ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 94 (2008) 1000 – 1015

Technical and Applied Papers
DOI 10.3813/AAA.918116

Framework for Real-Time Auralization in Architectural Acoustics

Markus Noisternig, Brian F. G. Katz
LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France. Markus.Noisternig@ircam.fr

Samuel Siltanen, Lauri Savioja
Department of Media Technology, Helsinki University of Technology, P.O. Box 5400, FIN-02015 TKK, Finland

Summary
Auralization is the process of making audible the acoustics of complex virtual architectural spaces in a realistic and accurate manner. This paper presents a novel real-time auralization software environment comprising a room acoustic modeler, a spatial renderer for auralization, and a visualization and scene graph unit for interactivity. The computation of early reflection paths within the geometric model is based on an efficient beam tracing algorithm capable of real-time detection of specular reflection paths in a static geometry with one or several moving listener(s). For "simple" rooms, the real-time performance is maintained even with dynamic geometries and sources. Results of the room acoustic model consisting of visible reflection paths and their accumulated material attenuation are sent to the audio renderer. From this geometrical and acoustical data, listener position-relative 3D room impulse responses are generated using a higher-order Ambisonics approach. Finally, the spatialized audio output is presented to the listener via multiple loudspeakers or binaurally rendered over headphones. As higher order reflections are more diffuse in nature, they are encoded using lower Ambisonic orders, thereby reducing computational load. The environment combines high quality audio with visual rendering realized using the open source platforms Pure Data and VirChor respectively. This open source auralization framework provides direct audio-visual interaction in real-time and is suitable for VR environments.
PACS no. 43.55.Ka, 43.55.Br, 43.20.Dk

1. Introduction

Auralization has become a useful tool for the acoustic design of three-dimensional architectural environments. In Kleiner *et al.* [1] auralization is defined as *the process of rendering audible, by physical or mathematical modeling, the sound field of a source in space [...] at a given position in the modeled space*. Begault *et al.* [2] categorize spatial audio systems into three levels of immersion, in which the highest level includes spatial encoding of sound reflections arriving at the listener from different directions and simulation of late reverberation. In order to simulate the different directions of arrival Meyer *et al.* [3] first suggested the use of a distributed loudspeaker system. Hence each loudspeaker represented a given reflection direction. Room reverberation is responsible for temporal and spectral smearing of the signal and is regarded as one of many cues, which affect sound source localization, distance perception, room dimension estimation, and orientation in a given space by a listener. In performing arts, the acoustic of the room is an essential component of the performance and listening experience. Convincing auralization in multimodal interactive virtual environment systems therefore requires accurate and computationally efficient (geometric) room acoustic modeling in real-time.

In general, auralization systems consist of two separable processing units. First, the room acoustic modeling algorithm computes the propagation paths of the sound waves traveling from source position to receiver position. In enclosed acoustic environments the acoustic waves are reflected multiple times before they arrive at the listener's position. Due to this multi-path propagation, the cornerstone of the acoustic model is to find all the relevant specular early reflections for a given listener position. This requires accurate and computationally efficient search algorithms in order to obtain the results in a reasonable time frame. Each specular reflection path is characterized by its direction of arrival (DOA) relative to the listener, propagation delay, absorption due to frequency-dependent material properties, and air absorption.

Within this article, the term *specular* refers to geometrical ray optics and defines a perfect mirror-like reflection.

In the following step, the information computed in the room acoustic modeling unit, *i.e.* the early reflection paths, is transmitted to the spatial audio signal processing unit to transform the reflection history data to a room impulse response which can be auralized. For auralization of a given source position, a dry sound signal, *i.e.* a sound source synthesized or recorded in an anechoic room, is delayed,

Received 4 March 2008, revised 6 August 2008,
accepted 2 September 2008.

filtered, and spatially positioned according to the room acoustic model parameters as described above. While it is straightforward to calculate the specular reflections, for interactive applications the late reverberation part of the room impulse response is most often estimated due to calculation time constraints. Finally, the spatialized audio information is rendered to the listener via multiple loudspeakers or binaurally over headphones after method appropriate processing.

The main contribution of this paper is a software environment for auralization of complex geometries in real-time comprising novel algorithms for accelerated beam tracing. The system is built on several separately developed software components, some of which have been partially presented earlier [4, 5, 6]. This paper provides a discussion on the interaction of the different processing units with the aim to create a complete processing environment for real-time auralization for use in interactive multimodal virtual environments. In addition to the two main processing modules generally used in auralization systems, the presented environment is extended through the inclusion of a graphical front-end capable of interfacing with different devices for user interaction. The room acoustic modeling is performed using a beam tracing technique, and the audio signal processing for 3D sound field reproduction is realized using a mixed order implementation of higher order Ambisonics, which is applicable to both loudspeaker and headphone reproduction. All of the programs presented in this paper have been made publicly available as open source and can be found through the following link: `http://auralization.tkk.fi`

The paper is organized as follows. Section 2 presents a review of previous work in room acoustic modeling and auralization. Section 3 contains a general description of the proposed auralization environment and the interaction of the different algorithmic units. It further presents a brief description of the visualization environment, details of the room acoustic modeling algorithm, and the spatial rendering unit. Section 4 presents performance evaluations and conclusions. A short description of the data transfer protocols is given in Appendix A1. Appendix A2 presents a brief theoretical overview on higher order Ambisonics.

2. Background

2.1. Room acoustic modeling

There are several techniques available to computationally model the room acoustics of a given geometry. A good overview has been presented by Svensson and Kristiansen [7]. A general classification of methods can be made between wave-based and ray-based techniques. The wave-based approach aims at numerically solving the wave equation, and is computationally intensive with the computational complexity increasing proportionally with the volume of the model and the upper resolution frequency. Therefore, this method is not suitable to many room acoustics problems, such as large spaces or real-time auralization covering the entire audible frequency range.

In ray-based methods, sound propagation and other behavior is approximated by geometric rays. Although this assumption is valid only at higher frequencies, it is typically used to improve calculation speed, and is very often used in real-time auralization. The effects caused by the wave-nature of sound, such as edge diffraction, must be incorporated into the models separately. There are basically two different families of ray-based techniques: ray tracing [8] and the image-source method [9, 10]. Both of these can be used to find the specular reflection paths in a given geometry. Ray tracing uses a Monte Carlo sampling to find paths such that the result will become more and more accurate with an increasing number of rays. Stochastic ray tracing methods can also model sound scattering in order to compute the diffuse sound field. However, this method is computationally intensive and therefore not applicable to real-time auralization with current computer hardware. The image source method attempts to find all specular reflection paths by recursively mirroring the sound source against all reflective surfaces facing towards the sound source being reflected.

The exploitation of spatial data structures aides in the efficiency of finding visible reflection paths. Among these data structures, binary-space partitioning (BSP) techniques have been often used [11, 12, 13]. The original image source method is quite inefficient with higher reflection orders due to the exponential growth of the number of image sources. In order to avoid this problem a number of more advanced techniques have been proposed to find the specular reflection paths. Among these methods and their variation are pyramid tracing [14], beam tracing [15, 16, 5], and frustum tracing [17].

The basis of beam tracing methods is the construction of a visibility structure called a beam-tree. The root of the tree is the sound source and the second level consists of the beams defined by the sound source position and the visible polygons in the model. These beams are further reflected on the plane of their defining polygons to produce the next level in the beam-tree, corresponding to first order reflections, see Figure 1 for details. Only the polygons inside the reflected beams need to be considered when constructing the following level of beams for each tree node. This process is repeated until the desired reflection order limit is reached. This beam-tree can be accessed efficiently at runtime, tracing it backwards from the listener position. Valid reflection paths can be constructed without performing extra computations on paths which are obstructed or invalid. Depending on the exact implementation of the algorithm, the beam-tree can be accurate [15, 16] or an approximation [5]. In the latter case the paths obtained using the tree must be validated, but the overall structure is typically smaller in size.

The goal of this work is to find and auralize all the specular reflection paths up to as high a reflection order as desired. Emphasis is placed on the early reflections, and for this reason the image-source approach is the most suitable. To optimize the use of computational resources a beam tracing technique was chosen such that the determination

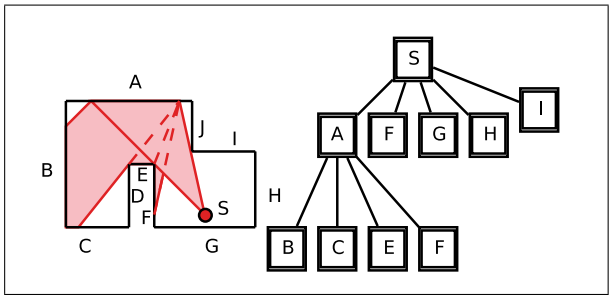

Figure 1. (left) A simple room with polygons A-J. The beam from source S to polygon A is shown. The following level of beams reflected from polygon A to polygons B, C, E, and F are also shown. (right) The corresponding beam-tree. For clarity, beams involving polygons F, G, H, and I are not shown. There is no inherent limit to the depth of the beam-tree.

of early reflection paths is as efficient as possible. However, since it is not possible to handle all the reflections separately to an infinite order, a reverberation algorithm has been implemented to estimate the late reverberation of the room impulse response.

2.2. 3D Sound reproduction

Spatial sound reproduction has been intensively studied and a variety of multi-channel spatial audio systems have been developed, *e.g.* quadrophony, vector base amplitude panning (VBAP) [18, 19], wave field synthesis (WFS) [20, 21, 22, 23, 24], and Ambisonics (described in more detail below). The proposed auralization environment benefits from certain properties available in Ambisonics to directionally encode the early reflections and to reconstruct the sound field independent of the listener's head orientation.

Ambisonics was originally introduced by Cooper and Shiga [25], Gibson *et al.* [26], and Gerzon [27] in order to recreate a 3D sound field at one central location of a spherical loudspeaker array. It is principally based on the spherical wave decomposition of the local sound field (cf. Appendix A2). The term Ambisonics primarily refers to systems that reproduce the spherical harmonics up to first order. The term higher order Ambisonics (HOA) is used for systems exceeding first order spherical harmonics. A comprehensive theoretical analysis of HOA is given in Daniel [28] and Poletti [29].

An ideal wave expansion in spherical coordinates consists of an infinite number of spherical harmonics [30, 31, 32]. The practical limitation of truncating the spherical harmonics at a given order results in limitation on the area of accurate sound field reproduction and the spatial resolution of the reconstructed sound field. Studies have shown that Ambisonics is asymptotically holophonic [33, 34] and increasingly accurate with increasing order [35]. Perceptual studies presented in Bertet *et al.* [36] show the advantage of using 4^{th} order Ambisonics compared to lower order systems.

Further improvements have been proposed by Daniel *et al.* [37] in their reformulation of HOA in order to remove the limitations implicit to the plane wave assumption. This results in the implementation of near-field compensation filters. Using high-order encoding the listener can receive distance cues from near-field sources equal to those that would be received from the real sound field, as the sound field around the listener is reproduced arbitrarily well. Near-field compensated higher order Ambisonics (NFC-HOA) also improves the reproduction quality for off-center listening considerably and allows one to reproduce spherical waves for proper distance perception. To guarantee stable compensation filters the encoder must be modified and is no longer generic, which is the main drawback of this approach with respect to the auralization framework proposed in this paper. In addition, in room acoustic modeling the assumption can be made that reflections are mainly coming from the far field. However, in generalized virtual environments with a predefined decoder configuration NFC-HOA could be used to obtain a more accurate rendering of the direct sound from nearby sources over loudspeakers.

The acoustic theory for multichannel sound reproduction systems usually assumes free-field conditions, where acoustics of the listening room are not considered. Spors *et al.* [38] derived a method for active listening room compensation for multi-loudspeaker sound reproduction systems. Corteel [39] proposed the use of MIMO optimization techniques to compensate for the listening room acoustics when using loudspeaker arrays.

Binaural signals for headphone listening can be easily derived by convolving each reflection with the appropriate head related transfer function (HRTF) according to their position in space [40, 41, 42, 4, 43, 44] thereby eliminating any effect of the reproduction room. Binaural simulation of loudspeaker rendering systems is also possible by rendering the loudspeaker feed signals as virtual sources in the same manner. Head-tracking can be used to further improve binaural rendering. These approaches can also be mixed, depending on reflection order and desired spatial precision. Transaural rendering of the binaural signal is also a possibility, reproducing the binaural audio over two loudspeakers, though interactivity is reduced as treating head rotation is difficult.

2.3. Auralization systems

There are several commercially available auralization systems that have been developed as part of room acoustic modeling packages. Examples of these systems are (in alphabetical order) CATT-Acoustic [45], EARS [46], ODEON [47], and RAMSETE [14]. Most of these packages are based on hybrid models using both the image-source method and ray tracing. Perceptual room models have also been developed at IRCAM [48].

Non-commercial packages also exist: The RAVEN software [49] from RWTH Aachen University has similar goals to that presented here and in the most recent version it is even possible to handle modifiable geometries [50]. The IKA-SIM software [51] from Ruhr-Universität Bochum is designed for psychoacoustic research with teleconferencing and networked virtual environments being

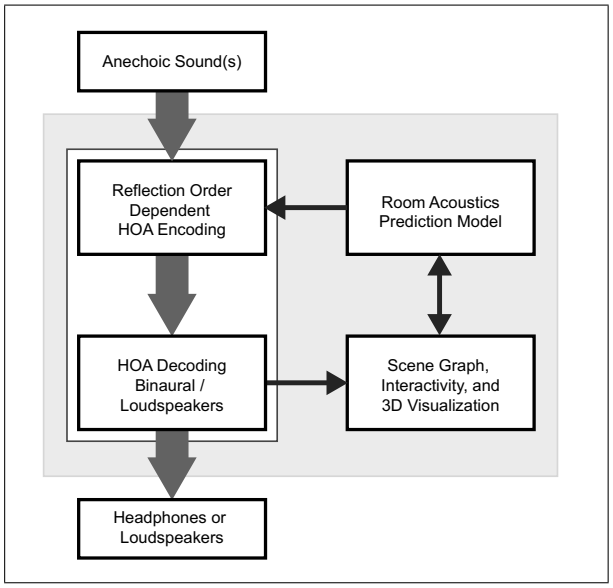

Figure 2. Functional units of the auralization environment.

the current major application areas [52]. The main focus areas of the system REVES [53] developed at INRIA are the efficient computation of scattering and perceptually optimized rendering of a large number of simultaneous sound sources and reflections. This is accomplished through preprocessing analysis of audio signals and the application of perceptual filtering and clustering in real-time. The AURALIAS software [54] from the University of Liege, and the DIVA system [11] from TKK, Finland have been designed for room acoustic modeling and auralization purposes.

However, all of the abovementioned software packages are closed source. The only open source software for room acoustic auralization, that the authors are aware of, is the result of the UniVerse project [13]. Although some of the developers of the UniVerse system also contributed to the room acoustic modeling unit of this project, EVERTims, these two software packages are separate and distinct. The main advantage of EVERTims over UniVerse is the use of more advanced algorithms both for beam tracing and for 3D sound reproduction.

One somewhat related system is the Sound Lab (SLAB) developed at NASA [55]. It is published as open source and it provides 3D spatialization but it does not include room acoustic modeling.

3. Description of the environment

The proposed auralization environment is divided into four main functional units: (1) geometric scene graph and 3D visualization, (2) room acoustic modeling, (3) auralization and spatial audio encoding, and (4) spatial audio decoding (cf. Figure 2).

In order to support the interchangeability of algorithms the main functional units are fully encapsulated and self-contained. Consequently, the interaction between the building blocks is completely defined by the communication protocol, allowing substitution of any element by a suitable replacement unit, which adheres to the defined protocol.

3.1. Geometric scene graph and 3D visualization

The scene graph unit is based on Virtual Choreographer (VirChor) [56], a real-time 3D graphics rendering engine published under GNU General Public License (GPL). The system handles the geometric room model definition as well as the positions of source(s) and listener(s). Graphic rendering provides interactive visualization of the room and propagation paths. In addition, VirChor incorporates a powerful scripting language allowing behavior interactions to be defined for any object.

During the first preprocessing phase the scene graph unit reads the three-dimensional geometry model defining the room's polygon structure. This is then sent to the room acoustic modeling unit. Each polygon has an associated acoustic material, in addition to its graphic properties. Each material is defined by its octave band absorption coefficients (from 32 Hz to 16 kHz) and a scattering coefficient (currently frequency independent). In the current version of the software the scattering coefficient is not applied; but it is defined in the protocol for future releases.

For the visualization of reflection paths, the room acoustics unit sends line segment coordinates. Continuous updates are transmitted as positional information varies, new paths become visible, or existing paths are no longer visible.

The listener's position and orientation can be modified interactively using for example a joystick or mouse. These control devices can be easily replaced by any human-computer interface (HCI) or client application, *e.g.* an optical body motion capture system using a simple messaging protocol. Direct interaction of the listener with the virtual environment improves immersion as compared to auralizations using pre-defined or pre-rendered walkthroughs.

3.2. Room acoustic modeling (EVERTims)

The goal of the room acoustic modeling unit is to determine all early specular reflection paths in a given geometry in real-time. The system is based on EVERT, an efficient beam tracing library [5]. It is optimized for static geometries and source positions. Under those conditions it can achieve an interactive update rate of 15 Hz or better for a moving listener while calculating up to seventh order reflections with a model containing less than one-thousand polygons.

3.2.1. The EVERT beam tracing library

The EVERT library uses an algorithm in which the constructed beam-tree is an approximation. While constructing the tree, accurate visibility information is used so that if a polygon is only partially inside a beam, it is cut against it and the part inside the beam is used when constructing the next levels of the tree. However, only the original polygon identifiers are stored in the final beam-tree structure.

Occlusion by other polygons is omitted. Thus, the structure is simple, because the beams are not split very often, but the resulting paths obtained by using the structure must still be validated.

The validation of the paths can be done efficiently by using two optimization techniques, which utilize the coherence in the paths validation. The first technique, *fail-planes*, aims to return a negative result early in the calculation for invalid paths. At the implementation level, the path validation can be considered as testing intersections of line segments against planes defining the beams. When an invalid result is obtained, the plane against which the test failed can be propagated up the tree by reflecting it against the planes of the polygons in the path, so that the tests for the following paths can be started by first testing against the propagated plane, which is likely to produce a negative result early.

The second optimization technique, *skip-spheres*, is based on grouping a small number of neighboring BSP nodes into "buckets". If all the paths in the bucket produce a negative result, one can compute a sphere defined by the smallest distance to the corresponding fail planes and the current listener position. Hence, while the listener's position remains inside a given sphere, the entire bucket of paths for that sphere remains invalid. Thus, the majority of intersection tests can be eliminated. For those intersection tests which are not eliminated, a kd-tree-based ray tracer is used to efficiently compute the results.

3.2.2. The EVERTims real-time beam tracing software

The EVERT library provides a beam-tree for a given geometry and source configuration up to a requested reflection order. The implemented software uses an iterative refinement procedure. The behavior is parameterized such that there is a minimum required order for a solution. Whenever there is a change in the geometry or a sound source position, an approximate beam-tree is constructed up to that order. The paths in the tree are validated, with the visible paths being sent to the auralization unit. The system then continues to compute the solution up to the next reflection order until the chosen maximum order is reached. For each source-receiver pair a separate beam-tree is generated.

When a listener moves, all the paths in the beam-trees for that listener are tested for validity and the changes are reported to the auralization unit. Changes in *orientation* of either the listener or source do not affect the tree or the visibility tests and therefore there is no need to perform any recalculation. Computationally, the most challenging operations are the movements of a source or changes in the geometry. A change to the geometry or the source position requires a new computation of the underlying BSP tree, reconstructing the beam-tree up to the minimum order, and beginning the iterative order refinement of the solution once again.

The implemented software is multi-threaded. There is one thread for the processing input, one for updating visibility changes, and one for calculating new solutions. This enables fast visibility updates while the listener is moving even in the case when the other thread is computing a new solution to higher reflection order or due to a change in the geometry or in the source position. The input-thread monitors incoming messages and asks for services of the other threads whenever there is a change that requires any recomputation.

An example of the real-time evolution of the room impulse response (RIR) is shown in Figure 3, illustrating the iterative refinement of the energy-time response as a function of reflection order (simple auditorium room model Figure 6c). The room modeler provides increasing reflection information over time in order to balance the need for real-time responsiveness and detailed results. The first panel (upper left corner) shows all first and second order reflections, with each subsequent panel showing the addition of higher order solutions, progressing finally to the last panel showing all specular reflections up to 9th order.

Both the room acoustic modeling software EVERTims and the EVERT library have been implemented in C++.

3.3. Spatial rendering and auralization

The room acoustics prediction model as described previously in Section 3.2 calculates the specular reflection paths for each sound source relative to the current listener's position. As the reflection paths can only be computed up to a limited order in real-time, a statistical model is used to approximate the late reverberation. The late reverberation is then added to represent the tail of the room impulse response. A generalized rendering approach is used to allow auralization of the calculated reflections using either loudspeakers or headphones [4].

Several methods for encoding the room impulse response or sound field exist. In order to provide reasonably correct angles of incidence toward the listener for the various reflections both direct encoding (*e.g.* VBAP, direct convolution with HRIRs) and sound field reproduction techniques (*e.g.* Ambisonics, WFS) can be used. A comprehensive comparison of the sound field reproduction techniques is given in Daniel [37].

In binaural rendering systems the listener is presented with signals derived and presented directly to each ear over headphones. A straightforward approach for creating binaural signals would model binaural room impulse responses (BRIRs) and convolve them with the *dry* sound source signals. An alternate approach would be to consider each reflection as an independent source (true image source implementation), which must be spatially processed individually. For binaural synthesis, each reflection would require a separate convolution calculation. For a large number of sources and time varying environments the direct binaural approach is not practical.

The main advantages of using HOA in auralization systems are the efficiency for encoding the room impulse response, the decoupling of encoder and decoder (allowing for user defined restitution configurations), computationally efficient rotation of the encoded sound field, the simple applicability for 3D sound field reproduction, the scal-

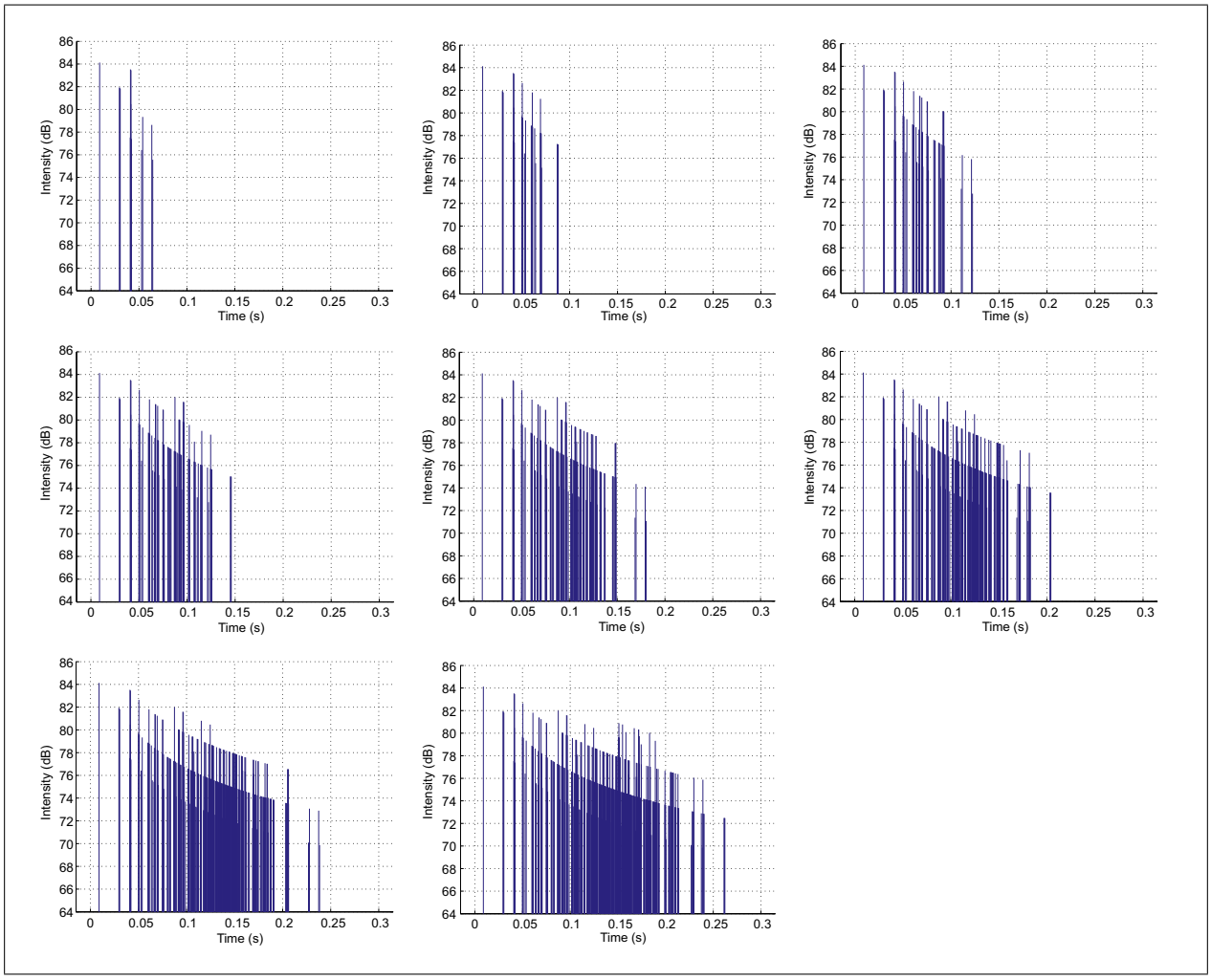

Figure 3. Image source model energy-time responses as a function of increasing reflection order from the room acoustic model. The first panel shows all first and second order reflections, the maximum order is increased by one in each panel such that the final panel shows specular reflections up to 9th order.

ability in terms of spatial resolution as a function of Ambisonic order, and highly similar reconstruction accuracy at higher orders when compared to WFS. In the following sections the application of Ambisonics to auralization is described.

As scattering is not taken yet into account the energy of each early reflection is somewhat overestimated. One solution would be to remove the energy associated with the scattered rays yielding a more accurate level of the early reflection. One could then either ignore this removed energy (discarding diffuse energy) or one could adjust the energy in the FDN late reverberation to account for this additional diffuse energy.

The spatial rendering and auralization unit has been implemented in the open source software Pure Data including external objects implemented in C.

3.3.1. Higher order Ambisonics theory

It can be shown [30, 31, 32] that any arbitrary function that is square-integrable over the unit sphere can be expanded into spherical harmonics coefficients by applying the spherical harmonics transform. As detailed in Gumerov *et al.* [32] and Williams [31], the eigen-solutions of the wave equation in spherical coordinates are given by sets of spherical harmonics $Y^{\sigma}_{mn}(\theta, \phi)$ of order n and degree m and spherical Bessel functions of the first and second kind, or their corresponding spherical Hankel functions. The mutually orthogonal spherical harmonics only depend on the direction-of-arrival (θ, ϕ) of the traveling wave, *i.e.* the angular dependencies of the solutions. The radial dependencies are given by the Bessel and Hankel terms. The Fourier-Bessel expansion of an 'incoming' sound field at the position $\mathbf{r} = (r, \theta, \phi)$ defined in spherical coordinates can be written as (cf. Appendix A2)

$$p(\mathbf{r}, \omega) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{0 \le n \le m, \sigma = \pm 1} A^{\sigma}_{mn} Y^{\sigma}_{mn}(\theta, \phi), \qquad (1)$$

where $i = \sqrt{-1}$, A^{σ}_{mn} are the spherical expansion coefficients, and $j_m(kr)$ is the spherical Bessel function of the first kind subject to the wave number $k = 2\pi f/c$. The

HOA encoding and decoding equations are derived from this Fourier-Bessel expansion assuming only plane waves (far-field sources). As such the radial terms can be omitted. A more elaborate mathematical description is provided in Appendix A2.

3.3.2. Encoding the sound field

The HOA encoder reduces to a simple weighting by spherical harmonics depending on the direction of arrival, which is computationally very efficient, cf. equation A6 in Appendix A2. An increase of the encoding order results in higher spatial precision. Ambisonics allows for encoding at mixed orders without any additional requirements on the decoder.

As mentioned previously, each time the listener or a sound source moves the reflection paths are calculated in real-time. The room modeler sends a list containing the reflection path relevant data to the spatial encoder. For each reflection path the direct impulse is delayed, attenuated, and filtered, before being encoded into the Ambisonic room impulse response; each visible reflection path is encoded independently. The encoder uses higher orders for lower order reflections to obtain higher spatial precision. A simple 'voice allocation algorithm' with a given maximum channel number limit provides smooth transitions when a reflection path becomes invisible or is replaced by another path. The propagation delay is calculated from the length l_i of the respective reflection path by $\Delta t = l/c$, where c is the speed of sound. For every reflection the product of all the attenuation coefficients along the corresponding path are calculated in octave bands. The current version of the encoder does not yet take into account air absorption. In order to maintain real-time performance on the current hardware, the present implementation groups octave bands into three broad bands (32 Hz–250 Hz, 500 Hz–2 kHz, and 4 kHz–16 kHz) implemented using IIR shelving filters. Exploitation of the scattering coefficient has not been incorporated in the current implementation. Use of the scatter coefficient requires a separate treatment of the early reflections and the diffuse part of the response.

The direct sound can be processed separately from the room impulse response to obtain the highest localization accuracy, or simply encoded at the highest order. For binaural systems the highest accuracy can be achieved by direct filtering with the HRTF, bypassing the HOA encoder. A similar approach could be used for loudspeaker restitution where a different method could be used for the direct sound and for the room response.

While the RIR is dense at higher orders, the image source response is too short to adequately represent the full room response for auralization, and therefore the late reverberation algorithm must be used to complete the response, completing the late reverberation tail of the room. The late reverberation response is computed from the early reflection data using feedback delay networks (FDN) [57]. The input signals are first decorrelated before the reverberators based on all-pass structures are applied. At every iteration the signal is filtered and attenuated in order to simulate the coloration by absorption and the decay time. In this instance the global average octave band absorption coefficients are used. This approach provides simple control of the reverberation pre-delay, the reverberation gain, the decay rate, and the frequency dependent absorption. One can assume that the late reverberation is more diffuse and therefore less spatially defined than the direct sound or early reflections. Therefore, the lowest order Ambisonic encoder is used for spatial encoding of the late reverberation. The reverberation time is estimated from the early decay time by creating reflectograms from the data coming from the room acoustic modeler. Sorting the averaged squared amplitudes in time and applying linear regression allows a rough approximation of the decay time. This estimated value controls the decay time of the FDN. The late reverberation algorithm produces four decorrelated output channels, which are then encoded into first order Ambisonics coming from different directions. Currently the actual spatial distribution of the late reverberation is not taken into account. An estimation of the spatial distribution of the late reverberation simply from the early reflection spatial distribution would only be valid in 'well tempered' rooms.

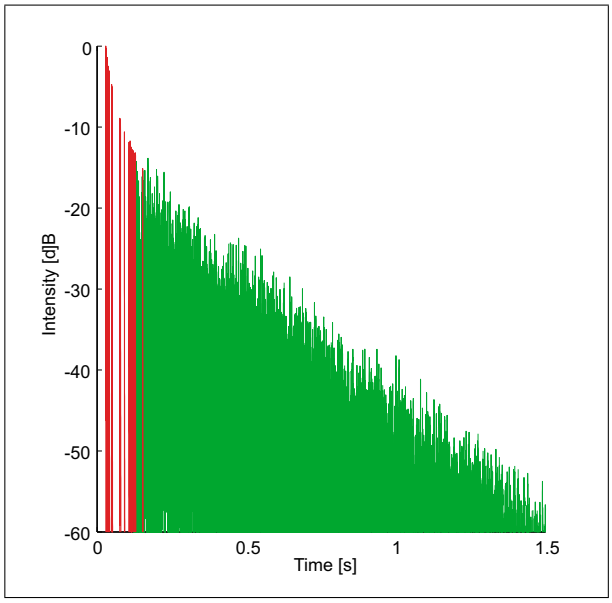

Figure 4. Room impulse response showing the early reflections and the FDN generated late reverberation.

Figure 4 shows an example of the RIR complemented by the estimated late reverberation (simple auditorium room model Figure 6c). For this example, 100 encoder channels were implemented for the early reflection paths, providing reflections up to 3rd order.

3.3.3. Decoding the sound field

Using HOA, the sound field reconstruction module depends only on the loudspeaker positions. Hence, definition of the playback configuration (*i.e.* loudspeaker numbers and positions, loudspeaker array *vs.* headphones) is only required for the decoder unit, separate from the multi-

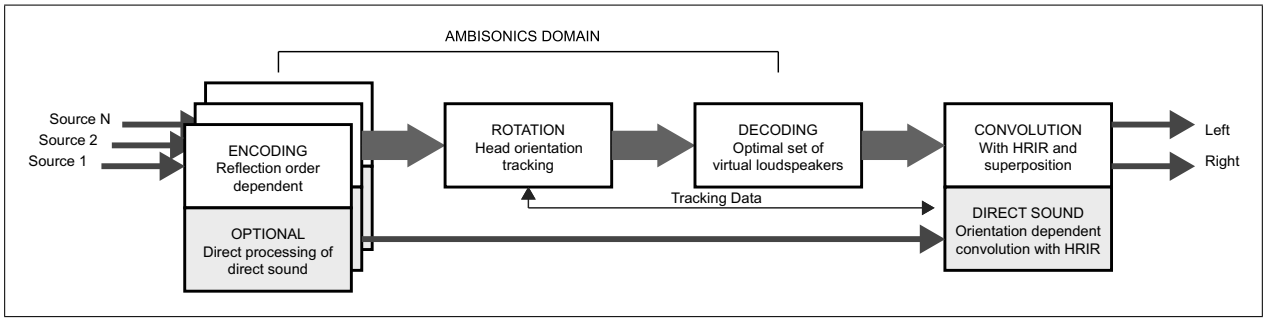

Figure 5. General structure of the binaural rendering using a virtual Ambisonics approach.

channel format implemented in the encoder. The optimal listening area for accurate sound field reproduction depends on the order, the upper frequency limit, and the decoder type (*i.e.* basic decoding, 'max r_E' decoding, 'in-phase' decoding [58], or user defined intermediary solutions).

The virtual Ambisonics approach allows for binaural rendering of HOA over headphones by filtering the decoded loudspeaker feed signals with HRTFs in order to create a binaural representation of the reconstructed sound field. As mentioned previously, the decoder is solely defined by the position of the virtual loudspeakers. Optimal decoders can be derived by uniform loudspeaker arrangements and the processing cost is fixed by the number of virtual loudspeakers rather than the number of processed reflections. Figure 5 outlines the general structure of the binaural rendering using a virtual Ambisonics approach.

A review on research related to headphone-delivered sound reproduction techniques shows the importance of adapting the rendered sound field to the listener's conscious and unconscious head-movements in order to avoid front-back confusions and thereby obtaining improved localization accuracy and presence [59, 60, 61, 62, 63, 64]. Encoded Ambisonics, prior to decoding, is easily rotated, which facilitates real-time compensation for head movement with a static set of HRTF functions. The use of non-individualized HRTFs in binaural rendering systems degrades the localization accuracy [65]. Therefore, the HRTFs are not fixed within the proposed environment and may be easily exchanged.

3.3.4. Further improvements

Recent research shows an increasing interest in directional sound sources within virtual acoustic environments [66, 44, 24, 67]. In addition to providing the direction of arrival at the listener, the communication protocol of the proposed framework also provides the first reflection point relative to the sound source. Therefore, spatial filters can be easily applied to allow for dynamic directionality of sound sources.

3.4. Inter-module communication

The four main functional units communicate via the Open Sound Control (OSC) protocol [68, 69] and UDP (User Datagram Protocol). The use of OSC allows the distribution of sub-tasks to different computers within a heterogeneous local area network, or to run the entire environment on a single computer. Not only does the clustering of computers provide more computational power, but it allows the use of different types of computers and operating systems, which might be optimized to perform their respective processing tasks.

During the initialization phase the geometric scene graph is constructed and sent to the room acoustic modeler. VirChor then monitors a predefined TCP/IP port to receive the reflection path coordinates as line segments for visualization. Each time an OSC message is received the line segments are updated. When a sound source or listener is moved the position data is sent to the room acoustic modeler, where the specular reflections are computed. All the reflection path parameters are then sent to the spatial encoder. The data structure of the protocol is described in more detail in Appendix A1.

Within the proposed environment head-tracking is captured using off-the-shelf motion sensors and orientation tracking devices. The communication protocol with VirChor allows any external software to update the source or listener positions via OSC. To guarantee minimal response times these sensors should be directly connected to the binaural rendering unit; latency and jitter are unavoidable in network based communications, even though OSC is optimized in respect thereof [68]. The current implementation enables Pure Data to act as the interface between the tracking devices and the scene graph.

4. Performance evaluation and discussion

A series of performance evaluations were carried out to examine different aspects of the proposed platform. For these evaluations, both the room acoustic modeling software (EVERTims) and the geometric scene graph modeler and 3D visualization (VirChor) have been run on a single machine equipped with an Intel Core 2 Duo 2.5 GHz processor, with 2 GB RAM and an NVIDIA GeForce 8600M GT / 512 MB VRAM graphics card. The spatial rendering and auralization unit (Pure Data) has been run on a Motorola PowerPC G4 1.67 GHz, with 1 GB RAM. The two machines were connected via a 100 MBit local area network (LAN).

Table I. Comparison of Image Source Method (ISM): Number of image sources as a function of order for a simple cube (6 polygons).

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ISM (analytic)	6	18	38	66	102	146	198	258	326	402	486	578	678	786	902
EVERTims	6	18	38	66	102	146	198	258	326	402	486	577	678	786	902
CATT-Acoustic	6	18	38	66	102	146	198	258	–	–	–	–	–	–	–

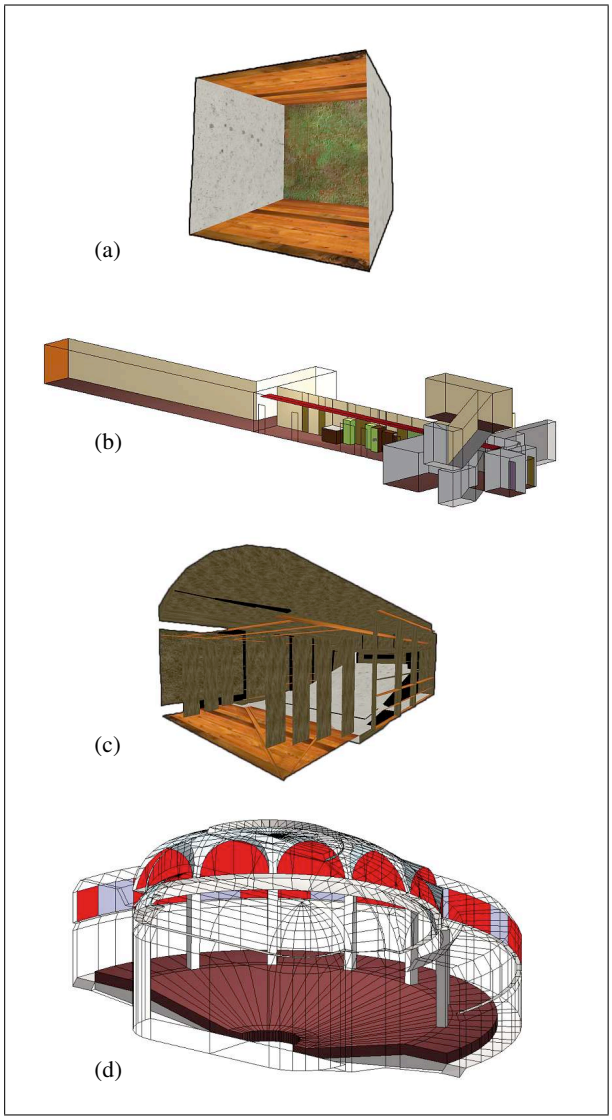

Figure 6. The architectural models used in the performance evaluation: (a) simple cube, (b) complex corridor, (c) simple auditorium, and (d) complex auditorium.

The four test models, as shown in Figure 6, range from a simple cube (6 polygons) to a complex auditorium (1626 polygons). The latter is a publicly available calibrated model of the "Fogg Art Museum Lecture Room", *i.e.* the lecture room that began Wallace Clement Sabine's work in architectural acoustics [70].

Comparison results have also been calculated for several configurations using a commercial room acoustics software package, CATT-Acoustic [71], in order to provide a point of reference. The reference results were calculated using CATT v.8.0e on an Intel Core 2 Duo, 2.13 GHz, with 2 GB RAM running WinXP. It is acknowledged that CATT-Acoustic is not optimized for real-time processing and that the timing results here are only provided as a point of reference to the reader and should in no way be considered as an evaluation on the efficiency or performance of the reference package. It is clear that, if desired, the reference package could also be optimized for speed resulting in optimized calculation times. As the reference package is used primarily for analysis, rather than real-time rendering, with the goals and data logging procedures being quite different from EVERTims, the presence of the reference results should not be considered here as a comparison of competitive software.

Table II. Comparison of Image Source Method (ISM): Number of image sources as a function of order for a complex auditorium model (1626 polygons).

Order	1	2	3	4	5	6
EVERTims	9	34	74	107	83	180
CATT-Acoustic	9	33	67	110	90	179

4.1. Image source calculation comparison

A basic validation of the room acoustic unit, EVERTims, has been performed using a simple cube model. The image source positions for a cube can be computed using the analytical model by Allen *et al.* [9], thereby providing a true solution as a reference. One has to note that with a simple cube model only a part of the algorithm is tested, as there are no obscuring surfaces. The purpose of this validation therefore is to verify the most basic functionality of image source calculations. For a simple cube (40 m edge length) and a given source-listener pair, image sources were calculated up to 15th order using the analytical solution and EVERTims. Image source positions were also calculated using the reference package up to 8th order. The number of calculated image sources as a function of order is provided in Table I. The results are nearly identical, with EVERTims only missing one 12th order reflection. An analysis of positions of the image sources shows that all results are coincident within a tolerance of 0.005°.

Following the case of a simple cube, a similar comparison was carried out for the complex auditorium model up to 6th order. For this case, no analytical solution was possible. As such, there is no *true* solution and EVERTims is only be compared to the reference package. The image source number tabulation is provided in Table II. While

the number of image sources for each order is not exactly the same, EVERTims provides results comparable to the reference.

As the number of image sources is not identical, it is not a trivial task to calculate error distances for any given image source between the two solutions. As such, the positions of the image sources are presented for visual examination for increasing reflection order, up to 3rd order, in Figure 7. The orientation of the listener is such that the direct sound originates from 0° azimuth, 0° elevation. The comparison in both number and location with the reference is quite good, with only a small number of discrepancies. As each subsequent order is based on the previous order's results, once a variation occurs the degree of variation naturally increases with increasing reflection order.

One must note that, while the brute force image source calculation can be used, software implementations commonly use various optimization methods or shortcuts to have even reasonably fast calculation times for complex models, thereby making the results susceptible to many subtle geometrical problems and errors. Within EVERTims, several parameters are used which affect the speed and accuracy of the results. First, there is a parameter defining skipping surfaces of "negligible" size used when clipping polygons against beams. An additional parameter affects the performance of the BSP tree. The configuration of these calculation constants depend on the scale of the model and while they could be determined automatically from the model data, their definition is currently the responsibility of the user.

The implication of each reflection, taking into account its accumulated absorption coefficient and polygon size, is an issue dealt with separately after the basic image source calculation in most packages. These different treatments are examined in Section 4.3, regarding the comparison of calculated basic room acoustic parameters from the results of the proposed package and the reference.

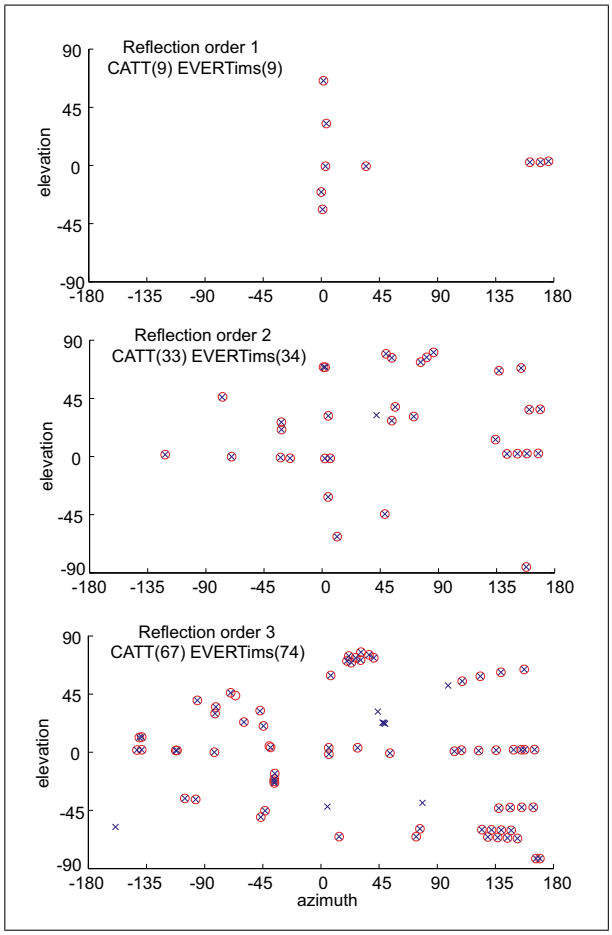

Figure 7. Comparison of the image source positions for a complex auditorium for EVERTims and CATT-Acoustic reference; results are presented for orders 1 to 3 (EVERTims (*x*), CATT-A (*o*)).

4.2. Reflection paths calculation time

An analysis of the algorithm's calculation time for the reflection paths was performed to validate if the package achieves interactive rates. The update times have been measured for different source–listener positions in both the full-calculation and the optimized update phase. Table III summarizes the results for when a new sound source is created in a static geometry, *i.e.* the update rate for the full-calculation phase. All values are averaged over different pre-defined measurement points distributed within the virtual space. The two values given in the table are for measurements with and without real-time visualization of the reflection paths. One can clearly see the downgraded performance when the visualization is updated in real-time. It should be noted that the calculation time not only depends on the number of polygons and the reflection order but also on the structure of the room model. An increase in the number of occluding elements reduces the speed of the beam-tree creation for a given number of polygons in the geometry model. Results for the reference package are also provided, where calculation times are reported relative to a zero order solution, which would include most preprocessing and initialization steps within the software.

It is noted that the minimum update rates in Table III are greater than those previously reported for the EVERT library [5]. This is due to the current frame-based processing structure of the VirChor implementation. A separate non-frame rate update thread is under development in order to improve performance for audio rate applications.

As the EVERT library is optimized for rapid updates for listener motion, not source motion, a simple test scenario was created for each room model. Table IV shows the results for a listener moving on a pre-defined path around a static sound source. The results are shown for rendering with real-time path visualization. A comparison of these results with the calculation times summarized in Table III clearly shows that there is a significant improvement in update times for complex room geometries using the proposed acceleration algorithms. Figure 8 shows the progression of the corresponding real-time early reflection visualization up to 3rd order. Performing the same task without real-time path visualization would further reduce update times.

Table III. Overall calculation time of the full-calculation phase after initialization steps. Results are shown for rendering with (on) and without (off) real-time path visualization.

Model	Polygons	Vis	Reflection order					
			1st (ms)	2nd (ms)	3rd (ms)	4th (ms)	5th (ms)	6th (ms)
Cube	6	off	12	14	14	19	27	47
		on	13	16	22	27	42	68
Complex corridor	235	off	12	155	1301	7616	35029	138710
		on	14	284	2092	11012	49033	179738
Simple auditorium	353	off	18	223	1669	11170	56240	230674
		on	22	322	2686	17473	86161	352432
Complex auditorium	1626	off	48	1377	17395	117615	528740	2095523
		on	52	1733	22555	135595	623288	2462240
Complex auditorium (CATT)	1626	–	90	2083	53443	219036	2961337	–

Table IV. Measured update time of 3rd order reflection paths during listener movements on a predefined path around the sound source. Results are shown for rendering with real-time path visualization.

Model	Polygons	Update (ms)
Cube	6	13
Complex corridor	235	24
Simple auditorium	353	33
Complex auditorium	1626	62

4.3. Room acoustic parameter comparison

While the various performance and speed evaluation methods are important to a real-time performance system, some comparisons of the quality of the acoustic auralization result is needed. A final comparison has been made relative to basic room acoustic parameters. The room model chosen for this comparison is the Fogg Art Museum Lecture Room (see complex auditorium room model Figure 6d). This is a complicated room, consisting of numerous curved and domed areas and it is not expected that the two solutions will converge. What is interesting to examine is to what degree a real-time solution can compare to a fully detailed more accurate non-real-time solution. With a generally high reverberation time, it exhibits many of the difficulties in room acoustics of small to medium sized rooms. Finally, one should note that Sabine's reverberation time equation, though developed during studies in this room, gives a poor indication of the reverberation time in this room. It is, in fact, not a "Sabinian" space.

A complete room acoustic model of this room had been developed and calibrated to measurements of the original room [70]. The affect of the scattering coefficient, and its importance in the calibration of the model was observed. This model has been adapted to the current framework. The impulse response (IR) has been measured within the system during the iterative image source calculation procedure, with 100 spatial encoders implemented in the auralization unit. From these measured IRs, the early decay time (EDT10) and reverberation time (T20) have been calculated for a given source and receiver position. Results from EVERTims are compared to those from the calibrated reference simulation, configured for 3rd order image sources, 1st order diffusion, and using the automatically determined number of rays for a 4 sec ray-trace calculation (calculation time approximately 1 hour). The results are shown in Figure 9. Figure 10 presents the broadband energy time curve (ETC) or reverse Schröder integration of the broadband impulse response. There is quite good agreement in the early portion of the ETC, and that while there is some difference in the octave band reverberation times, the ETC curves and hence the broadband reverberation times are quite similar. Due to the fact that the current package does not yet include air absorption, the resulting high frequency decay rates will be somewhat exaggerated, as is evident in the results.

There is reasonably good agreement for the early response analysis. It is possible to see the progression of the ISM results with each processing step, resulting in a mean frequency error of -0.2 sec, on the order of the just-noticeable difference for EDT as well as the observed variations due to parameter measurement errors [72]. The difference between the proposed system and the reference is most noticeable in the region of 50 – 200 ms. Calculating only reflections up to 3rd order means that this region is not well populated with calculated reflections and the response is heavily dominated by the performance of the FDN reverberation algorithm, which is clearly not accurate enough for this complex room in this early region of the IR. More processing power, allowing for a greater number of encoders, would improve these results.

For the late reverberation time calculation the Sabine reverberation time calculation is also shown. In contrast to EDT, the reverberation time comparison shows a more pronounced difference in relation to the reference. There are several possible explanations for these differences. The first is the late reverberation estimation method used here. In order to have rapid results, as discussed in section 3.3.2, the late reverberation time is estimated from the early image source decay response. This is in contrast to the off-

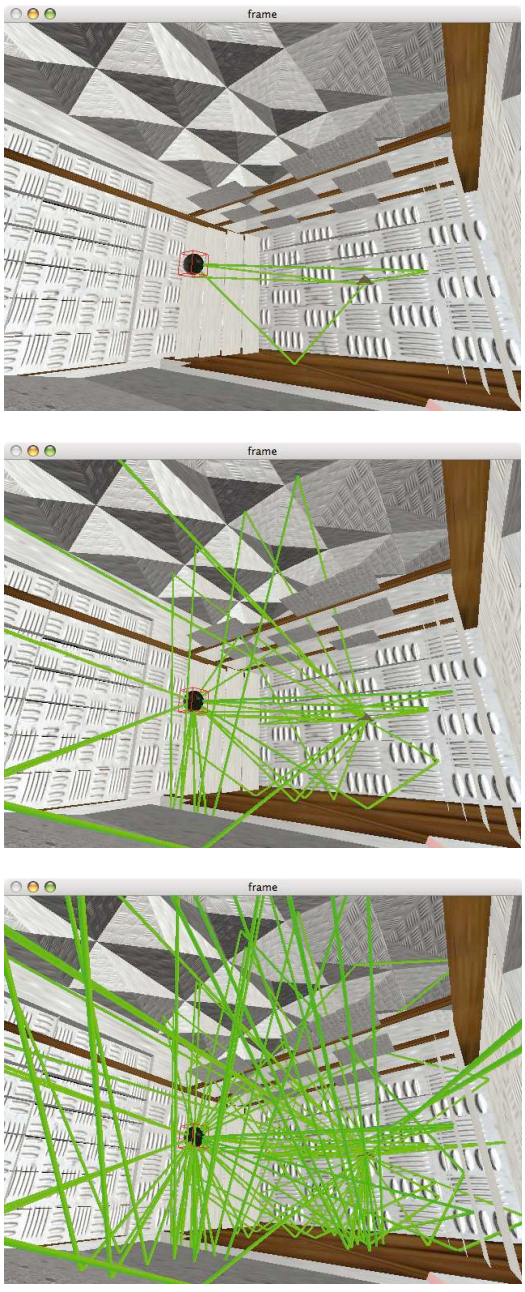

Figure 8. Time evolution of early reflections of 3rd order displayed in the 3D visualization unit.

line reference, which calculates the response through ray tracing up to several hundreds of milliseconds before estimating the late reverberation response. The second factor is the frequency band limitation used in the current real-time algorithm. To allow for real-time performance and a high number of encoders, the IR has been divided into three broad frequency bands rather than in octave bands. As a result, the reverberation time between adjacent octave bands will be more homogeneous than in a solution where narrower frequency bands are used. Examining the results shows that on average over these frequency bands the results are comparable. There is still some discrepancy at the upper and lower bands, which could be further improved. The frequency division limitation can be easily modified in the future, as more powerful processors are available. In addition, the frequency limits of the band divisions are user definable.

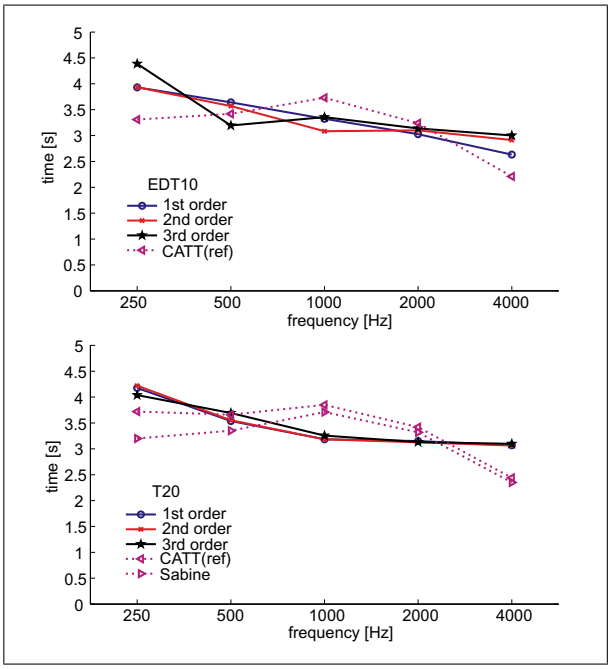

Figure 9. Early decay time (EDT10) and reverberation time (T20) of impulse responses (IR) measured in EVERTims during the iterative image source calculation procedure. The measured results are compared against results from a CATT-Acoustics simulation (CATT) and standard Sabine reverberation time calculation.

5. Conclusion

There is a growing need for real-time room acoustic auralization software for interactive and immersive environments. The aim of this project was to create a platform to obtain accurate rendering of complex rooms in interactive time. This paper presents a framework for performing this task in a novel manner. The framework has been constructed in a modular way, using self-contained units and a defined open protocol. Each unit has been made individually available as open source projects, available to the public community. This approach allows for future improvement, integration, and alternatives to be more freely and openly developed between research institutions. The authors have provided this framework as open source with the hope that it will be used by other researchers and developers to test their own algorithms, for any unit, and contribute any improvements to the research community.

The update rate of the auralization system has been shown to be within the limits for 'real-time' interaction for reflections up to 3rd order in complex rooms, achieving latencies of 62 ms for a complex room with over 1600 polygons for listener movement. When translational movement

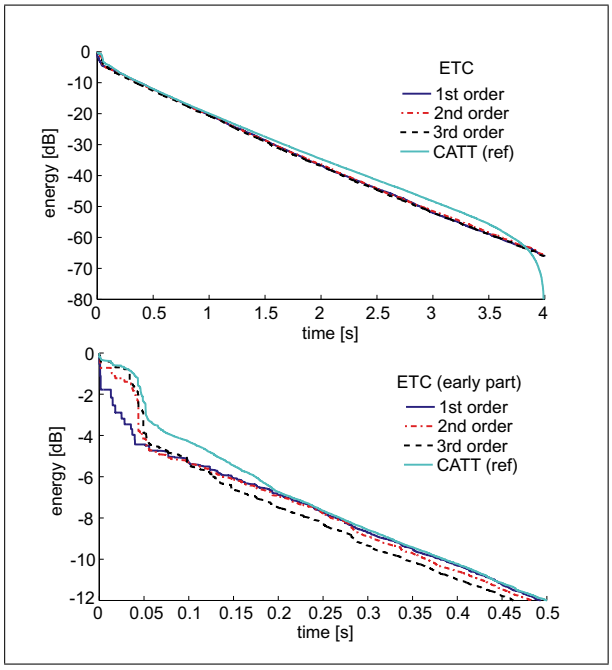

Figure 10. Comparison of broadband ETC for different reflection orders simulated in EVERTims and CATT-Acoustics.

is not occurring, the iterative solution continues to refine the auralization by providing higher order results. In addition, the iterative approach allows for latencies on the order of 50 ms for 1st order reflection for source or geometry modifications in the same complex room model. Additional options exist for further reducing the update latency, currently related to the frame-rate operation model of the graphical and scene rendering unit. The ability to distribute the computational load by allowing for different processing units to be on different machines also provides for potential speed improvements.

The current implementation presented here has been tested in terms of speed and accuracy performance relative to an analytical solution for a simple cube room and also in comparison to an existing commercial program, not intended for real-time use, for a highly complex room. Comparisons of calculated room acoustic parameters between the proposed platform and the commercial reference provided satisfying results, given the different intentions, uses, and constraints of the two packages. Early decay times were within the EDT JND. Late reverberation times varied more, with reasons being linked to the use of late reverberation estimation based on linear regression of the early decay. This is not as accurate as a full detailed ray-trace solution, but results were satisfactory for a fully interactive solution. Other limitations were seen to be linked to the use of only three frequency bands, a choice related to computation power in order to achieve real-time encoding of 100 reflection channels. Further improvements can be made in the optimization of encoder channels and frequency band decomposition to better balance processing power, result precision, and newer and more powerful hardware.

Appendix

A1. Protocols

A communication protocol has been defined for the inter-unit messaging of the different units. The communication is handled via OSC messages. The exact syntax of each protocol is described in detail in the documentation accompanying the software, but here is a short outline of them.

A1.1. From *scene graph* to *acoustic room modeler*

The room is composed of faces, which are sent as separate OSC messages containing the face identifier (`/face`), coordinates of the vertexes, and name of the material. A face can be any convex polygon. The absorption and diffusion characteristics of each material are defined in a separate file. For each message the face identifier is checked to see if it is new or describes a change of a defined face. A special message (`/facefinished`) indicates that the geometry model is complete and the beam-tree computation is launched. This concept is applicable for both initialization of the geometry and for real-time updates of an existing geometry. At present, the protocol does not handle the removal of faces from a model.

Sound source and listener data are sent in separate messages (`/source` and `/listener`) containing a unique identifier and information concerning location and orientation. The same message protocol is used for describing the creation of a listener and a change in an existing listener. The same applies to the sound sources. Examples of this protocol section format are as follows:

```
/face id material_id p0_x p0_y p0_z ... p3_y p3_z
/facefinished
/source id 4x4-float-matrix
/listener id 4x4-float-matrix
```

A1.2. From *acoustic room modeler* to *scene graph*

Communication from EVERTims to VirChor is only used for visualization of the reflection paths. The messages are used to turn on (`/line_on`) and turn off (`/line_off`) a line-segment with given coordinates. Lines have identifiers to enable switching off a certain segment. Examples of this protocol section format are as follows:

```
/line_on id x0 y0 z0 x1 y1 z1
/line_off id
```

A1.3. From *acoustic room modeler* to *spatial encoder*

EVERTims communicates to Pure Data information concerning all the visible reflection paths. Each time there is a change in visibilities they are updated via a message bundle containing all the changes for a certain source-listener combination. The bundle contains messages for source (`/source`) and listener (`/listener`) positions and their identifiers. They are followed by messages for each affected reflection path. The message can describe either a new visible reflection path (`/in`), update an already visible path (`/upd`), or tell that a reflection path has turned invalid (`/out`). The reflection path message contains the path identifier, source and receiver identifiers, the reflection order, the

first reflection point, last reflection point, length of the path, and accumulated absorption coefficients. The first reflection point is needed in implementation of the source directivity and the last reflection point is for 3D sound reproduction. This communication is one-way only and there is no feedback from Pure Data to EVERTims. The format of the reflection path message is a follows:

```
/refl path_id source_id listener_id order ...
r1_x r1_y r1_z rN_x rN_y rN_z dist ...
abs_0 abs_1 ... abs_9
```

A1.4. From *spatial encoder* to *spatial decoder*

The spatial encoder output format is a HOA multi-channel audio stream. This could be transmitted to any HOA decoder on the same computer or any distant computer. In the event where a head-orientation tracker is used and connected at the decoder, it is necessary for a supplemental message to be used from the *spatial decoder* to the *scene graph* to update the listener position. The direct sound can be sent on a separate audio stream for applications where an alternate rendering method is desired.

A2. Higher order Ambisonics (HOA) fundamentals

In the following the solution of the wave equation in spherical coordinates is shown, and the higher order Ambisonics (HOA) encoding and decoding equations are given. The equations presented refer to a spherical coordinate system in which the elevation θ is defined from the x–y plane. This coordinate system is related to Cartesian coordinates through

$$\begin{aligned} x &= r\cos\theta\cos\phi, \\ y &= r\cos\theta\sin\phi, \\ z &= r\sin\theta, \end{aligned} \tag{A1}$$

where r denotes the radius, ϕ the azimuth, θ the elevation, and the position vector is given as $\mathbf{r} = (r, \theta, \phi)$.

The general solution of the Helmholtz equation for the sound pressure $p(\mathbf{r}, \omega)$ in spherical coordinates writes as a Fourier-Bessel series as [31, 30, 37]

$$\begin{aligned} p(\mathbf{r},\omega) = &\sum_{m=0}^{\infty} i^m j_m(kr) \sum_{0\le n\le m,\sigma=\pm1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\phi) \\ &+ \sum_{m=0}^{\infty} i^m h_m^-(kr) \sum_{0\le n\le m,\sigma=\pm1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\phi), \end{aligned} \tag{A2}$$

A_{mn}, B_{mn} are the expansion coefficients, $i = \sqrt{-1}$. The wave number k is given as $k = 2\pi f/c$ where c is the speed of sound. The spherical Bessel functions of the first kind $j_m(kr)$ and the divergent Hankel functions $h_m^-(kr)$, represent the 'through-going' and 'outgoing' sound field, respectively. The angular solutions are given by the spherical harmonics

$$\begin{aligned} Y_{mn}^{\sigma}(\theta,\phi) = &\sqrt{(2m+1)(2-\delta_n)\frac{(m-n)!}{(m+n)!}} P_{mn}(\sin\theta) \\ &\cdot \begin{cases} \cos(n\theta) & if \quad \sigma = +1, \\ \sin(n\theta) & if \quad \sigma = -1, \end{cases} \end{aligned} \tag{A3}$$

which are defined by the associated Legendre functions $P_{mn}(sin\theta)$ of order n and degree m. The term δ_n is the Kronecker delta. Values of higher order associated Legendre functions can be found using recurrence relations [73]. For the interior problem, *i.e.* with all sources located outside the sphere, the solution reduces to

$$p(\mathbf{r},\omega) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{0\le n\le m,\sigma=\pm1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\phi). \tag{A4}$$

It can be seen from the above equations that the angular solution is separated from the radial solution.

A2.1. HOA encoding

The Ambisonics approach uses the plane-wave base to decompose the sound field, where the magnitude and phase is given by a complex signature function on a sphere. Note that the plane-wave base, or Herglotz wave function, is different from the far field representation of the sound field. The expansion coefficients and the signature function are related by integration over the sphere. For practical reasons the series expansions is limited to a maximum order M

$$p(\mathbf{r},\omega) = \sum_{m=0}^{M} i^m j_m(kr) \sum_{0\le n\le m,\sigma=\pm1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\phi). \tag{A5}$$

For an incident plane wave from direction (θ_S, ϕ_S), *i.e.* assuming the source with driving signal $S(\omega)$ in the far field, the Ambisonics encoding reduces to a simple weighting of the signal with spherical harmonics (real valued gain functions) [28]

$$A_{mn}^{\sigma} = S.Y_{mn}^{\sigma}(\theta_S, \phi_S). \tag{A6}$$

The B-format convention to term the Ambisonics channels uses capital letters arranged in some kind of reverse alphabetical order. These are not easily applied to HOA systems. The authors adopt the nomenclature used in geodesy and mathematics, which directly refers to the order n and degree m of the spherical harmonics. For example the B-format signals $[\ W\ X\ Y\ Z\ ]$ are written as $[\ X_0^0\ X_1^1\ X_1^{-1}\ X_1^0\ ]$, which is implemented in the software environment as [X0_0 X1p1 X1m1 X1_0].

A2.2. HOA decoding

The sound field in the center of the array is re-synthesized by the superposition of sound waves emitted from loudspeakers arranged on a sphere, also assuming plane waves. This can be expressed in matrix notation as

$$\tilde{\mathbf{A}} = \mathbf{CS}, \tag{A7}$$

where the elements of matrix $\mathbf{C}$ are defined by the corresponding spherical harmonics Y_{mn}^{σ}. Thus the loudspeaker driving signals can be derived from the Ambisonics coefficient vector $\mathbf{A} = [\ A_{00}\ A_{11}\ \ldots\ A_{mn}\ ]$ as

$$\mathbf{S} = \mathbf{DA}. \tag{A8}$$

From this equation the decoding matrix $\mathbf{D}$ derives by inverting the 're-encoding' matrix $\mathbf{C}$ under the assumption that the system is not under-determined

$$\mathbf{D} = \mathbf{C}^T(\mathbf{CC}^T)^{-1}. \tag{A9}$$

One can easily see from the above equation that there need to be at least as many loudspeakers as coefficients to decode.

To preserve the original curvature of encoded wave fronts, compensation filters can be used to take into account the radial components of equation A4, *i.e.* near-field compensated higher order Ambisonics (NFC-HOA) [74]. To assure stable digital compensation filters the encoder must be modified and becomes dependent on the decoder configuration. This means that the loudspeaker arrangement needs to be known at the time of encoding.

Acknowledgments

The authors would like to acknowledge many informative discussions and much helpful assistance on all aspects of VirChor from Christian Jacquemin and Rami Ajaj at LIMSI-CNRS. The authors would also like to thank Bengt-Inge Dalenbäck for his comments on the working manuscript.

References

[1] M. Kleiner, B. I. Dalenbäck, P. Svensson: Auralization – an overview. J. Audio Eng. Soc. **41** (1993) 861–875.

[2] D. R. Begault, R. Ellis, E. M. Wenzel: Headphone and head-mounted visual displays for virtual environments. Proc. AES 15th Int. Conference, Copenhagen, Denmark, 1998, 213–217.

[3] E. Meyer, W. Burgtorf, P. Damaske: Eine Apparatur zur elektroakustischen Nachbildung von Schallfeldern. Subjektive Hörwirkungen beim Übergang Kohärenz–Inkohärenz. Acustica **15** (1965) 339–344.

[4] M. Noisternig, A. Sontacchi, T. Musil, R. Höldrich: A 3D ambisonic based binaural sound reproduction system. Proc. AES 24th Int. Conferencde, Banff, Canada, 2003.

[5] S. Laine, S. Siltanen, T. Lokki, L. Savioja: Accelerated beam tracing algorithm. Applied Acoustics (2008) in press. Available at http://dx.doi.org/10.1016/j.apacoust.2007.11.011.

[6] M. Noisternig, L. Savioja, B. F. G. Katz: Real-time auralization system based on beam-tracing and mixed-order ambisonics. Proc 2nd ASA-EAA Joint Conference, Paris, France; J. Acoust. Soc. Am. **123** (2008) 3935.

[7] U. Svensson, U. Kristiansen: Computational modeling and simulation of acoustic spaces. Proc. AES 22nd Int. Conference, Espoo, Finland, 2002, 11–30.

[8] A. Krokstad, S. Strom, S. Sorsdal: Calculating the acoustical room response by the use of a ray tracing technique. J. Sound Vib. **8** (1968) 118–125.

[9] J. B. Allen, D. A. Berkley: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. **65** (1979) 943–950.

[10] J. Borish: Extension of the image model to arbitrary polyhedra. J. Acoust. Soc. Am. **75** (1984) 1827–1836.

[11] L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen: Creating interactive virtual acoustic environments. J. Audio Eng. Soc. **47** (1999) 675–705.

[12] D. Schröder, T. Lentz: Real-time processing of image sources using binary space partitioning. J. Audio Eng. Soc. **54** (2006) 604–619.

[13] R. Kajastila, S. Siltanen, P. Lunden, T. Lokki, L. Savioja: A distributed real-time virtual acoustic rendering system for dynamic geometries. Proc. AES 122th Int. Convention, Vienna, Austria, 2007, preprint nr. 7160.

[14] A. Farina: Verification of the accuracy of the pyramid tracing algorithm by comparison with experimental measurements of objective acoustic parameters. Proc. 15th Int. Congress Acoust. (ICA'95), Trondheim, Norway, 1995, 445–448.

[15] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, J. West: A beam tracing approach to acoustic modeling for interactive virtual environments. Proc. ACM Computer Graphics, SIGGRAPH'98, 1998, 21–32.

[16] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, A. Ngan: A beam tracing method for interactive architectural acoustics. J. Acoust. Soc. Am. **115** (2004) 739–756.

[17] C. Lauterbach, A. Chandak, D. Manocha: Interactive sound rendering in complex and dynamic scenes using frustum tracing. IEEE Trans. on Visualization and Computer Graphics **13** (2007) 1672–1679.

[18] V. Pulkki: Virtual source positioning using vector base amplitude panning. J. Audio Eng. Soc. **45** (1997) 456–466.

[19] V. Pulkki: Localization of amplitude-panned virtual sources II: Two- and three dimensional panning. J. Audio Eng. Soc. **49** (2001) 753–767.

[20] A. J. Berkhout, D. de Vries, P. Vogel: Acoustic control by wave field synthesis. J. Audio Eng. Soc. **93** (1993) 2764–2778.

[21] M. M. Boone, E. N. G. Verheijen, P. F. van Tol: Spatial sound-field reproduction by wave field synthesis. J. Audio Eng. Soc. **43** (1995) 1003–1012.

[22] D. de Vries: Sound reinforcement by wavefield synthesis: Adaptation of the synthesis operator to the loudspeaker directivity characteristics. J. Audio Eng. Soc. **44** (1996) 1120–1131.

[23] A. J. Berkhout, D. de Vries, J. J. Sonke: Array technology for acoustic wave field synthesis in enclosures. J. Acoust. Soc. Am. **102** (1997) 2757–2770.

[24] E. Corteel: Synthesis of directional sources using wave field synthesis, possibilities and limitations. EURASIP J. on Applied Signal Processing (2007) 188–206.

[25] D. H. Cooper, T. Shiga: Discrete-matrix multichannel stereo. J. Audio Eng. Soc. **20** (1972) 346–360.

[26] J. J. Gibson, R. M. Christensen, A. L. R. Limbert: Compatible FM broadcasting of panoramic sound. J. Audio Eng. Soc. **20** (1972) 816–822.

[27] M. A. Gerzon: Periphony: With-height sound reproduction. J. Audio Eng. Soc. **21** (1973) 2–10.

[28] J. Daniel: Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Ph.D. thesis, Université Paris 6, Paris, 2000.

[29] M. A. Poletti: Three-dimensional surround sound systems based on spherical harmonics. J. Audio Eng. Soc. **53** (2005) 1004–1025.

[30] P. M. Morse, K. U. Ingard: Theoretical acoustics. Princeton University Press, 1987.

[31] E. G. Williams: Fourier acoustics. Academic Press, 1999.

[32] N. A. Gumerov, R. Duraiswami: Fast multipole methods for the Helmholtz equation in three dimensions. Elsevier, 2004.

[33] R. Nicol, M. Emerit: 3D sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonics. Proc. AES 16th Int. Conference, Helsinki, Finland, 1999, 436–453.

[34] M. Poletti: A unified theory of horizontal holographic sound systems. J. Audio Eng. Soc. **48** (2000) 1155–1182.

[35] M. Poletti: The design of encoding functions for stereophonic and polyphonic sound systems. J. Acoust. Soc. Am. **44** (1996) 1150–1182.

[36] S. Bertet, J. Daniel, E. Parizet, L. Gros, O. Warusfel: Investigation of the perceived spatial resolution of higher order ambisonic sound fields: A subjective evaluation involving virtual and real 3D microphones. Proc. AES 30th Int. Conference, Saariselkä, Finland, 2007, 26–35.

[37] J. Daniel, R. Nicol, S. Moreau: Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. Proc. AES 114th Int. Convention, Amsterdam, Netherlands, 2003.

[38] S. Spors, H. Buchner, R. Rabenstein, W. Herbordt: Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering. J. Acoust. Soc. Am. **122** (2007) 354–369.

[39] E. Corteel: Equalization in an extended area using multichannel inversion and wave field synthesis. J. Audio Eng. Soc. **54** (2006) 1140–1161.

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

[41] J. M. Jot, S. Wardle: Approaches to binaural synthesis. Proc. AES 105th Int. Convention, San Francisco, USA, 1998.

[42] D. Menzies: W-panning and O-format, tools for object spatialization. Proc. AES 22nd Int. Conference, Helsinki, Finland, 2002.

[43] R. Duraiswami, D. Zotkin, Z. Li, E. Grassi, N. Gumerov, L. Davis: High order spatial audio capture and its binaural head-tracked playback over headphone with HRTF cues. Proc. AES 119th Convention, New York, NY, USA, 2005, preprint 6540.

[44] D. Menzies, M. Al-Akaidi: Nearfield binaural synthesis and ambisonics. J. Acoust. Soc. Am. **121** (2007) 1559–1563.

[45] B.-I. Dalenbäck: Room acoustic prediction based on a unified treatment of diffuse and specular reflection. J. Acoust. Soc. Am. **100** (1996) 899–909.

[46] W. Ahnert, R. Feistel: EARS auralization software. J. Audio Eng. Soc. **41** (1993) 894–904.

[47] G. M. Naylor: ODEON – another hybrid room acoustical model. Applied Acoustics **38** (1993) 131–143.

[48] J.-M. Jot, V. Larcher, O. Warusfel: Digital signal processing issues in the context of binaural and transaural stereophony. Proc. AES 98th Int. Convention, Paris, 1995, preprint 3980.

[49] T. Lentz, D. Schröder, M. Vorländer, I. Assenmacher: Virtual reality system with integrated sound field simulation and reproduction. EURASIP Aedvances in Sig. Proc. (2007) 187–187.

[50] D. Schröder, I. Assenmacher: Real-time auralization of modifiable rooms. Proc. 2nd ASA-EAA Joint Conference, Paris, France, 2008, J. Acoust. Soc. Am. **123** (2008) 3936.

[51] A. Silzle, P. Novo, H. Strauss: IKA-SIM: A system to generate auditory virtual environments. Proc. AES 116th Int. Convention, Berlin, Germany, 2004, preprint 6016.

[52] C. Borß, A. Silzle, R. Martin: Internet-based interactive auditory virtual environment generators. Proc. AES 14th Int. Conference, Paris, France, 2008.

[53] T. Moeck, N. Bonneel, N. Tsingos, G. Drettakis, I. Viaud-Delmon, D. Aloza: Progressive perceptual audio rendering of complex scenes. Proc. ACM SIGGRAPH'07 Symposium on Interactive 3D Graphics and Games.

[54] J. J. Embrechts: Broad spectrum diffusion model for room acoustics ray-tracing algorithms. J. Acoust. Soc. Am. **107** (2000) 2068–2081.

[55] E. M. Wenzel, J. D. Miller, J. S. Abel: Sound lab: A real-time, software-based system for the study of spatial hearing. Proc. AES 108th Int. Convention, Paris, 2000, preprint 5140.

[56] Virtual choreographer (virchor). Software released under GPL 2008, http://virchor.wiki.sourceforge.net/.

[57] J.-M. Jot, A. Chaigne: Digital delay networks for designing artificial reverberators. Proc. AES 90th Int. Convention, Paris, France, 1991, Preprint 3030.

[58] J. Daniel, J. B. Rault, J. D. Polack: Encoding of other audio formats for multiple listening conditions. Proc. 105th Int. Convention, San Francisco, CA, USA, 1998, preprint 4795.

[59] H. Wallach: The role of head movements and vestibular and visual cues in sound localization. J. Exp. Psychol. **27** (1940) 339–346.

[60] W. R. Thurlow, P. S. Runge: Effect of induced head movement in localization of direction of sound. J. Acoust. Soc. Am. **42** (1967) 480–488.

[61] F. L. Wightman, D. J. Kistler: Factors affecting the relative salience of sound localization cues. Lawrence Erlbaum, Mahwah, NJ, 1997.

[62] S. Perret, W. Noble: The effect of head rotations on vertical plane sound localization. J. Acoust. Soc. Am. **102** (1997) 2325–2332.

[63] F. L. Wightman: Resolution of front-back ambiguity in spatial hearing by listener and source movements. J. Acoust. Soc. Am. **105** (1999) 2841–2953.

[64] D. R. Begault, E. M. Wenzel, M. Andersen: Direct comparison of the impact of head tracking, reverberation and indiviualized head-related transfer functions on the spatial perception of a virtual speech source. J. Audio Eng. Soc. **49** (2001) 904–916.

[65] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman: Localization using nonindividualized head-related transfer-functions. J. Acoust. Soc. Am. **94** (1993) 111–123.

[66] J. Ahrens, S. Spors: Rendering of virtual sound sources with arbitrary directivity in higher order ambisonics. Proc. AES 123rd Int. Convention, New York, NY, USA, 2007, preprint 7243.

[67] B. F. G. Katz, F. Prezat, C. d'Alessandro: Human voice phoneme directivity pattern measurements. Proc. 4th ASA-ASJ Meeting, Honolulu, Hawaii; J. Acoust. Soc. Am. **120** (2006).

[68] M. Wright, A. Freed: Open sound control: a new protocol for communicating with sound synthesizers. Proc. Int. Computer Music Conference ICMC, Thessaloniki, Hellas, Greece, 1997, 101–104.

[69] M. Wright: Open sound control specifications. http://www.cnmat.berkeley.edu/OSC/OSC-spec.html, 2002.

[70] B. F. G. Katz, E. A. Wetherill: Fogg art museum lecture room, a calibrated recreation of the birthplace of room acoustics. Proc. 4th Acoustical European Meeting Forum Acusticum, Budapest, 2005, 2191–2196.

[71] Acoustic: Software release. http://www.catt.se/, 2008.

[72] B. F. G. Katz: International round robin on room acoustical impulse response analysis software 2004. Acoustics Research Letters Online **5** (2004) 158–164.

[73] G. Arfken: Mathematical methods for physicists, 2nd ed. Academic Press, New York, 1970.

[74] J. Daniel: Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. Proc. AES 23rd Int. Conference, Copenhagen, Danmark, 2003.