RENDERING WALK-THROUGH AURALISATIONS USING WAVE-BASED ACOUSTICAL MODELS

Alex Southern, Jeremy Wells, and Damian Murphy

Audio Lab, Department of Electronic
University of York, YO10 5DD, York, UK
aps502, jjw100, dtm3
@ohm.york.ac.uk

ABSTRACT

Room acoustic simulations using wave-based models offer several advantages over more established geometric acoustic approaches such as simulation accuracy at low frequencies and wave-propagation effects such as diffraction and occlusion. However, geometric approaches can be adapted for real-time implementation and hence are more appropriate for applications that demand real-time interactivity such as walk-through auralisation or scene rendering for computer game audio. This paper presents a method that facilitates real-time dynamic auralisation for wave-based approaches based on a pre-computed spatial subset of B-format encoded impulse responses for a given environment.

Different methods are explored for post-simulation processing of this dataset to enable real-time dynamic auralisation based on both interpolation and Ambisonic-style manipulation of the resulting B-format impulse responses. Comparisons are made against a simple pre-computed walkthrough and demonstrate that this approach has significant potential for facilitating dynamic auralisation for wave-based room acoustic simulations.

1. INTRODUCTION

Room acoustic simulations using wave-based models such as the finite difference time domain method (FDTD) [1] or Digital Waveguide Mesh (DWM) [2] offer several advantages over more established geometric acoustic approaches. In particular they offer improvements in simulation accuracy at low frequencies and inherent wave-propagation effects such as diffraction and occlusion. The main disadvantages of wave-based modelling strategies are the excessive computation times and memory requirements if full audio-bandwidth simulations are demanded, with the former running to days and potentially weeks for anything other than the smallest of 3D spaces. Hence wave-based room acoustics modelling is always considered as an offline process, where the goal is to obtain the source/receiver room impulse response (RIR) which can then be implemented in real-time for any audio input signal via convolution. The consequence of this process is that the static RIR imposes significant limitations on interactivity and dynamic auralisation. Virtual room acoustic simulations have significant potential in areas such as audio for computer game environments and virtual reality applications, but interactive dynamic auralisation in real-time is a necessity for effective, immersive and realistic results, and this clearly cannot be dealt with in current wave-based approaches.

With geometric acoustic algorithms, it is generally accepted that there are limitations as to what can be achieved in terms of absolute objective accuracy, although the RIRs produced can give a good perceptual representation of a particular source/receiver/room combination. However real-time, interactive dynamic auralisation is still not possible without imposing further limitations or assumptions on the basic algorithm. The DIVA system [3] employed first reflection synthesis via the image-source method [4] together with simple directional filtering, combined with HRTF processed direct sound and reverberant tail synthesis using a recursive delay network. Although real-time binaural auralisation was possible, the Binaural RIRs that result do not represent an exact representation of the geometry being modelled. The acoustic radiance transfer approach [5] offers a significant improvement over simple early-reflection modelling as it enables sound propagation paths between surfaces within the space to be pre-computed for the whole RIR. The result is the synthesis of a moving receiver and hence the potential for a dynamic walk-through simulation of the virtual space although source positions must remain fixed.

An alternative approach that does not impose such limits, either in terms of source/receiver movement or in the capabilities of the underlying algorithm, is the pre-computation of a RIR dataset for different source/receiver combinations, sufficient to cover the range of possibilities required in the final simulation. Real-time walkthroughs can then be synthesized via convolution with the appropriate RIRs for a given time and position, with interpolation between RIR spatial sampling points. This method has been used in some room modelling auralisation schemes [6][7] as well as with real, as opposed to synthesized, RIR datasets [8][9]. This method can be further enhanced through the use of B-format encoded RIRs to allow post-simulation manipulation of spatial attributes, such as receiver head-rotation [7].

Recent developments with wave-based simulations have introduced the use of high-order B-format spatial encoding strategies based on RIRs captured across multiple spatial sample points [10]. It therefore becomes possible to introduce receiver interactivity with an otherwise static RIR obtained from an FDTD/DWM simulation. This paper explores for the first time the potential for synthesizing dynamic walk-through style auralisation using this approach and the remainder of this paper is organized as follows: Section 2 gives a brief overview of the FDTD scheme used in this work; section 3 compares the computational resources required to implement dynamic auralisation via wave and geometric acoustic methods; section 4 presents some simple test case scenarios to demonstrate the potential for this work and section 5 considers the conclusions drawn and identifies the most promising paths for furthering this work.

2. ACOUSTICAL MODELLING

Both FDTD and DWM methods are multi-dimensional physical modelling techniques that have shown to be suitable in room acoustics simulations [1][2]. The two methods differ in that the numerical approximation of wave propagation within modelled enclosures is approached differently. In a FDTD scheme, wave propagation is simulated by numerically approximating the wave equation using 2^{nd}-order finite differences to determine acoustic pressure over a regularly spaced spatio-temporal sampling grid as shown in Figure 1(a). The calculation of acoustic pressure at a grid point at the next time sample is performed using the update equation given in (1) in conjunction with Figure 1(b):

$$p_{i,j}^{n+1} = 0.5 \cdot [p_{i+1,j}^{n} + p_{i-1,j}^{n} + p_{i,j+1}^{n} + p_{i,j-1}^{n}] - p_{i,j}^{n-1} \qquad (1)$$

where i and j are grid positions, indexed over the x and y axes, while n is the current time step. Therefore $p_{i,j}^{n+1}$ is the pressure for the next time step at the grid point location i,j.

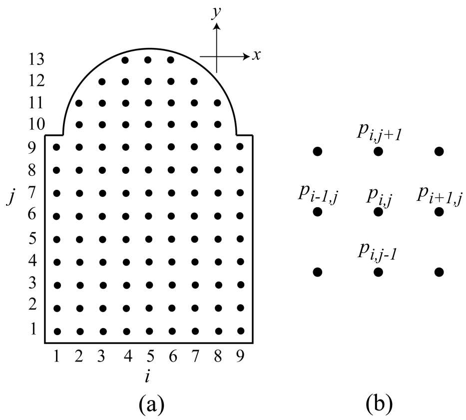

Figure 1: *A regularly spaced grid of points used to discretely model a continuous propagation medium within an enclosure (a) and (b)*

In the DWM technique wave propagation across the saptio-temporal sampling grid is achieved via travelling waves moving between grid/mesh points based on the principles of Digital Waveguides [11]. The reader is referred to [2] and [11] for a full description of this approach.

In this paper only 2D representations of rooms are considered, yet both the FDTD/DWM schemes can be extended to the 3D case [12]. It may be observed that some grid points have less than 4 neighbours. These can be defined as either *boundary nodes* or *corner nodes* and are implemented differently for calculating acoustic pressure at their location. Although beyond the scope of this paper such grid points may also be attributed with frequency dependent and diffusive characteristics as discussed in [13] and [14].

2.1 Spatial Encoding

The spatial encoding of wave-based acoustical models for auralisation over a multichannel sound system has been discussed previously in [10]. The modelled soundfield is encoded into B-Format and therefore provides flexibility in choosing the auralisation technique i.e. 2D/3D Ambisonics [15], Binaural [16] and SIRR/DirAC [17]. In this paper, only 1^{st}order B-Format is required and so a brief overview for obtaining such signals is provided here.

1^{st}order B-Format consists of 4 audio channels WXYZ that correspond to a spherical harmonic decomposition, or sampling, of the soundfield at a central point over time. A pressure microphone is used to capture the W channel and 3 pressure gradient microphones orientated along the *xyz* Cartesian axes are used for the XYZ channels respectively. For the purposes of this work only 2D soundfields are considered and as such the vertical axis z and therefore the Z channel are not required, the 2D microphone arrangement is provided in Figure 2.

In a real environment it is not physically possible to have multiple microphone capsules occupy the same point in space. Therefore, alternative arrangements are employed such as a tetrahedral array of sub-cardioid capsules whose 4 outputs are collectively known as A-Format. It is then possible to perform an A-Format to B-Format conversion to obtain the WXYZ signals [18]. In contrast, when using a FDTD/DWM scheme for acoustical modelling, B-Format may be attained directly from the soundfield by using multiple omnidirectional pressure sensitive receivers to form the orthogonally orientated pressure gradient microphones for X and Y. The general receiver arrangement for capturing 2D B-Format is provided in Figure 3 and processed using (2),(3) and (4):

$$W(t) = p_1(t) \tag{2}$$

$$X(t) \approx \left(\frac{1}{\rho_0 \cdot d}\right) \int_{-\infty}^{t} [p_2(\tau) - p_3(\tau)]\, \delta\tau \tag{3}$$

$$Y(t) \approx \left(\frac{1}{\rho_0 \cdot d}\right) \int_{-\infty}^{t} [p_4(\tau) - p_5(\tau)]\, \delta\tau \tag{4}$$

where $W(t)$, $X(t)$ and $Y(t)$ make the 2D B-Format signal and are formed using receivers $p_1(\tau)$ to $p_5(\tau)$, ρ_0 is the particle density in air and distance d is depicted in Figure 3. Equations (3) and (4) are based on the finite difference approximation of sound velocity between two closely spaced pressure sensors, adapted from [19]. Processing this array of receivers by considering (2),(3) and (4) for discretely sampled signals, a 1^{st}order spatial auralisation of an FDTD/DWM acoustically modelled soundfield may be realized.

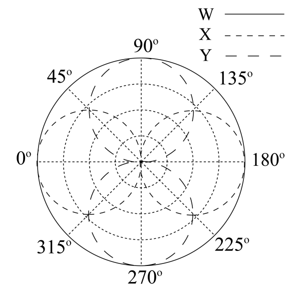

Figure 2: *The 2D microphone arrangement required to encode the WXY channels of a B-Format signal*

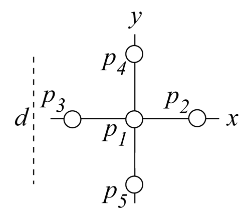

Figure 3: *The 2D receiver arrangement required to encode the WXY channels in a FDTD/DWM scheme*

3. COMPARING COMPUTATIONAL EFFICIENCY

In a previous paper, performance benchmarking was conducted to compare the computation times associated with both typical leading ray-based methods and the more recent wave-based FDTD/DWM modelling approaches [20]. In this paper we conduct similar performance testing but from the viewpoint of generating RIR datasets suitable for rendering dynamic walk-through auralisations.

In Figure 4 a test geometry is defined that has an approximate volume of $27m^3$ and pressure measurements are made using the array of 359 receivers with both ray-based and wave-based acoustical modelling. In Table 1 the typical processing time taken to measure each dataset is presented. Ray-based results are generated using ODEON, wave-based results are generated using [21], both on a standard desktop PC. Although Table 1 does not represent a

Table 1: *Performance comparison between ray-based and wave-based generation of RIR dataset. Time is expressed as (hours:mins)*

	Ray	Wave (2D)	Wave (3D)
Single Receiver	00:13	00:37	14:18
359 Receivers	70:00+	00:37	14:18

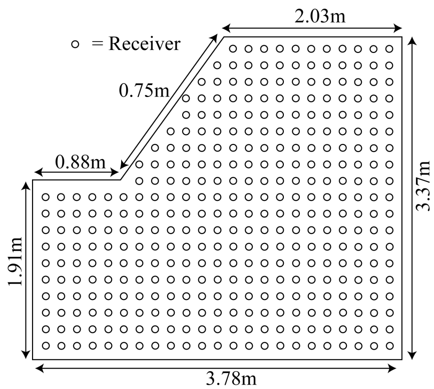

Figure 4: *The performance test setup.*

rigorous benchmarking test, it is clearly apparent that the processing times for the wave-based methods do not vary if more receivers are added. Furthermore, though the ray-based approach is faster than the wave-based 2D/3D case for a single receiver it is clear that if many receiver outputs are required then the wave-based approaches in general are more suitable. The results also suggest that 3 source/receiver combinations for the ray-based case will take approximately the same computation time for one 2D wave-based case that can potentially yield a RIR for every mesh point for one source with little additional overhead. In the 3D case 66 source/receiver combinations could be dealt with using ray-tracing before the 3D mesh becomes a more efficient approach. Note that exact ratio for these relationships is geometry specific.

4. A SIMPLE WALK-THROUGH DEMONSTRATION

As real-time computation of RIR datasets using FDTD/DWM is not currently feasible for full audio bandwidth and complex room geometries, pre-calculation is necessary. The required amount of receivers and the chosen inter-receiver spacing is dependent on both the application and efficiency of the real-time RIR interpolation as the listener navigates through the virtual space. In the previous section it was established that at no extra computational cost than that for one RIR, hundreds of RIRs may be modelled using a FDTD/DWM scheme to generate a suitable dataset. For this reason, this paper is not concerned with a particular interpolation or optimization technique for reducing the number and size of RIR measurement points as in [9][8]. Therefore a simple demonstration of the wave-based approaches for facilitating a moving listener who has the freedom to make horizontal head rotations is provided here by incorporating the previously discussed method for recording a B-Format RIR. An objective evaluation of a RIR dataset is performed by the estimation of perceived source location using time-frequency domain directional analysis.

The directional analysis is performed using the technique introduced with SIRR/DirAC spatial audio rendering [17]. Briefly, the XY channels are converted to the time-frequency domain using a short time fourier transform (STFT). Active intensity describes the net flow of energy over a period of time and is calculated for each time-frequency element of XY and therefore used to indicate average direction of propagation/arrival. The results are visualized by overlaying a quiver plot on to a spectrogram of the W channel, indicating the directional properties of the audio signal.

4.1 Demonstration procedure and setup

Three demonstrations were conducted. Case 1 consisted of a stationary monopole sound source and three 1storder B-Format arrays in a row that represent the chosen walk-though path, from point A to C, of the listener. In case 2 the same stationary monopole source and walk-through path are taken however the central B-Format array is omitted. Instead this central array is linearly interpolated. The directional analysis is performed on the three recorded B-Format RIRs and the single interpolated B-Format RIR. The complete setup is depicted in Figure 5.

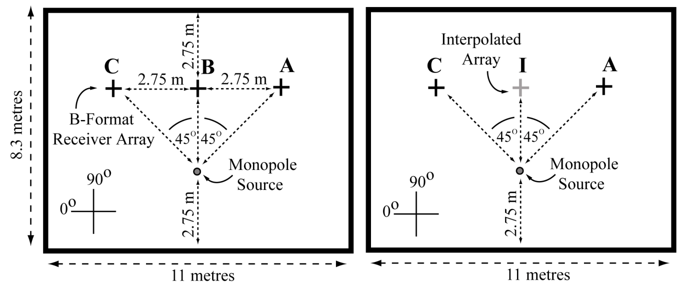

Figure 5: *The chosen setup geometry of the walk-through demonstration. Case 1, no interpolated B-Format RIRs and Case 2 a linearly interpolated B-Format RIR*

The final 3rdcase demonstrates the inherent support for head rotations that is available when using B-Format to capture the acoustically modelled soundfield. At point C the listener rotates their head by 45° such that they are pointing directly towards the source. This is implemented through a mathematical rotation applied to the RIR XY channels to simulate the head rotation according to (5) and (6) from [22]:

$$X' = X\cos\theta - Y\sin\theta \tag{5}$$

$$Y' = X\sin\theta + Y\cos\theta \tag{6}$$

where θ is the amount of positive rotation desired, while X' and Y' are the new X and Y channels. Therefore the apparent direction of the source should change by 45° which will be apparent from the directional analysis performed as for cases 1 and 2.

4.2 Results

The directional analysis results for array positions A, B and C in test case 1 are given in Figures 6, 7 and 8. The directional analysis for the interpolated array in case 2 is provided in Figure 9. Each of these four figures provide a SIRR/DirAC time-frequency analysis of the sound pressure at the identified locations in the form of a spectrogram. In addition, a quiver plot has been superimposed to indicate the average direction from which the sound originates for each time-frequency element, the 'x' marks the time-frequency element while the line indicates the direction.

Considering the walk-through cases from point A to point C as the listener looks straight ahead towards C, it is apparent in each case that the direct sound that arrives between 10ms and 20ms would be perceived as arriving from the correct direction, if the B-Format RIR was suitably decoded to a loudspeaker arrangement. However, it is also clear that the time of arrival of the direct sound is delayed in the interpolated case (Figure 9) as are all the following first reflections. In practice interpolation would not be performed linearly over such a relatively large distance, instead a higher resolution RIR dataset would be compiled therefore alleviating this issue, alternatively more rigorous interpolation techniques could be applied [9]. Despite this, the correct directional information can still be retrieved even from a simple interpolation of a B-Format RIR. Case 3 in Figure 10 is the directional analysis at point C for a listener that has rotated their head by 45° such that they are pointing directly towards the source. In Figure 7 the direct sound arrives from 225°, after the simulated head rotation in Figure 10 the apparent source direction is now 270° which indicates that the expected 45° head rotation results in the listener now effectively facing the sound source.

5. CONCLUSIONS

This paper has reviewed the methods and practices that have been employed for walk-through auralisations. Ray-based acoustic mod-

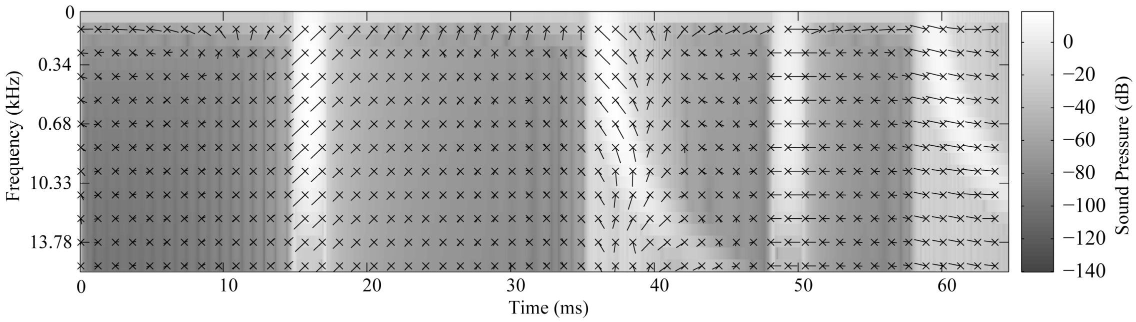

Figure 6: *Directional analysis for Array A in Cases 1 and 2*

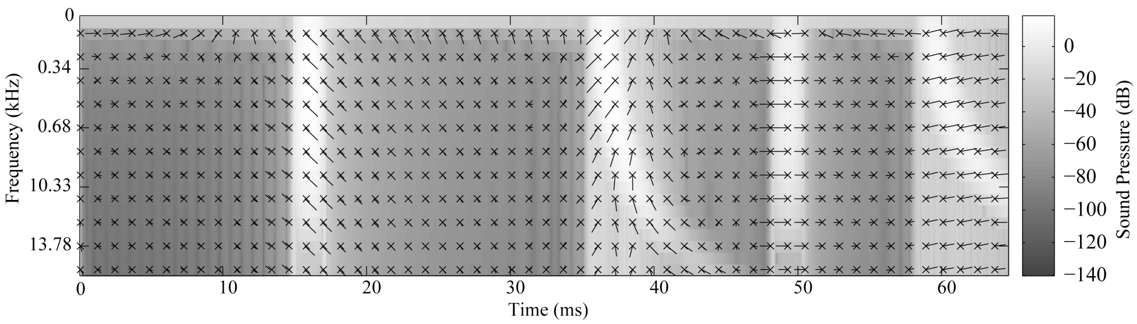

Figure 7: *Directional analysis for Array C in cases 1 and 2*

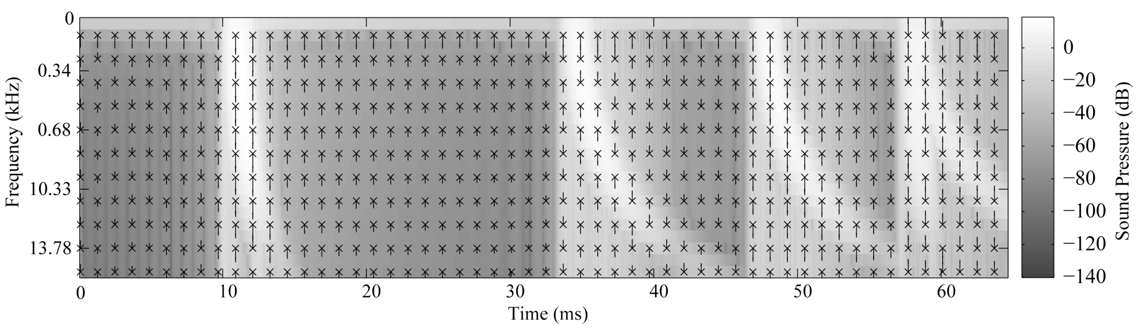

Figure 8: *Directional analysis for Array B in case 1*

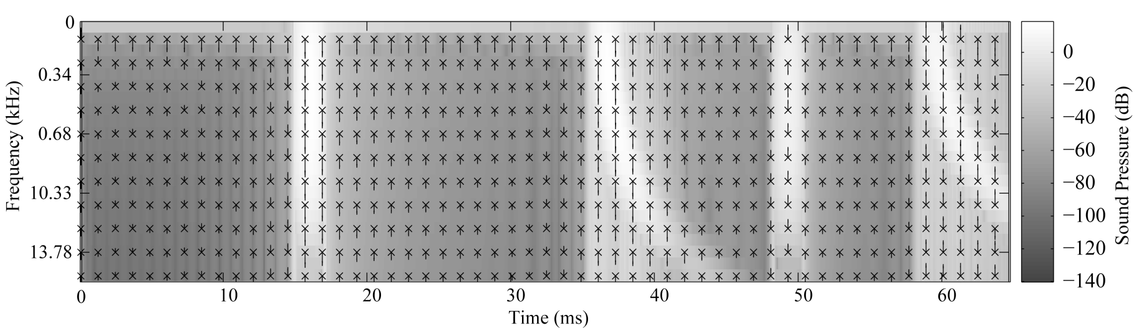

Figure 9: *Directional analysis for the interpolated array in case 2*

els have predominantly been used to physically relate what the listener sees via an appropriate visualisation scheme, to the soundfield they hear. In this paper a new method for facilitating walk-through auralisations is considered using a wave-based acoustical model. The FDTD/DWM schemes were briefly introduced and a benchmarking test compared a 2D and 3D wave based model to a typical ray-based model. Overall, if the complete RIR is desired then the wave-based approach is typically the most efficient option as the computation time does not increase when large receiver arrays are employed. A technique for encoding 1^{st}order B-Format from a wave-based model was also reviewed and used to demonstrate the suitability of a waved-based model for rendering walk-through au-

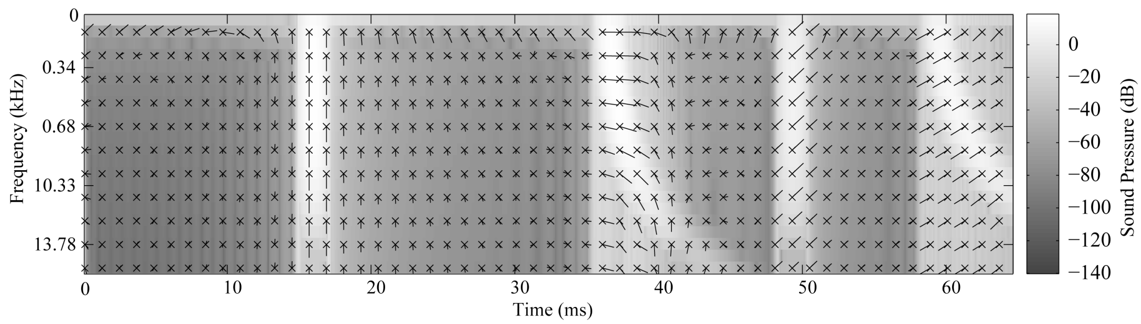

Figure 10: *Directional analysis for the rotated array in case 3*

ralisations that supported listener head rotations. A time-frequency directional analysis technique was employed to objectively indicate the location of a sound source using 3 encoded B-Format RIRs taken at positions along the listener's walk-through path. A comparison made between the centrally measured B-Format RIR and a linearly interpolated B-Format RIR show that even for this exaggerated case, the apparent location of the source was correct although time delayed. This shows that if a more dense set of RIRs is measured a much more accurate auralisation will be possible, but at no extra computational cost. In addition, if a reduced dataset size is required due to memory constraints at run-time, then more advanced interpolation techniques can be employed during the auralisation to alleviate this.

Overall this paper has outlined how wave-based acoustical models can be used as the basis for pre-computing a full RIR dataset suitable for real-time walk-through auralisation against more established ray based approaches.

6. ACKNOWLEDGMENTS

With thanks to EPSRC for their funding support and to Alastair Moore for aid with the quiver plot scaling.

REFERENCES

[1] D. Botteldooren, "Finite difference time domain simulation of low frequency room acoustic problems," *J. Acoust. Soc. of Am.*, vol. 98, no. 6, pp. 3302–3308, 1995.

[2] D. T. Murphy, A. Kelloniemi, J. Mullen, and Shelley S., "Acoustic modeling using the digital waveguide mesh," *IEEE Signal Processing Magazine*, vol. 24, no. 2, pp. 55–66, 2007.

[3] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Virtual environment simulation - advances in the diva project," in *International Conference on Auditory Display (ICAD'97), Palo Alto CA, USA*, 1997, pp. 43–46.

[4] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. of Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[5] S. Siltanen, T. Lokki, and L. Savioja, "Frequency domain acoustic radiance transfer for real-time auralization," 2009.

[6] B. Champagne, A. Lobo, and P. Kabal, "Efficient methods for simulating a moving talker in a rectangular room," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA*, 1991.

[7] B. Dalenbak and M. Stromberg, "Real-time walkthrough auralization - The first year," in *Institute of Acoustics*, 2006.

[8] R. Stewart and M. Sandler, "Generating a spatial average reverberation tail across multiple impulse responses," in *35th International Audio Engineering Society Conference, London, UK*, 2009.

[9] C. Masterson, G. Kearney, and F. Boland, "Acoustic impulse response interpolation for multichannel systems using dynamic time warping," in *35th International Audio Engineering Society Conference, London, UK*, 2009.

[10] A. Southern and D. Murphy, "A second order differential microphone technique for spatially encoding virtual room acoustics," in *Proc. of the AES 124th Convention: New Horizons in Audio, Amsterdam, The Netherlands*, May, 2008.

[11] J. O. Smith, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, no. 4, pp. 74–87, 1992.

[12] L. Savioja, T. J. Rinne, and T. Takala, "Simulation of room acoustics with a 3-D finite difference mesh," in *Proc. International Computer Music Conf., Denmark*, 1994, pp. 463–466.

[13] S. Shelley and D. T. Murphy, "The modeling of diffuse boundaries in the 2-D digital waveguide mesh," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 3, pp. 651–665, 2008.

[14] K. Kowalczyk and M. van Walstijn, "Modeling frequency-dependent boundaries as digital impedance filters in FDTD and K-DWM room acoustics simulations," *J. Audio. Eng. Soc.*, vol. 56, no. 7/8, pp. 569–583, 2008.

[15] M. A. Gerzon, "Periphony: With height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[16] A. McKeag and D. McGrath, "SoundField format to binaural decoder with head tracking," in *6th Australian Audio Engineering Society Convention, Melbourne*, 1996.

[17] V. Pulkki, "Directional audio coding in spatial sound reproduction and stereo upmixing," in *Proc. of the AES 128th International Conference: The future of surround and beyond, Piteå, Sweden*, July 2, 2006.

[18] A. Farina, "A-format to B-format conversion," Online at http://pcfarina.eng.unipr.it/Public/B-Format/A2B-conversion/A2B.htm, 2006.

[19] F. J. Fahy, *Sound Intensity*, Elsevier Applied Science, London, 1989.

[20] M. Beeson, A. Moore, D. Murphy, S. Shelley, and A. Southern, "Renderair - room acoustics simulation using a hybrid digital waveguide mesh approach," in *Proc. of the AES 124th Convention: New Horizons in Audio, Amsterdam, The Netherlands*, May, 2008.

[21] D. Murphy, M. Beeson, S Shelley, A. Moore, and A. Southern, "Hybrid room impulse response synthesis in digital waveguide mesh based room acoustics simulations," in *Proc. of the Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland*, September, 2008.

[22] D. Malham, "Spatial hearing mechanisms and sound reproduction," Online http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm, 2006.