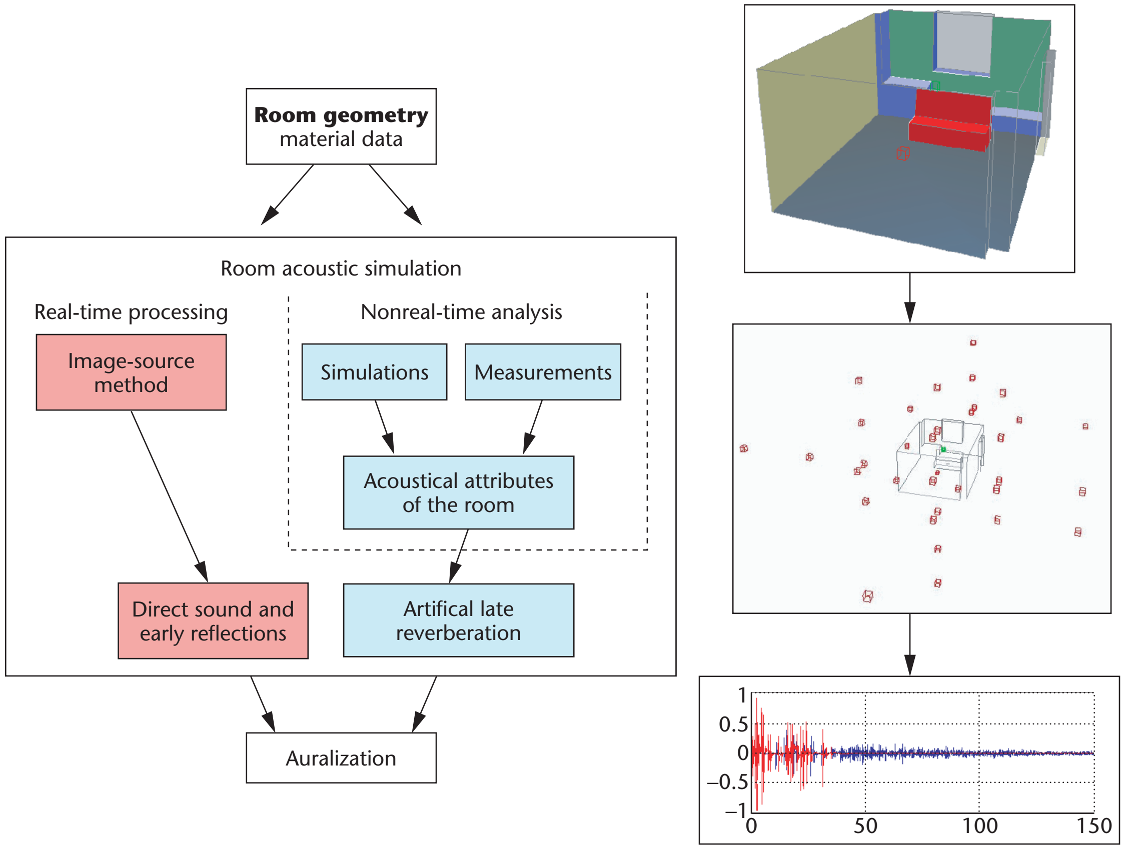
Despite the lack of diffraction modeling, the most commonly used methods in room acoustic simulation are ray based. It's interesting to note that the history of ray tracing dates back to the late 1960s in both graphics[5] and sound rendering.[6] Other ray-based techniques are the image-source method[7] and beam tracing. In addition, researchers have applied wave-based algorithms, such as finite-element (FEM), boundary element (BEM), and finite-difference-time-domain (FDTD) methods. Due to their computational complexity, these methods aren't practical through the whole audible frequency range, and we can't currently use them in real-time applications.

Ray-based methods typically neglect diffraction, but we can roughly model diffusion by dividing reflections into specular and diffusive components, as in graphics. Despite these simplifications, we can do quite reliable room acoustic simulations, and today's concert halls and other acoustically interesting spaces are often designed with the aid of computers.

Auditory displays and modeling of spatial hearing

To experience VR, we need both visual and auditory displays with 3D capability. Rendering high-quality images for both eyes independently gives a stereoscopic view. In interactive situations, we can achieve the correct projection with a head-tracking device combined with either a head-mounted display (HMD) or shutter glasses within a spatially-immersive display (SID) such as the Cave Automatic Virtual Environment.

Three-dimensional sound reproduction means that users can perceive the direction of an incoming sound. As in graphics, techniques fall into two categories: surrounding loudspeakers and headphone reproduction. A straightforward approach would be to place a loudspeaker in the direction of each virtual sound source. However, this is impractical for arbitrarily moving



2 **The three parametrization levels and hybrid method for real-time sound rendering. We derive the room impulse response (lower right) from image sources (red) and statistical reverberation (blue).**

sources. With a limited set of loudspeakers, we must tune each speaker so that the resulting sound field in the listener's ear canals is approximately correct. Both amplitude panning techniques (such as vector base amplitude panning [VBAP][8]) and decoded 3D periphonic systems (such as Ambisonics[9]) require many loudspeakers and several cubic meters of echoless space.

Headphones are more flexible than multichannel loudspeaker setups, but they make sound processing more cumbersome. A head-tracking device and modeling of human spatial hearing are necessary for this approach. The main cues for lateral directional hearing are the interaural level and time differences (ILD and ITD, respectively). These alone don't resolve elevation or the front-back confusion,[10] so a more complex head-related transfer function (HRTF) is necessary. It models the reflections and filtering by the listener's head, shoulders, and pinnae (part of the external ear), which provide more elaborate directional cues. We can measure the HRTFs on human subjects or dummy heads or derive them from mathematical models. Binaural reproduction techniques using digital filters derived from HRTFs are widely applied today, and they provide quite natural 3D sound, despite the fact that head and pinnae dimensions vary between individuals. In interactive situations, head movements also help resolve the perceived sound's direction.

Besides sound direction, the distance of a virtual sound source should also be audible. However, rendering this cue isn't straightforward because the sensation of distance results from many environmental cues simultaneously, including reverberation, attenuation of high frequencies caused by air, and the relative movement between the source and the listener.

Interactive sound rendering

We capture the essential information of sound propagation from a source to a listening point in an impulse response that contains all the information about the sound source's radiation and a room's reverberation. The most straightforward way to auralize this response is to convolve it with the stimulus signal, usually anechoic sound (free from echoes and reverberation). In this way, we add the sound propagation and reflection information to the sound—that is, render the sound through the modeled space.

A single impulse response doesn't contain information about the sound's incoming directions. In addition, an impulse response describes only sound propagation from one source point to a single listening point. For example, if the listener moves, the whole impulse response changes. For these reasons, the direct convolution approach is impractical for creating dynamic virtual auditory environments.

In dynamic sound rendering, we parameterize the impulse response used in convolution for two reasons: to make dynamic rendering possible and save computational load. In this article, we present a general three-level approach to the parameterization (see Figure 2). To give a practical example, we describe how we implemented it in the DIVA auralization system.[2] The three levels of parametrization are

- defining the scene,
- calculating sound propagation and reflections in the space, and
- auralization and audio signal processing.

Defining a scene

Similar to graphics, a scene's acoustic model contains 3D geometry. Typically, polygonal modeling is sufficient and the number of surfaces is much less than in graphics rendering. Each polygon is associated with a material described by absorption and diffusion coefficients. These factors depend on frequency and direction, but in practice, we only give them direction independently in octave bands. This established custom in acoustics is caused by the impractical and laborious measurement of direction-dependent coefficients. In addition, the model contains location and orientation of sound sources and the listener. Both the sources and listener have directivity characteristics, but there's no common practice for describing them.

The sidebar "Audio Scene Description APIs" briefly explains one possible way to use 3D audio scene description languages to hierarchically organize the necessary (acoustical and architectural) data to build up a scene's textual representation.

Real-time room acoustic simulation

Several wave-based and ray-based methods calculate sound propagation and reflections in a space. For real-time simulation, ray-based methods are widely used. The image-source method[7] is especially suitable for calculating early reflections, but we must model the late reverberation separately due to the exponential growth of the computational load when tracing higher-order image sources. A common assumption in acoustics is that the late reverberation is diffuse (direction independent) and exponentially decaying. This lets us use computationally efficient recursive filter structures. We can control these with statistical late reverberation parameters such as reverberation time and energy.

The image-source method is a ray-based method where each reflection path is replaced by an image source. We find the image sources by reflecting the original source against all surfaces. The reflecting process can be recursively continued to find higher-order image sources. In a typical room geometry, most of them are invalid because the reflection path isn't realizable. In addition, some of the image sources are invisible to the listening point because of occlusion. Thus, a visibility check similar to culling in computer graphics is needed. (We use ray casting in our DIVA implementation.) For each image source, the reflection path from the source to the receiver through all the reflecting surfaces is reconstructed inside the 3D model. Then possible occlusions (intersections with other surfaces) are calculated and checked that all reflecting points actually lie inside the reflecting surface's boundaries. To reduce the number of required intersection calculations, we use a spatial directory with adaptive spatial subdivision and hash-based addressing.

The input data for the image-source calculation is the room geometry and the material data. Based on the location and orientation of the sound sources and the listener, the calculation provides information of audible specular reflections. The parameters for each visible image source are

- order of reflection,
- orientation (azimuth and elevation angles) of sound source,

Audio Scene Description APIs

Researchers have developed several application programming interfaces (APIs) to help define 3D sound scenes, named audio scene description APIs. Here, we briefly describe common principles in different 3D sound APIs and concentrate on the sound scene description interface of the MPEG-4 standard.[1]

Defining a room acoustic model includes writing the room's geometry in terms of polygon surfaces, associated with sound reflecting properties that depend on the materials. The model definition also contains positions of the sound sources and possibly their directivity properties. The 3D sound APIs are designed to facilitate the process of writing this data in a hierarchical and object-oriented way so that information that intuitively belongs together is grouped under the same sound scene object.

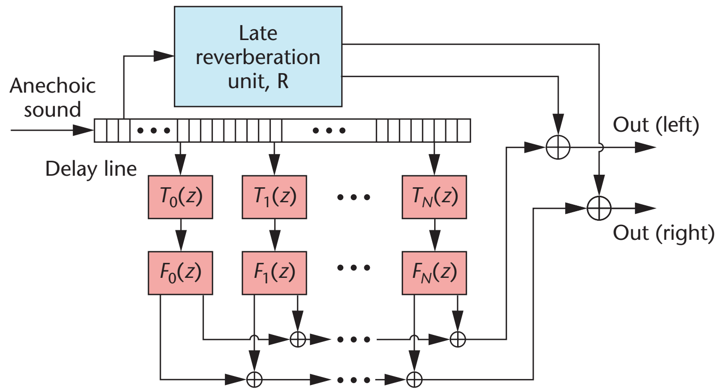
The MPEG-4 standard[1] enhances modeling of 3D sound scenes. Unlike its predecessors, the MPEG-4 standard defines the coding and specifies the presentation of multimedia objects to the end user. In this context, the standard includes a set of controls enabling interactive and dynamic audio–visual 3D rendering. These specifications are in the Systems part of MPEG-4, called BIFS (Binary Format for Scenes).

BIFS is an extension of VRML, with various nodes added to make the specification more compatible with the goals of MPEG-4 (such as playing streamed audio and visual content in virtual scenes and presenting advanced 2D content). AudioBIFS is a set of BIFS nodes for mixing and preprocessing sound streams before their playback at the MPEG-4 decoder.[2] The Sound node (familiar from VRML) adds sound sources in 3D scenes in defined positions and with simple directivity patterns. For modeling sound propagation in 3D spaces, the MPEG-4 standard adds a new set of nodes (called Advanced AudioBIFS) that describe the effect of the room, air absorption, obstruction of sound caused by walls, and the Doppler effect, in a manner described in this article.

MPEG-4 adopts both physical and perceptual approaches to 3D sound environment modeling. By physical approach, we mean the room acoustic modeling we present in this article. Acoustic properties are associated with the geometry objects. The perceptual approach, on the other hand, relies on a set of parameters that defines the perceived quality of a room acoustic effect. In this approach, these qualitative parameters characterize a room impulse response that is rendered at an MPEG-4 decoder. The parameters are associated with each sound source separately, and they don't depend on other objects in the scene.

References

1. *ISO/IEC 14496-1:1999, Coding of Audiovisual Objects*, Int'l Organization for Standardization, Geneva, Dec. 1999.
2. E.D. Scheirer, R. Väänänen, and J. Huopaniemi, "AudioBIFS: Describing Audio Scenes with the MPEG-4 Multimedia Standard," *IEEE Trans. Multimedia*, vol. 1, no. 3, Sept. 1999, pp. 237-250.



3 The DIVA auralization signal processing structure.

- distance from the listener,
- identification of all the materials involved in the reflection path, and
- incoming direction of the sound (azimuth and elevation angle in relation to the listener).

We calculate the parameters of late reverberation offline, as Figure 2 illustrates. This lets us tune the reverberation time and other reverberation features according to the modeled space.

In our hybrid model, the actual impulse response isn't constructed for the convolution (auralization) process. Instead, we use a signal-processing structure that modifies the input signal as it would be modified in convolution. With the image-source method, we calculate these parameters, with which we define the signal-processing parameters for the auralization module.

Auralization and signal processing

Image-source calculation provides the auralization parameters that are finally converted to signal-processing parameters. We use this two-level process because auralization parameters don't need to be updated for every audio sample. However, we must recalculate the signal-processing parameters on a sample by sample basis. For efficiency, we look them up from precalculated tables or create them by interpolating the auralization parameters. The practical update rate (for example, 30 Hz) of auralization parameters depends on the available computational power and maximum tolerable latency.

We implement the final auralization process as a signal-processing structure (see Figure 3). The applied parameters are filter coefficients or the lengths of delays. Most of them are calculated beforehand and stored to simple data structures that are easily accessed with the auralization parameters. For example, we calculate material absorption as follows. First, the absorption coefficients in octave bands are translated to reflection coefficients. Then, a digital filter, which covers the whole audible frequency range, is fitted to this data and used in signal processing. For each material and material combination, one filter is calculated offline. In real-time processing, the only information needed per one image source is the material's identification, and the correct filter is picked from a table.

The signal-processing structure in Figure 3 contains a long delay line, which we feed with anechoic sound to process it. The image source's distance from the listener defines the pick-up point to the filter blocks $T_{0\ldots N}(z)$ where N is the image source's ID ($N = 0$ corresponds to direct sound). Blocks $T_{0\ldots N}(z)$ modify sound signal with the sound source directivity filters, distance dependent gains, air absorption filters, and material filters (not for direct sound). The sound's incoming direction is defined with blocks $F_{0\ldots N}(z)$ containing directional filtering or panning depending on the reproduction method. The superimposed outputs of the filters $F_{0\ldots N}(z)$ are finally summed with the outputs of the late reverberation unit R, which is a complex recursive algorithm. Note that the filter blocks $F_{0\ldots N}(z)$ produce binaural output. For other reproduction methods, we need different filters.

The whole signal-processing structure contains hundreds of filter coefficients and delay line lengths, so a detailed description isn't possible in this article. Find more information in a previous article[2] in which we also discuss signal-processing issues of dynamic auralization.

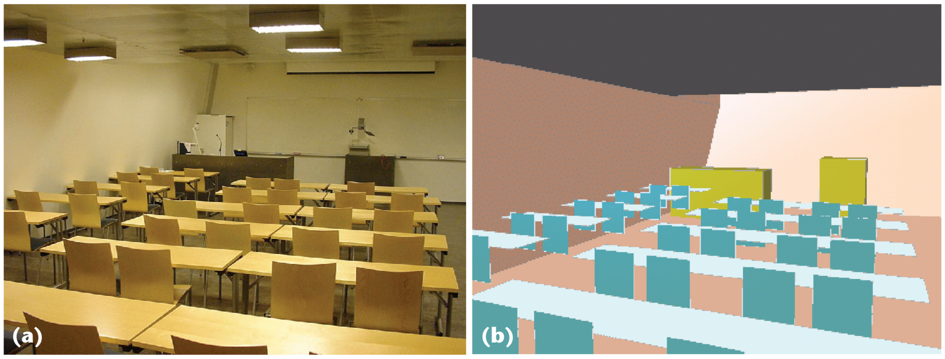
Assessing auralization quality

Acousticians traditionally judge the objective quality of room acoustics with criteria calculated from impulse responses. Almost all these attributes—for example, reverberation time and clarity—are based on energy decay times at certain frequency bands. Although these criteria aren't very accurate, we can use them to predict the room acoustics and design of concert halls.

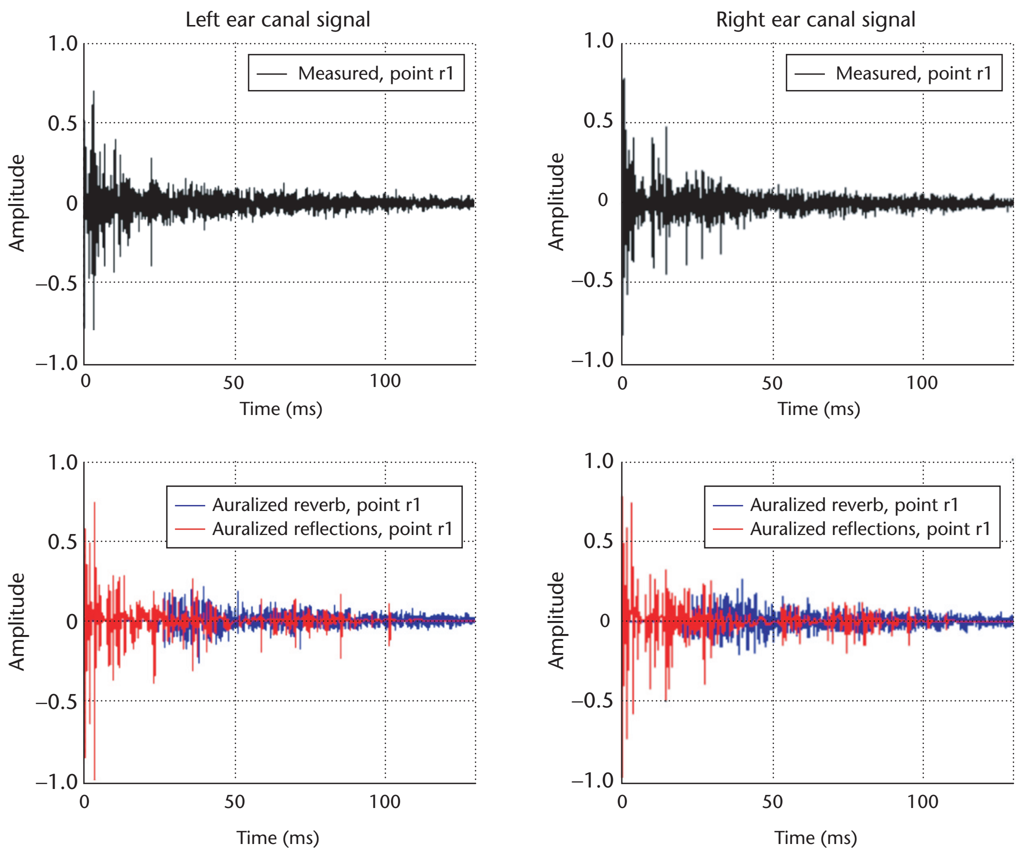
The more relevant evaluation of auralization is based on subjective judgments. Ultimately, the only thing that matters in auralization is that the produced output sounds are plausible for the application under study. Acoustics is also a matter of taste and thus an overall optimum quality might be difficult to define. In dynamic situations, the quality measurement is even harder to define because generally accepted objective criteria don't exist. However, we can obtain subjective quality judgments with listening tests.

For example, we evaluated the quality of our DIVA auralization system by comparing measured and rendered responses and sounds. We chose an ordinary lecture room (dimensions 12 m × 7.3 m × 2.5 m, see Figure 4), for which we performed both objective and subjective evaluation. Figure 5 depicts an example binaural impulse response pair. The early parts of the responses slightly differ, but they are close to each other. To find out the audible differences, we also conducted listening tests.

The results were surprisingly good, and we noticed no perceived differences with sustained tonal sound signals. However, transient-like signals, such as snare drum hits, were slightly different.



4 **(a) The photograph and (b) 3D model of the lecture room studied in evaluation of the DIVA auralization system. Note that for acoustic rendering a much simpler model is sufficient than for photorealistic visual rendering.**

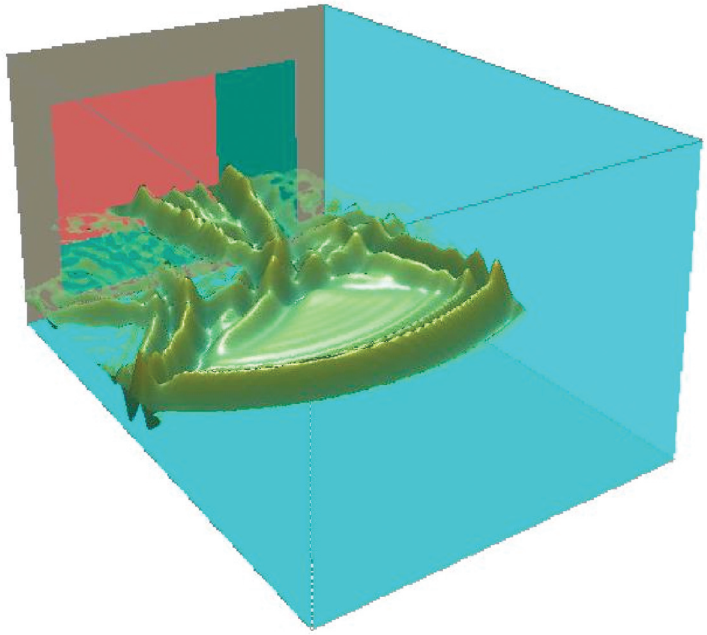


5 **The measured (top) and rendered (bottom) binaural impulse responses of the lecture room in Figure 4.**

Future trends in sound rendering

Although current auralization systems produce natural sounding spatial audio and are quite plausible, these systems apply simplified modeling methods. One major defect in current real-time physics-based auralization systems is the lack of diffraction. Recently, a few articles have reported augmenting ray-based methods with diffraction. For example, Tsingos et al.[11] presented one solution based on Uniform Theory of Diffraction that's well suited for complex geometries. Svensson et al.[12] presented another method based on the exact Biot–Tolstoy solution. However, these diffraction models aren't yet applicable in real time.

Wave-based methods aim at solving the wave equation in a space. This approach seems promising because diffraction and diffusion are automatically included in the simulation. From the auralization point of view, the most promising wave-based method is the 3D digital waveguide mesh.[13] Mathematically, it's a finite difference method, and the computation is done in the time domain. The computational load grows in $O(f^4)$ with frequency (3D mesh plus sampling frequency), and therefore, with current computational capacity, the waveguide mesh simulations can be done only at low frequencies. As an example, we've simulated a large hall with a stage having a total volume of approximately 13,000 cubic meters up to



6 **A visualization of a 3D waveguide mesh simulation in a simplified concert hall model.**

500 Hz. Figure 6 illustrates the sound field in the hall at 63 ms after an excitation impulse. The surface presents the sound pressure at a height of 1.5 m above the floor while the excitation was on the stage.

There has also been much progress in the room acoustic modeling and auralization during the past decade. Commercial room acoustic modeling systems exist, and computers are in everyday use in the design of acoustically challenging spaces such as concert halls and music studios. However, a real challenge is modeling dynamic environments in which we can interactively change everything, including the geometry and materials. Such a tool could be practical for architects and other designers. Although accurate sound rendering methods exist already, they are still computationally too laborious. As computers get faster and algorithms more efficient, we gradually approach the ultimate goal of real-time simulation of sound and light behavior. ■

Acknowledgments

This work has been partly financed by the Academy of Finland through the Helsinki Graduate School in Computer Science, the Pythagoras Graduate School, and project funding.

References

1. T. Takala and J. Hahn, "Sound Rendering," *Computer Graphics* (Proc. Siggraph 92), vol. 26, no. 2, ACM Press, New York, 1992, pp. 211-220.
2. L. Savioja et al., "Creating Interactive Virtual Acoustic Environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, Sept. 1999, pp. 675-705.
3. M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization: An Overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, Nov. 1993, pp. 861-875.
4. J.-M. Jot, "Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human–Computer Interfaces," *Multimedia Systems*, vol. 7, no. 1, Jan./Feb. 1999, pp. 55-69.
5. A. Appel, "Some Techniques for Shading Machine Renderings of Solids," *AFIPS 1968 Spring Joint Computer Conf.*, Thompson Books, Washington, D.C., vol. 32, 1968, pp. 37-45.
6. A. Krokstad, S. Strom, and S. Sorsdal, "Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique," *J. Sound and Vibration*, vol. 8, no. 1, 1968, pp. 118-125.
7. J.B. Allen and D.A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoustical Soc. of America*, vol. 65, no. 4, Apr. 1979, pp. 943-950.
8. V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, Jun. 1997, pp. 456-466.
9. M. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1-2, Jan./Feb. 1973, pp. 2-10.
10. J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd ed., MIT Press, Cambridge, Mass., 1997.
11. N. Tsingos et al., "Modeling Acoustics in Virtual Environments Using the Uniform Theory of Diffraction," *Computer Graphics* (Proc. Siggraph 2001), ACM Press, New York, 2001, pp. 545-552.
12. U.P. Svensson, R.I. Fred, and J. Vanderkooy, "Analytic Secondary Source Model of Edge Diffraction Impulse Responses," *J. Acoustical Soc. of America*, vol. 106, no. 5, Nov. 1999, pp. 2331-2344.
13. L. Savioja and V. Välimäki, "Interpolated 3-D Digital Waveguide Mesh with Frequency Warping," *Proc. 2001 IEEE Int'l Conf. Acoustics, Speech, and Signal Processing* (ICASSP 01), IEEE Press, Piscataway, N.J., vol. 5, 2001, pp. 3345-3348.

Tapio Lokki *is a PhD student in the Laboratory of Telecommunications Software and Multimedia at the Helsinki University of Technology, Finland. He has a MSc in electrical engineering from the Helsinki University of Technology. His research interests include room acoustic modeling, auralization, and virtual reality.*

Lauri Savioja *is an acting professor in the Laboratory of Telecommunications Software and Multimedia at the Helsinki University of Technology, Finland. He has a PhD in computer science from the Helsinki University of Technology. His research interests include virtual reality, room acoustics, and human–computer interaction.*

Jyri Huopaniemi *is a research manager at Nokia Research Center's Speech and Audio Systems Laboratory in Helsinki, Finland. He has a PhD in electrical engineering from the Helsinki University of Technology. His professional interests include 3D sound and interactive audio, digital audio signal processing, virtual audio–visual environments, middleware, and multimedia.*

Riitta Väänänen *is a PhD student in the Laboratory of Acoustics and Audio Signal Processing at the Helsinki University of Technology. She has a MSc in electrical engineering from the Helsinki University of Technology, Finland. Her research interests include room reverberation modeling and parametrization of 3D sound in virtual reality systems. She participated in the MPEG-4 standardization process from 1998 to 2001 and spent 2001 at IRCAM (Institute for Research and Coordination of Acoustic and Music), Paris, as a visiting researcher.*

Tapio Takala *is a professor of interactive digital media in the Laboratory of Telecommunications Software and Multimedia at the Helsinki University of Technology, Finland. He is currently on a leave of absence, working for interaction development at Nokia Ventures Organization. He has a PhD in computer science from the Helsinki University of Technology. His interests range from modeling and design theory to interactive multimedia performance, including artificial life simulations. His research group specializes in sound for VR.*

Readers may contact Tapio Lokki at the Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, PO Box 5400, FIN-02015 HUT, Finland, email Tapio.Lokki@hut.fi.

For further information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.