Digital Waveguide Networks for Room Modeling and Auralization

Matti Karjalainen
Helsinki University of Technology, P.O.Box 3000, FI-02015 TKK, Finland e-mail: **matti.karjalainen@hut.fi**,

Digital waveguide networks (DWNs) were proposed by Smith (1985) for the simulation of room reverberation. Since then they have been applied extensively to modeling of 1-D and 2-D resonators in musical instruments and to sound synthesis. Special cases of DWNs, digital waveguide meshes, have been studied also for 3-D modeling of acoustic spaces, and the research remains active until today. While mesh structures are computationally too complex for real-time simulation and auralization, DWNs are highly efficient for that purpose. DWNs simplify physical reality considerably, but may still be attractive from a perceptual point of view. The present paper discusses efficient DWN structures that can simulate source-receiver paths in rooms of considerable complexity if the model parameters are tuned appropriately. The model topology is based on the geometry of the room to be modeled, and thus many of the parameters are inherently related to real room acoustics. Bi-directional delay lines connect scattering nodes that are placed at the walls and reflecting surfaces. For best results, early reflections can be realized separately by delay paths according to the image source principle, but for generic room models all reflections and late reverberation are realized by the same DWN structure. Auralization by loudspeakers or headphones can be easily integrated into the models. The paper presents some basic DWN structures and discusses the results achieved with them in real-time modeling and auralization.

1 Introduction

Modeling and simulation of acoustic spaces can be carried out in many ways. Some methods, such as finite element (FEM) and boundary element methods (BEM), are based on numerical approximation of solving the wave equation. They can be made accurate, but on the other hand they are computationally expensive and cannot be applied to real-time simulation. For improved efficiency the ray-based geometrical methods are applied with useful results at mid-to-high frequencies. The image source approach is useful for solving the direct path and early reflections in real-time simulation, while the family of ray-tracing techniques is more an offline approach. Statistical methods are useful to compute the late reverberation. Artificial reverberation algorithms are used to do this in real time.

Real-time simulation and auralization of room acoustics is needed in virtual reality applications and it is useful also in rendering audible the results of acoustics designs. The real-time requirement sets strict limitations to the applicable simulation techniques so that the perceivable quality of the result is preferred over physical accuracy. A typical approach to real-time computation is to use the image source method to get parameters for delays and filters that simulate the direct sound and a number of early reflections, while reverb algorithms simulate the late part of the room impulse response [1]. Thus the early and the late parts are realized separately by very different techniques.

In this paper we explore possibilities to approximate both early and late parts with a single modeling technique, using the *digital waveguide networks* (DWNs). The principle was proposed by Smith already in 1985 [2], but surprisinly few publications have discussed the details of applying it to room simulation. The goal in the present paper is to formulate simple yet useful models that combine the geometrical rules of early sound and statistical behavior of late reverberation.

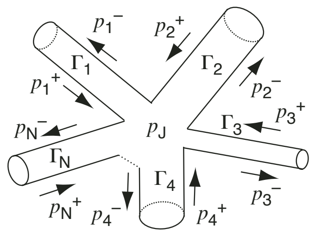

Figure 1: A scattering junction of connected acoustic tubes. Traveling pressure waves are denoted by '+' for incident and by '-' for scattered wave components.

2 Digital Waveguide Modeling and Synthesis of Acoustic Spaces

Digital waveguides are computational structures that are composed of bi-directional delay-lines for simulating wave propagation, filters to simulate losses and dispersion, and scattering junctions to connect delay-lines at nodes of wave scattering [3]. Figure 1 depicts a scattering junction of acoustic tubes with acoustics admittances Γ_i and traveling waves to and from the junction. Junction pressure p_J is obtained by formula

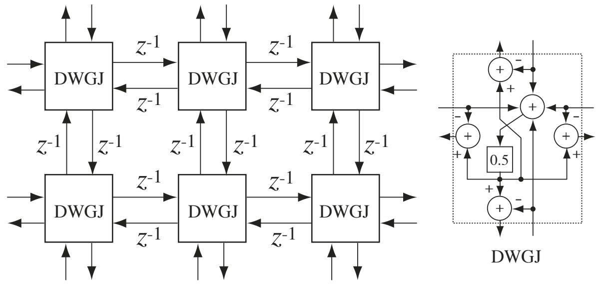

Figure 2: Left: part of a rectilinear 2-D waveguide mesh structure. Bi-directional unit delays z^{-1} connect scattering junctions (square blocks). Right: computation of a DWG scattering junction, Eq. (1).

$$p_J(n) = \frac{2}{\sum_{i=1}^{N} \Gamma_i} \sum_{i=1}^{N} \Gamma_i p_i^+(n), \qquad (1)$$

which can be computed as a signal flow diagram of Fig. 2 for a regular 2-D mesh-like rectilinear structure.

An equivalent but computationally more efficient mesh structure is obtained by *finite difference time domain* (FDTD) formulation [4] shown in Fig. 3, corresponding to state recursion formula

$$p_J(n+1) = \frac{2}{\sum_{i=1}^{N} \Gamma_i} \left(\sum_{i=1}^{N} \Gamma_i p_i(n) \right) - p_J(n-1) \quad (2)$$

Mesh-based models require in practice a spatial density of about 6 mesh points per wavelength to be accurate enough by physical criteria [5]. For an audio bandwidth of 10 kHz this means a mesh point distance of approximately 0.5 cm. For a medium room size of 4 x 6 x 3 m^3 this means 800 x 1200 x 600 = 576000000 mesh points, and for a concert hall of 30 x 60 x 10 m^3 this means 144000000000 mesh points. While the former one requires “only” a few gigawords of memory, the time consumed for simulation is far beyond real time even in the forseeable future, and for the concert hall hopelessly impractical even as a non-realtime simulation. Sparse DWN structures are thus needed for real-time simulation.

2.1 Realization of losses and dispersion

Dispersion and diffuseness[1] in artificial reverberation is obtained through the use of allpass and comb filter structures. In physical rooms these acoustic effects as well as losses take place typically in reflections and scattering at boundaries, including edge diffractions. Therefore, it is interesting to investigate DWN cases where scattering junctions at wall positions are designed to implement dispersive and lossy reflections to DWN delay lines by approximating physical reflections. The *reflectance* $S(z)$ is defined here as the ratio of reflected (-) and incident (+) pressure signals

$$S(z) = P^-(z)/P^+(z). \qquad (3)$$

If the wave admittance of a delay line attached to a junction is $\Gamma_R(z)$, then a given pressure wave reflectance $S(z)$ can be realized by loading the junction by admittance $\Gamma(z)$ so that

$$S(z) = \frac{\Gamma_R(z) - \Gamma(z)}{\Gamma_R(z) + \Gamma(z)} \leftrightarrow \Gamma(z) = \frac{1 - S(z)}{1 + S(z)} \Gamma_R. \quad (4)$$

For example a measured or synthesized reflectance can be used by loading a reflection junction by a proper wave digital admittance [7] or by realizing scattering coefficients according to Equation (1).

[1] Dispersion means here the temporal spreading of frequency components (allpass transfer function with frequency-dependent delay) of reflections or wave propagation. That is what can be done with filters in the DWN structures. Diffusion refers to the spreading of wavefront directions at boundary reflections. This does not take place in simplified DWNs due to a small number of possible wave propagation directions. For an overview of diffusion modeling, see [6].

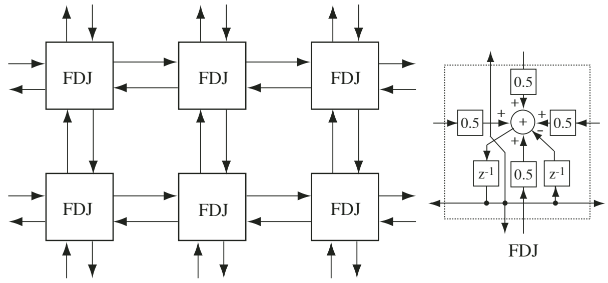

Figure 3: Left: part of a rectilinear 2-D FDTD mesh structure. Bi-directional delay-free connections link the junctions (square blocks) and the memory z^{-2} is a part of the junction. Right: computation of an FDTD junction.

2.2 Simple DWN for a rectangular room

To gain intuition about the possibilities of WDN modeling we will start from simple cases with a sound source and a receiver in arbitrary positions in a rectangular room. The conceptual framework for the early sound is based on the image source approach. For simplicity, only the horizontal plane is considered. Extending to the vertical dimension is conceptually straightforward.

Figure 4 shows a sound source (S) and a receiver (R) along with the direct sound path (thick arrow line S-R) and the first-order reflection paths (thin dotted lines). A highly simplified 2-D waveguide network consists of two waveguides (double-lines N1-N2 and N3-N4) through the receiver point and perpendicular to the walls of the room. This configuration is geometrically simple for real-time auralization and dynamic control of model parameters for moving sources and receivers.

It is easy to notice that the direct sound always needs special treatment, i.e., a delay from S to R with a proper gain factor for attenuation. This path is totally independent of

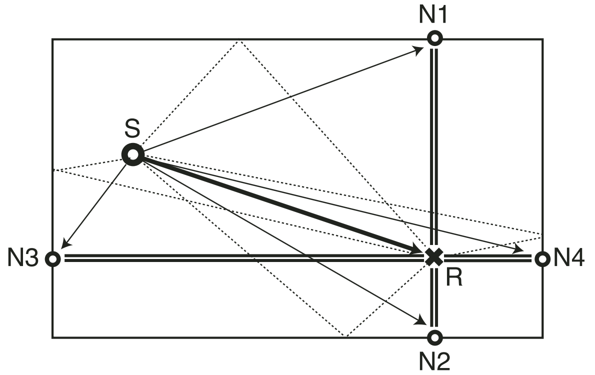

Figure 4: Simple DWN configuration in a rectangular room in the horizontal plane. S = sound source, R = receiver, and N1-N4 are scattering junctions. Thick arrow line: direct sound path; thin dotted lines: real acoustic paths of first-order reflections; and thick double axial lines: waveguides (bi-directional delay lines). The thin solid arrow lines are for feeding junctions to approximate the first-order reflection paths.

the walls and size of the room so it cannot utilize DWN nodes that are on the walls, which are required to realize the temporal structure of reflections between walls.

Let's study first a simple case where the waveguides N1-N2 and N3-N4 are uncoupled. Both waveguides are simple bi-directional delay lines with terminations at the walls, so they can be designed independently for frequency-dependent decay and dispersion by adding proper lowpass filtering and allpass structures within each delay line loop. The waveguides also need to be fed by delayed and scaled excitation inputs from the source S (thin arrow lines) to nodes N1-N4. The receiver R just senses the incident wave components in the delay lines for auralization.

The inherent limitations of simple DWNs become evident from this configuration. While the direct sound is treated separately and is therefore no problem, already the first-order reflections arrive from incorrect directions (the arrival times can be adjusted by the delays from source S to nodes N1-N4). Another general problem in addition to the direction error is the signal level error of reflections. This can be noticed for example when the source is near any of the nodes, say N2. For the first reflection the source coupling to that node should be strong, but this leads to the fact that the later reflections along that waveguide N1-N2 will remain almost as strong (unless decay time is very short). This problem comes from the spherical nature of the wavefront near the source, which doesn't fit naturally to the sparse DWN structure.

The next limitation of the simple case in Fig. 4 is that wave components and thus also modes are possible only axially to the wall directions, and no cross-modes can

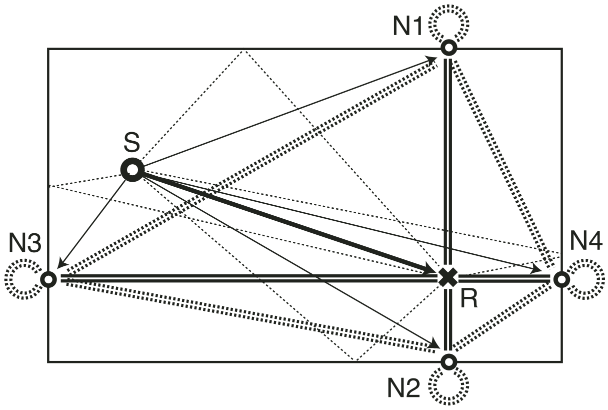

Figure 5: DWN configuration in a rectangular room with added non-axial waveguides (dotted double lines) and loading of nodes for losses and dispersion/diffusion (circular dotted double lines).

emerge. This may be even desirable for special room effects, such as strong flutter echos, but not for yielding perceptually valid rendering of normal room acoustics. By adding enough dispersive elements within the simple DWN model, however, it can be made a generic room simulator without too disturbing artifacts.

One possible improvement to the simple model is shown in Fig. 5. The waveguides marked by dual-dotted lines between nodes N1-N4 have been added to allow for coupling of the axial waveguides[2] and to allow the build-up of "cross-modes", although the receiver still perceives reflections as coming only from axial directions.

Another minor detail is the possibility to move the nodal points N1-N4 at the walls so that the first-order reflections come from directions closer to the arrival angles in the real room. However, this means that all later reflections also come from these directions. If the accuracy of first-order reflections is crucial, they can be realized in a way similar to the direct sound, i.e., by separate delays from the source. This kind of special treatment of early reflections means, however, that the DWN structure is used more or less for late reverberation only.

Compared to the case of Fig. 4, the waveguides in Fig. 5 are truly connected. Thus the design for losses and dispersion/diffusion is more complex. From a physical point of view, most of these phenomena take place in reflections at the walls. Thus it is natural to keep the waveguides lossless and dispersionless (except for minor air absorption loss), and simulate the phenomena in the nodes at the walls. Losses can be added by loading the junctions by wave digital admittances [7] that have a proper resistive component for absorption simulation. Disper-

[2] Another way to achieve coupling between the axial waveguides is to make a scattering junction at the receiver position R with weak scattering coefficients between the main directions. This is, however, not motivated from real physical behavior of rooms.

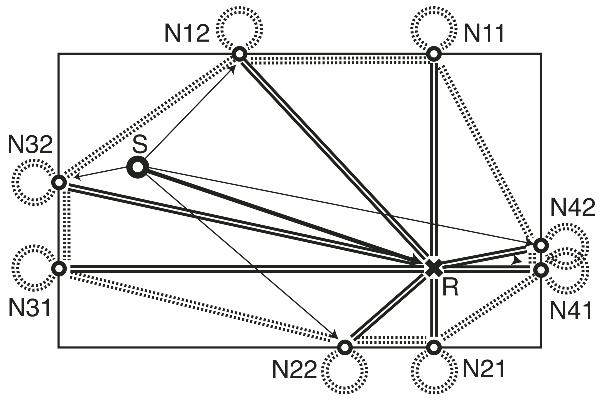

Figure 6: A DWN structure where secondary waveguides (N12-R-N22, N32-R-N42) are added to correct arrival directions of first-order reflections. For improved cross-modes and reflections, all nodes are connected into a polygon path (dotted double lines). Losses and dispersion/diffusion at the walls is increased by circular connections at each node (circular dotted double lines).

sion/diffusion can be introduced by reactive loading. This can be realized with different recursive structures, such as star-like DWNs [8], wave digital admittances, or circularly connected waveguides as characterized in Fig. 5.

2.3 Adding nodes to DWN models

DWN room models can be made closer to physical reality by adding nodes and delay lines between them. There is an unlimited number of different configurations from the simple models discussed above to a full-density digital waveguide mesh. Here we briefly mention another relatively simple DWN structure, shown in Fig. 6. Secondary waveguides (N12-R-N22 and N32-R-N42) have been added which allow correct first-order reflection directions and improve directional diversity of wave arrivals. Notice that these lines make an angle at the receiver point and thus do not simulate physical wavefront behavior. The DWN nodes at the walls are connected into a polygon structure of waveguides (dotted double lines) to improve the cross-mode behavior. Each node is also circulated to itself through a dispersive structure as was already shown in Fig. 5.

If the geometry of a room to be simulated is more complex or there are reflecting and diffracting objects inside the room, the image source approach for approximating the waveguide junction positions is still generally useful. Diffraction from edges can also be taken into account. In fact, edge diffraction modeling and image source modeling can be integrated together [9]. If the first order (and second order) reflections and diffractions need accurate directional and level estimates, it is probably better to implement them as separate image source paths than to try to realize them as part of a DWN structure.

2.4 Late reverberation

Since the times of Schroeder [10] and Moorer [11] the goal of artificial reverberation has been naturalness without artifacts, but typically in a generic form and not calibrated in detail for any specific room. The complexity of reverberation in real spaces is high and the auditory system is not very sensitive to fine details. Therefore modeling the early part with more accuracy by various techniques and combining this to generic late reverberation works fine in many cases.

Important features of generic late reverberation are the lack of undesirable spectral peaks or valleys and a response envelope without periodicities in any auditory critical band. The quality and artifacts of reverberation can be evaluated to some degree by auditory modeling analysis [12, 13]. In the future this could also be a way to automatically adjust room model parameters to fit perceptually to given (measured) room responses.

In DWN room simulation discussed in this paper the late reverberation is closely integrated with the early response. Therefore there is need to compromise between these requirements. Because the topology of DWNs is in large part dictated by room geometry and positions of the source and the receiver, the temporal and spectral richness of reverberation has to be controlled mainly by dispersive scattering elements in the model.

3 Auralization of DWN models

For virtual acoustics applications the result of room simulation needs to be rendered, i.e., auralized by loudspeaker or headphone reproduction [14, 15, 1, 16]. For loudspeaker reproduction there are two basic cases: multichannel techniques or the binaural (transaural) technique. In binaural reproduction by headphones or a pair of loudspeakers the goal is to get proper signals into the listener's ear canals. In multichannel loudspeaker reproduction the goal is to obtain the desired wave field at the position of the listener's head. Here we discuss separately two cases: utilizing the wave components and simulation of the sound pressure field.

When the wave components arriving at the receiver position are computed, including the wavefront direction, auralization can be done by panning techniques or HRTF (head related transfer function) processing. Amplitude panning, particularly the *vector base amplitude panning* (VBAP) [17], is a straightforward and computationally efficient method to create virtual sound sources in desired directions both in 3-D and 2-D (horizontal) reproduction by loudspeakers. *Ambisonics* [18] is another method where a similar approach can be applied. Referring to the figures in Section 2, for each delay line entering at the receiver point, the corresponding signal needs to be panned to the direction of signal arrival.

A problem with loudspeaker reproduction is that the listener cannot move in a space without leaving the sweet-spot area. For a single subject, listener tracking can be done and the varying position can be taken into account for example in the VBAP method. However, for multiple listeners or for cases where the subject comes close to a single loudspeaker, the method may not work well. In such applications *wave field synthesis* [19] is a better solution, but at the cost of a high number of loudspeakers and large amount of related signal processing.

For headphones and binaural loudspeaker reproduction the directions of arrival can be realized by HRTF processing. For directional accuracy the first wave front is the most important one, and should therefore be processed with highest precision HRTFs. Due to the precedence effect the early reflections contribute mostly to timbre and perceived source width, and practically not at all to perceived source direction. Thus the accuracy of HRTFs for those paths can be compromised to some degree. To make auralization applicable to subjects moving in a virtual space, head-tracking has to be applied and the signal paths need to be updated accordingly.

One possibility for headphone auralization, particularly with digital waveguide mesh structures, is to model also the head as a mesh. This is like simulating virtual ears on a virtual dummy head. The head can be assumed to be a hard object, spherical or more detailed in shape. To obtain a detailed HRTF behavior a highly detailed DWN model is needed, which means that real-time processing is probably not possible.

4 DWN modeling experiments

The principles of simplified DWN room models discussed in Section 2 were simulated in BlockCompiler [20], a block-based modeling environment for real-time simulation and synthesis. Simple graphical user interfaces were applied to visualize the room geometry and to move the source and the receiver by mouse during real-time simulation.

The goal was to gain basic understanding of the possibilities of integratinf the early and late parts of room response through the use of geometrically intuitive DWN models. Binaural HRTF-based headphone auralization was applied to listen to signals such as speech and music when processed through the models. Here we present some basic experiences from the simulations.

Sparse digital waveguide networks used here are computationally efficient. The room models of Section 2 take about 5-15% of CPU time on a 1 GHz G4 PowerPC processor when a single source and a single receiver are simulated and no specific optimization of the software implementation is done. HRTF-based auralization, however, is more computation intensive. The five directions of auralization in Fig. 4 take about 30 % of CPU time of the processor mentioned above.

Adding new sources increases the computation load less than by the number of sources, while each receiver needs in most cases duplicating the DWN structure. The use of multiple sources and receivers expands the complexity rapidly. In this sense these DWNs are computationally more expensive than algorithms where the DWN principle is used only for reverberation common to every receiver in the room.

Simulation of true true geometry of room in the models of Section 2, though in a simplified manner, cannot fully compete in efficiency with source-filter type of models, such as feedback delay networks, or even with simple DWNs, such as the star-like DWN reverberator [8]. This is due to the overhead of using bi-directional dual delay lines and non-vectorizable data structures. However, for a single-source single-receiver case the difference in computational load is not radical, and the speed of modern computers helps using complex DWN topologies. An advantage of the DWN models discussed in Section 2 is the direct connection of model parameters to the geometry of the room.

For high quality early reflections there is probably only one way to go, that is, to realize them separately as delay lines form source to receiver according to the image source principle. This does not increase radically the computational load unless there are many reflecting surfaces or complex dispersion filters. Even when realizing early reflections separately, it is important to excite the DWN structure as early as possible for the build-up of reflections and modes.

The selection of model parameters comes partly from the room geometry, partly it can be derived from measurements of a real room, but quite much it is art of guessing and perceptual tuning, as with any approximate modeling of rooms. The lengths of delay lines is an easy case, determined by the geometry of the room and positions of source and receiver. The coupling factors of the source to the scattering nodes through delays is already a question where good compromise rules are needed.

Losses and reverberation time are fairly easy to tune in the simplest model of Fig. 4, because each waveguide branch is independent. The reflections at the walls can be designed to realize losses and dispersion. In more complex configurations the control of reverberation time as a function of frequency becomes less easy.

Designing good dispersion structures at the scattering nodes (or as part of waveguide delay lines) is one of the most important and difficult factors to make the models sound good and realistic. This is where good guesswork and careful experimentation yields the most natural sounding results in reverb design. An interesting challenge in the room-related DWNs is to utilize measured data from a given room. In the model of Fig. 4 the reflectance at the scattering node could be made to correspond to an average measured reflectance of the corresponding real surface.

When listening to the DWN room models described above through real-time headphone auralization it was found that it is fairly easy to make realistic sounding simulation for small to medium sized rooms having relatively short reverberation times. For large spaces and long reverberation times the design of dispersion becomes demanding, and the simple DWN topologies may not model details of more complex geometries, at least unless more of the early reflections are implemented separately according to the image source principle.

5 Summary

This paper has explored the use of digital waveguide networks for real-time room simulation. The methods discussed combine the principles of digital waveguide networks, image source principles, reverberation algorithms, and HRTF-based rendering. Examples were given on synthesizing room responses for simple geometries.

Acknowledgments

The work has been supported by the Academy of Finland (project 105629). Special thanks are due to Patty Huang and Julius O. Smith of Stanford University, who collaborated on an overview paper to be published in AES 118th Convention in Barcelona, 2005.

References

[1] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, 1999.

[2] J. O. Smith, "A new approach to digital reverberation using closed waveguide networks," in *Proc. 1985 Int. Computer Music Conf.*, (Vancouver, Canada), pp. 47–53, 1985.

[3] J. O. Smith III, *Physical Audio Signal Processing: Digital Waveguide Modeling of Musical Instruments and Audio Effects*. Available online at `http://ccrma.stanford.edu/~jos/-pasp04/`, August 2004.

[4] M. Karjalainen and C. Erkut, "Digital waveguides versus finite difference structures: Equivalence and mixed modeling," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 7, pp. 978–989, 2004.

[5] L. Savioja, *Modeling Techniques for Virtual Acoustics*. Helsinki University of Technology, 1999. Doctoral Thesis, Report TML-A3.

[6] H. Nironen, *Diffuse Reflections in Room Acoustics Modelling*. Helsinki University of Technology, 2004. MSc Thesis.

[7] A. Fettweis, "Wave digital filters: Theory and practice," *Proc. IEEE*, vol. 74, pp. 270–327, Feb. 1986.

[8] J. O. Smith and D. Rocchesso, "Connections between feedback delay networks and waveguide networks for digital reverberation," in *Proc. 1994 Int. Computer Music Conf.*, (Århus, Denmark), pp. 376–377, 1994.

[9] P. Svensson, R. I. Fred, and J. Vanderkooy, "An analytic secondary source model for edge diffraction impulse responses," *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2331–2344, 1999.

[10] M. R. Schroeder, "Natural-sounding artificial reverberation," *Journal of the Audio Engineering Society*, vol. 10, no. 3, pp. 219–223, 1962.

[11] J. A. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, no. 2, pp. 13–28, 1979.

[12] M. Karjalainen and H. Järveläinen, "More about this reverberation science: Perceptually good late reverberation," in *Preprint of 111th AES Convention*, (New York, USA), 2001. Paper 5415.

[13] P. Huang and J. O. Smith, "Loudness-based display and analysis applied to artificial reverberation," Computer Music Association, Nov. 2004.

[14] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.

[15] M. Kleiner, B.-I. Dalenbäck, and U. P. Svensson, "Auralization — An overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, 1993.

[16] T. Lokki, *Physically-Based Auralization – Design, Implementation, and Evaluation*. Helsinki University of Technology, 2002. Doctoral Thesis.

[17] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

[18] M. A. Gerzon, "Panpot laws for multispeaker stereo," in *Preprint of 92th AES Convention*, (Vienna, Austria), 1992. Preprint 3309.

[19] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, 1993.

[20] M. Karjalainen, "BlockCompiler: Efficient simulation of acoustic and audio systems," in *Preprint of 114th AES Convention*, (Amsterdam), 2003. Convention paper 5756.