Proceedings of the Institute of Acoustics

A MULTI-CHANNEL SPATIAL SIMULATION SYSTEM FOR COMPUTER MUSIC APPLICATIONS

Damian T Murphy and David M Howard
Department of Electronics, University of York, Heslington, York YO10 5DD

1. INTRODUCTION

Every sound we hear has associated with it an environmental context and a sense of location for the acoustic space within which it is heard. For hundreds of years composers have manipulated and used these properties of sounds in space as a fundamental part of their music. However, it is only in the relatively recent electronic age that composers and in particular computer musicians have sought complete control over this particular musical element. The processing of audio signals by simulating the acoustics of an enclosed space such as a room or hall is a technique widely used to enhance recorded and synthesized sounds in the computer music and recording industries. In particular, digital reverberation has been a standard post-synthesis sound enhancement tool for digital music creation since Schroeder's original paper in 1961 [1]. The acoustic properties of a real enclosed space can be uniquely defined by measuring the Room Impulse Response (RIR) at a specific listener location for an input signal applied at a given sound source location. The purpose of digital reverberation is to derive a digital signal processing operation that simulates the effect of a "good" room or hall when applied to a musical sound. Ideally, the impulse response of the final filtering scheme resembles the measured RIR. This is made difficult by the fact that typical listening spaces are inherently large order systems - every enclosed space calculates and outputs an infinite number of reflections as opposed to the finite limitations of the digital room.

Two methods commonly used for modelling a RIR based on a description of the room geometry, are ray tracing [2] and the image source method [3]. These methods have individual limitations although common to both is the fact that they are only valid for high frequencies. At low frequencies, or for small enclosed spaces, where the wave-like behaviour of sound and the effects of room resonances or modes are more noticeable these methods are less appropriate. These methods also require some expert knowledge on behalf of the user and are difficult to visualise. Often the computer musician will have an intuitive idea as to the type of acoustic space they require but not necessarily its particular acoustic properties. Therefore it would be desirable to have a method that enabled the user to investigate acoustic principles from simple concepts such as room size and shape, and provided visual as well as aural feedback as to how waves would propagate through this defined space.

This problem has been addressed in a number of ways, usually based around a direct time domain model of wave motion within the space, but these methods are often computationally intensive. Digital waveguide models - essentially an implementation of the Finite Difference Time Domain

(FDTD) method [4] - have been used successfully as a partial alternative to time domain models based on the wave equation in the fields of musical acoustics and sound synthesis. They have also been used to model the low frequency behaviour of an enclosed listening space and are capable of providing visual feedback to the user [5]. Current digital waveguide models of room acoustics are limited to providing only a partial solution to an accurate RIR, as they are less effective at high frequencies where execution time becomes prohibitive for a full 3-D implementation.

This paper introduces and describes the ongoing implementation of the *WaveVerb* Multi-channel Spatial Simulation System which uses a digital waveguide model as an alternative method of generating a multi-channel RIR valid for both low and high frequencies. This will be accomplished by limiting the model to the horizontal plane only. Results are presented that validate the current model for low frequencies, and demonstrate how audio can be processed using multi-channel RIRs for presentation to a listener in a multi-speaker surround-sound environment.

2. THE 2-D DIGITAL WAVEGUIDE MESH

A waveguide is any medium in which wave motion can be characterised by the one dimensional wave equation. In the lossless case, all solutions can be expressed in terms of left-going and right-going travelling waves and can be simulated using a simple bi-directional digital delay line. A digital waveguide model is obtained by sampling, both in space and time, the one-directional travelling waves which occur in a system of ideal lossless waveguides. The sampling points in this case are called scattering junctions, and are connected by bi-directional unit-delay digital waveguides. The theory behind digital waveguides is well covered in [5] and [6] but is included here as further background information. Figure 1 shows the general case of a scattering junction J with N neighbours, $i = 1,2,\ldots N$.

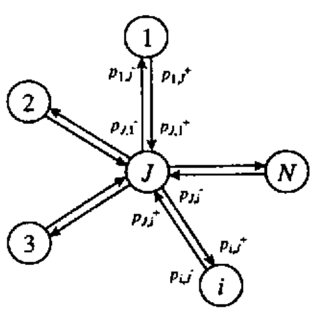

Figure 1: Scattering junction J with N neighbours connected via bi-directional unit delay elements.

The sound pressure in a waveguide is represented by p_i, the volume velocity by v_i and the impedance of the waveguide by Z_i. The input to a waveguide is termed p_i^+ and the output p_i^-. The signal $p_{i,J}^{+}$ therefore represents the incoming signal to junction i along the waveguide from the opposite junction J. Similarly, the signal $p_{i,J}^{-}$ represents the outgoing signal from junction i along the waveguide to the opposite junction J. The volume velocity v_i is equal to pressure, p_i, divided by impedance, Z_i. The delay elements are bi-directional and so the sound pressure is defined as the sum of its input and output:

$$(1) \qquad p_i = p_i^{+} + p_i^{-}$$

At a lossless scattering junction with N connected waveguides the following conditions must hold:

1. $\sum_{i=1}^{N} v_i^{+} = \sum_{i=1}^{N} v_i^{-}$ (Flows add to zero)
2. $p_1 = p_2 = ... = p_i = ... = p_N$ (Continuity of impedances)

Using these conditions the sound pressure at a scattering junction can be expressed as:

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

As the waveguides are equivalent to bi-directional unit-delay lines, the input to a scattering junction is equal to the output from a neighbouring junction into the connecting waveguide at the previous time step. This can be expressed as:

$$(3) \qquad p_{J,i}^{+} = z^{-1} p_{i,J}^{-}$$

To model the propagation of a wave on the horizontal plane within an enclosed space, a 2-D rectilinear mesh is constructed using unit delay waveguides and lossless scattering junctions according to equations (1), (2) and (3) with N=4. Each junction will therefore have four similar neighbours each connected by a unit waveguide. Boundary nodes only have one neighbour. A signal representing acoustic pressure introduced to a waveguide will propagate in either direction along the bi-directional delay lines until it comes to a junction. The signal then scatters according to the relative impedances of the connected waveguides as determined by equations (1) and (2). In the current model all impedances are equal. After the pressure at each junction has been calculated, together with the associated incoming and outgoing pressures for each connected waveguide, the signals are passed along from junction to junction according to equation (3).

3. SPATIALLY ENCODING THE RIRs

It is possible to encode a soundfield by decomposing it into spherical harmonic components [7]. The zero order pressure component is termed W and is omnidirectional, picking up all sounds from all directions equally. The first order velocity components are figure-of-eight responses pointing forward, left and up, and are termed X, Y and Z respectively. The four signals, W, X, Y and Z are known as B-format. Reproducing a B-format signal is possible using an appropriate Ambisonic decoding scheme and a multi-speaker array [8]. It is possible to derive horizontal only B-format signals - W, X and Y only - from the mesh at any point resulting in a B-format multi-channel RIR. Using this encoding method it is possible to recreate the soundfield in the modelled room complete with acoustic cues associated with the location of a sound source in relation to the listener. This is in addition to the cues relating to source distance and environmental context that are already present in a single channel RIR.

4. IMPLEMENTATION - THE *WAVEVERB* SYSTEM

Two versions of the waveguide mesh have currently been implemented. The first enables the user to observe wave propagation on a graphical display with a number of different views and realisations. This provides excellent visual feedback on the behaviour of the propagating wave. This module forms the basis of the *WaveVerb* system and allows the user to make RIR measurements for various source-listener positions. User control is given over variables such as room width, length and mesh density. In this way the user does not require any in depth knowledge of room acoustics in order to generate a RIR for a reverberant effect - the parameters the user has control over correlate readily to real world variables such as room size, source location and object position. The second implementation runs without graphical feedback or user interaction, allowing much higher mesh densities to be used.

4.1 The Wave Propagation Module

Figure 2 is a screen shot from *WaveVerb* showing a wave propagating along the mesh for a simple rectangular room. Notice the phase reversed reflections from the walls.

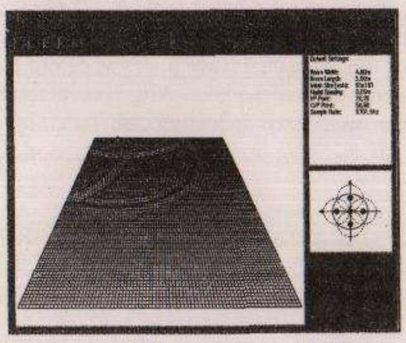

Figure 2: WaveVerb Main Window - Wave propagation along the mesh with phase reversing reflections at the walls.

Figure 3 is a close up shot of a similar rectangular room. This shows a wave propagating along the mesh and passing through gaps in a dividing wall. Notice how the waves spread out into the second half of the room showing that the model is capable of demonstrating diffraction effects. Clearly from Figure 3 if a listener was situated immediately behind one of the dividing wall partitions the direct sound from a source in the other half of the room would reach them. This is due to the curved leading edge of the diffracted wave that has passed through the gaps in the wall. If this room were modelled using a basic ray tracing algorithm, as there is no direct line-of-sight path between the source and the listener, the direct sound would not be recorded until the sound "particles" or "rays" being tracked had been reflected off one or more surfaces. This would lead to an inaccurate RIR measurement.

Figure 3: Diffraction effects due to gaps in a wall placed in the modelled room.

4.2 The Analysis Module

Figure 5 shows part of *WaveVerb*'s analysis module where RIRs can be displayed and examined in both the time domain and the frequency domain. Conversion to and from various file formats including AIFF soundfiles can also be performed. The example in figure 5 shows a measured RIR in the frequency domain up to the approximate critical frequency of a rectangular room of length 6.6m and width 5.5m. Below this critical frequency the behaviour of the room's modes is the dominant acoustic factor. A smooth impulse was applied at one corner of the room and a 2 second RIR was measured at the opposite corner with a mesh density - corresponding to the inter-junction spacing - of 0.011m. Boundary conditions were set to give phase preserving reflections with no absorption. The RIR has the first 0.2s removed to eliminate any transient response, and the remaining 1.8s has an FFT applied to it. The analytical room modes have been calculated and plotted on the same graph. In this case only axial (reflections between two surfaces) and tangential (reflections between four surfaces) are valid. The figure shows that in this low frequency region there is an exact correlation between the analytical room modes and the resonant frequencies highlighted by examining the frequency response of the measured RIR. It is assumed in both ray tracing and image source models that the wavelength corresponding to the lowest frequency of the sound is small compared to the linear dimensions of the room. If a room is large, the results may be valid for frequencies above 100Hz [2]. For smaller rooms, as in this example, the lower bound on the valid frequency range would be much higher.

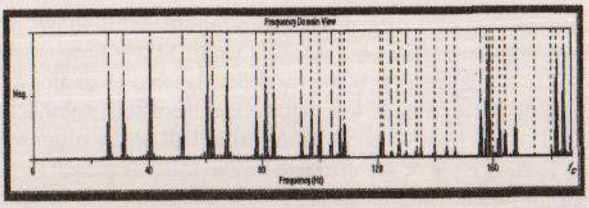

Figure 5: Frequency response of modelled room below approximate critical frequency, f_c. Analytical axial modes are shown as long dashed lines. Analytical tangential modes are shown as short dashed lines.

4.3 RIR Measurements

Figure 6 shows a model of two differently shaped rooms separated by a wall with an open doorway.

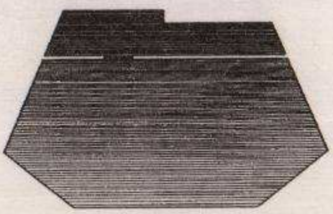

Figure 6: Two adjoining rooms separated by a wall with an open doorway.

Figure 7 shows RIR measurements taken from this modelled room. Figure 7(a) is the RIR measured with both source and receiver placed in the larger of the two rooms. Figure 7(b) is the RIR measured with the source in the larger room and the receiver behind the wall in the smaller room. The north and south walls have been set to be more absorptive than the east and west walls.

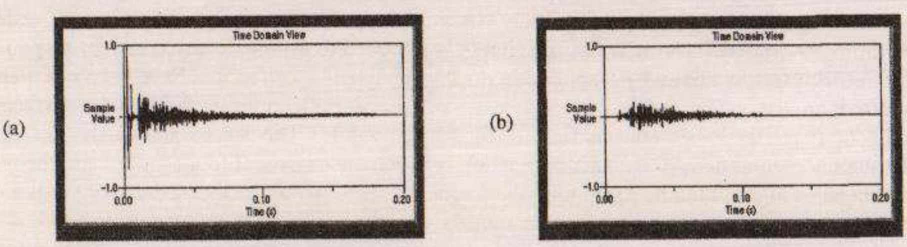

Figure 7: RIR measurements taken from the modelled room as shown in Figure 6; 7(a) - source and receiver in larger room; 7(b) - source located in larger room with receiver behind the wall in the smaller room.

Note that as with real RIRs there is an initial time delay before the sound reaches the listener, a single large impulse representing the *direct sound,* and a number of clearly discrete impulses generated by *early reflections* from the walls in the room followed by the slowly decaying *reverberant* sound.

4.4 The Decode Module

Using the decoding module as shown in Figure 8, anechoic or synthesized sound can be processed with a B-format RIR for playback over a multi-speaker array. Further processing using a set of measured Head Related Transfer Functions (HRTFs) allows binaural surround sound monitoring over headphones removing the necessity for a large multi-speaker array.

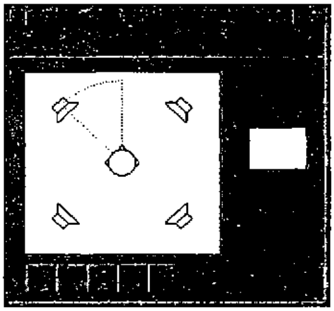

Figure 8: The WaveVerb Decoding Module

4.5 Mesh Sampling Rate

The sampling rate of the mesh, which is determined by the spacing of the scattering junctions and the mesh topology, is given by $f_{update} = c\sqrt{2}/d$, where c is the speed of sound, and d is the distance between scattering junctions. The effective sampling rate, and so the wave of highest frequency that can be propagated successfully by the mesh, is well below the value given by f_{update}. An inherent problem with lattice type structures such as the digital waveguide mesh is that they are *dispersive*. That is, the velocity of a wave will vary according to its frequency. For a 2-D rectilinear mesh there is no dispersion at any frequency when travelling diagonally to the mesh coordinate system, but high frequency signals are delayed along the directions of the waveguides themselves. This limitation can be improved upon by using a triangular based mesh consisting of 6-port scattering junctions [9]. This mesh topology has as yet only been used for modelling 2-D membranes [10] and not for room acoustic problems. Work is currently ongoing in the development of this basic rectilinear model into a 6-port triangular mesh resulting in RIRs with a greatly improved high frequency response due to the minimisation of dispersion effects. Initial results are promising and show that this appears to be the case.

5. CONCLUSIONS AND FURTHER WORK

A method of generating multi-channel reverberant effects by modelling acoustic wave propagation in an enclosed space using a 2-D digital waveguide mesh has been presented. Initial results from the *WaveVerb* system show that diffraction is clearly visible and that the modal behaviour of a theoretical 2-D room is closely approximated at low frequencies. Audio samples, when convolved with a B-format RIR obtained from the mesh, can be presented to a listener over a multi-speaker array. The effect of the B-format RIR is to add a reverberant quality to the anechoic or synthesized audio complete with appropriate audio cues for sound source localisation, as if the sound had originated from within the modelled space. Possible extensions to this basic model include a detailed study of the 6-port mesh, a full 3-D implementation, and user defined boundary conditions that relate to real world materials.

6. REFERENCES

[1] M. R. Schroeder and B. F. Logan, "'Colorless' Artificial Reverberation", *Journal of the Audio Engineering Society*, vol. 9, no. 3, pp 192-197, 1961.

[2] A. Krokstad, S. Strøm and S. Sørsdal, "Calculating the Acoustical Room Response by use of a Ray Tracing Technique", *Journal of Sound Vibration*, vol. 8, no. 1, pp 118-125, 1968.

[3] U. Stephenson, "Comparison of the Mirror Image Source Method and the Sound Particle Simulation Method", *Applied Acoustics*, vol. 29, pp 35-72, 1990.

[4] S. A. Van Duyne and J. O. Smith III, "Physical Modelling with the 2-D Digital Waveguide Mesh", *Proc. International Computer Music Conference (ICMC) 1993*, pp 40-47, 1993.

[5] L. Savioja, T. J. Rinne and T. Takala, "Simulation of Room Acoustics with a 3-D Finite Difference Mesh", *Proc. ICMC 1994*, pp 463-466, (1994).

[6] L. Savioja, M. Karjalainen and T. Takala, "DSP Formulation of a Finite Difference Method for Room Acoustics Simulation", *Proc. NORSIG'96 IEEE Nordic Signal Processing Symposium*, Espoo, Finland, Sept 1996.

[7] J. S. Bamford and J. Vanderkooy, "Ambisonic Sound for Us", *Presented at the 99th Audio Engineering Society Convention, October 1995*, Preprint 4138 (O-4), 1995.

[8] M. A. Gerzon, "General Metatheory of Auditory Localisation", *Presented at the 92nd Audio Engineering Society Convention, March 1992*, Preprint 3306, 1992.

[9] S. A. Van Duyne and J. O. Smith III, "The 3D Tetrahedral Digital Waveguide Mesh with Musical Applications", *Proc. ICMC 1996*, pp 9-16, 1996.

[10] D. Rocchesso and F. Fontana, "A New Formulation of the 2D-Waveguide Mesh for Percussion Instruments", *Colloquium on Musical Informatics*, Bologna, November 1995.