Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

The prediction of synthesised wavefields within realistic room acoustics scenarios

I. A. Drumm (1), R. G. Oldfield (1)

(1) The University of Salford, UK

PACS: 43.55.KA 43.38.VK

ABSTRACT

This paper presents a technique to predict the output from an in-house developed 128 channel wavefield synthesis system deployed within realistic room acoustic scenarios. Based on Finite Difference Time Domain the acoustic prediction detailed here utilises ray based voxellisation and a least pth norm filter design approach to import hall data in popular formats that specify complex room geometries and the frequency dependant absorption profiles of surface materials. Using this approach the objective and subjective fidelity of wavefield synthesis within realistic deployment scenarios can be better understood.

INTRODUCTION

BACKGROUND

Wavefield Synthesis

Originally developed by Berhout [1], Wavefield synthesis uses loudspeaker arrays to recreate the wavefronts of multiple arbitrarily positioned virtual sources thus emulating the distribution of sound waves in a volume as would be experienced by a user at any given position within that volume. Notably wavefield synthesis can auralise virtual sources both outside and within the confines of a speaker array and doesn't tie users to a sweet spot. It is worth at this point briefly considering Wavefield Synthesis theory as this will have a bearing on the scope of the later prediction technique. According to Huygen's Principle such wave-fronts can also be created by the cumulative effect of secondary sources just behind the wave-front.

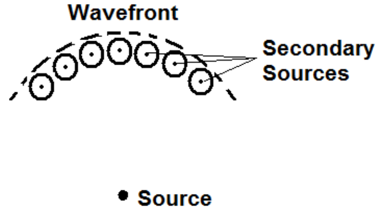

Figure 1. The generation of a wavefront with secondary sources

So in theory a set of nearby sources can be used to emulate the wave-front of a far away source. Consider a notional plane surface S within a volume where $\mathbf{r}_v$ is a point within the volume, $\mathbf{r}_S$ is a point on the plane and $\mathbf{n}$ is a normal to the plane. If a sound source was placed at $\mathbf{r}_S$ with known pressure and particle velocity we could predict the magnitude and phase of sound pressure at $\mathbf{r}_v$.

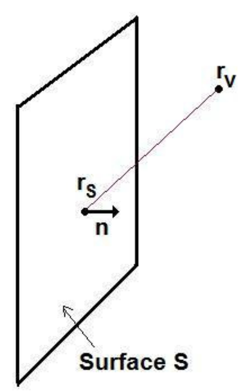

Figure 2. Considering a cross-section of a sound field as a surface

Similarly if the plane consisted on an infinite number of mono-pole and dipole sources then the pressure at the receiver point in the volume can be found by adding secondary source contributions using the Kirchhoff-Helmholtz integral.

$$p(\mathbf{r}_v) = \int_S p(\mathbf{r}_S)\frac{\partial G(\mathbf{r}_S, \mathbf{r}_v)}{\partial n(\mathbf{r}_S)} - G(\mathbf{r}_S)\frac{\partial p(\mathbf{r}_S)}{\partial n(\mathbf{r}_S)}ds \tag{1}$$

Where $p(\mathbf{r}_v)$ is pressure at a point $\mathbf{r}_v$ in volume, $p(\mathbf{r}_S)$ is pressure at a point $\mathbf{r}_S$ on the surface, $\mathbf{n}(\mathbf{r}_S)$ is the unit normal of a point on surface and G is a complex function of source distance $r = |\mathbf{r}_v - \mathbf{r}_S|$ and wave number k.

$$G = \frac{e^{-jkr}}{4\pi r} \tag{2}$$

Referring to the above we can in theory synthesise the sound at all points within a volume if we surround it with a surface of infinite mono-pole and dipoles sources. The planes themselves represent the cross-sections of a desired sound field. If we can predict the pressures and particle velocities through these cross-sections we can render the entire volume. By considering propagation into half space equation 1 can be simplified to Rayleigh integral I or II that indicate wavefields to be constructed in terms of either just monopoles or dipoles [1]. Practical limitations dictate the use of finite line arrays of a discrete number quasi mono-pole sources (i.e. loudspeakers). By considering the integrals of vertical contributions in the plane as coming from a horizontal line hence by applying a stationary phase approximation, wavefields can be constructed using linear arrays. Hence equation 1 becomes

$$p(\mathbf{r}_v) = \sum_{n=1}^{N} D_n \frac{e^{-jk|\mathbf{r}_v - \mathbf{r}_n|}}{|\mathbf{r}_v - \mathbf{r}_n|} \delta x \tag{3}$$

where N is the number of loudspeakers, each positioned at $\mathbf{r}_n$ and separated by the distance δx. D_n is the spectrum of the loudspeaker's driving function.

In practice virtual sources are positioned within the world space coordinate system and their contributions to the respective driving functions evaluated in the time domain. By considering a virtual source located at $\mathbf{r}_{vs}$ with source signal vs the loudspeakers driving function is

$$D \propto vs(\omega)\sqrt{\frac{jk}{2\pi}}\, f\left(|\mathbf{r}_n - \mathbf{r}_{vs}|\right) \tag{5}$$

Rendering the time domain requires appropriately delayed and attenuated version of the source signal contribution. From equation 5 the source signal should also be high pass filtered to compensate for cylindrical wavefronts from the line arrays.

Given the use of finite discretely located sources in none anechoic environments there are a number of rendering pathologies including spatial aliasing, array truncation errors and secondary wave fronts from room reflection (as well as diffusion and modes). A key challenge of current research it to better understand and mitigate for the perceptual effect of rendering pathologies, notably their effect on source localisation and user presence.

The authors have developed two 128 channel wavefield systems and a corresponding C++/WASAPI based API and C# wrapper classes. Both systems use one dimensional arrays of dual concentric loudspeakers at head height with typically 12.5cm loudspeaker separations.

Figure 3. A prototype 128 channel Wave-field synthesis system at the University of Salford.

Beyond initial development in our anechoic chamber our systems will be deployed within environments that are non anechoic (e.g. an eight screen cave, a listening room and a large multipurpose hall).

Finite Difference Time Domain

Originally developed as a computational electrodynamics modelling technique [2], FDTD has found popularity in modelling broadband sound propagation phenomena [3] including frequency dependant absorption, diffraction and interference. The acoustic implementation of FDTD considers sound propagation as coupled equations 5 and 6.

$$\frac{1}{K}\frac{\partial p}{\partial t} = -\nabla . w \tag{6}$$

$$\rho \frac{\partial w}{\partial t} = -\nabla p \tag{7}$$

where p is pressure, w is particle velocity, ρ is the density of medium and K the bulk modulus. Writing equations 6 and 7 in discrete centre difference forms gives equations 8, 9, 10 and 11 where i, j and k represent grid locations and n represents time steps.

$$w_x^{n+1}\left(i+\tfrac{1}{2}, j, k\right) = w_x^{n}\left(i+\tfrac{1}{2}, j, k\right) - \frac{\delta t}{\rho \delta x}\left\{p^{n+\frac{1}{2}}(i+1, j, k) - p^{n+\frac{1}{2}}(i, j, k)\right\} \tag{8}$$

$$w_y^{n+1}\left(i, j+\tfrac{1}{2}, k\right) = w_y^{n}\left(i, j+\tfrac{1}{2}, k\right) - \frac{\delta t}{\rho \delta y}\left\{p^{n+\frac{1}{2}}(i, j+1, k) - p^{n+\frac{1}{2}}(i, j, k)\right\} \tag{9}$$

$$w_z^{n+1}\left(i, j, k+\tfrac{1}{2}\right) = w_z^{n}\left(i, j, k+\tfrac{1}{2}\right) - \frac{\delta t}{\rho \delta z}\left\{p^{n+\frac{1}{2}}(i, j, k+1) - p^{n+\frac{1}{2}}(i, j, k)\right\} \tag{10}$$

$$\begin{aligned} p^{n+\frac{1}{2}}(i, j, k) = p^{n-\frac{1}{2}}(i, j, k) - K\delta t \Bigg\{ & \frac{w_x^n\left(i+\frac{1}{2}, j, k\right) - w_x^n\left(i-\frac{1}{2}, j, k\right)}{\delta x} \\ & + \frac{w_y^n\left(i, j+\frac{1}{2}, k\right) - w_y^n\left(i, j-\frac{1}{2}, k\right)}{\delta y} \\ & + \frac{w_z^n\left(i, j, k+\frac{1}{2}\right) - w_z^n\left(i, j, k-\frac{1}{2}\right)}{\delta z} \Bigg\} \end{aligned} \tag{11}$$

Hence the medium is effectively represented as interleaved 3D grids of pressure and velocities respectively with neighbouring pressure and velocity nodes separated by a half step. To avoid reflections at grid terminations, perfectly matched layers (PMLs) [4] are implemented to introduce gradual absorption whilst keeping the impedance constant. By adding or setting pressures within specified grid positions various source types can be emulated.

Voxelisation of hall data

Three dimension hall data with specified material properties is parsed to extract planes and their frequency dependant absorptions, which are hence used to set corresponding elements within the 3D FDTD grid. The primary author's software parses ODEON, CATT and FBX file formats. FBX is a popular export format for common 3D file creation software and an important format for XNA / Managed C# which in conjunction with msbuild provides a means to include complex 3D models of people, furniture, etc within room models. Although there are some fast voxelisation techniques documented in computer graphics literature the author uses a simple ray based approach [5][6]. Essentially each plane can be described by the equation

$$D = \hat{\mathbf{n}} \cdot \mathbf{p} \tag{12}$$

where p is a point the plane, D is a constant scalar and $\hat{\mathbf{n}}$ is the plane's unit normal found using the vector cross product of plane edges. A succession of rays

$$\mathbf{r} = \mathbf{s} + \mathbf{c}t \tag{13}$$

are created along all axes where s is the start vector position, c is the direction vector, and *t* a scalar distance. Ray / plane intersects given are by

$$t = \frac{D - \hat{\mathbf{n}} \cdot \mathbf{s}}{\hat{\mathbf{n}} \cdot \mathbf{c}} \tag{14}$$

and used with an 'in boundary' condition that tests the sense consistency of the cross products formed between the intersection point and successful plane vertices. For each valid intersection the corresponding FDTD grid position object is set to index the material properties of the surface.

Emulating frequency dependant absorption

The implemented approach to emulating frequency dependant absorption utilises octave banded absorption coefficients of the voxellised hall data to specify the magnitude response of 4th order IIR filters. Detailed in [5] the technique finds the filter's coefficients $h_0, a_1, b_1, a_2, b_2\ a_3$, etc via a Newton based optimisation technique. An assumption was made that a velocity component at a surface element can be given as a function of the history of normal velocity components incident and the history of previous outputs from this function, with for example…

$$w = f_n\left(n_x \cdot w_x^n, n_x \cdot w_x^{n-1}, n_x \cdot w_x^{n-2} \ldots, f_{n-1}(\ldots), f_{n-2}(\ldots), ..\right) \tag{15}$$

Where w_x^n is given by equation 8 and n_x is the x component of the surface's absolute unit normal. The output velocity becomes an input argument to equation 11. The equation 16 is in effect the difference equation of the element's IIR filter which could written as

$$w = \left[h_0 + \sum_{i=1}^{N} b_i w_{n-i} - \sum_{j=1}^{N} a_j f_{n-j}\right] \tag{16}$$

By considering the magnitude response

$$H_d(\omega) = 1 - \sqrt{1 - \alpha(\omega)}$$

from octave banded absorption energy coefficients $\alpha(\omega)$

$$\omega = {}^{125}\!/_{2\pi}, {}^{250}\!/_{2\pi}, \ldots, {}^{4000}\!/_{2\pi}$$

a least pth norm based filter design approach was employed to find the desired coefficients. The approach employ's Newton method in optimisation to find from the minimised the Lp-norm objective function

$$E(\mathbf{x}) = \int \left|H(\omega) - H_d(\omega)\right|^2 d\omega \tag{17}$$

where frequency response

$$H(\omega) = h_0 \frac{B}{A} \tag{18}$$

$$B = 1 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_N z^{-n}$$

$$A = 1 + a_1 z^{-1} + b_2 z^{-2} + \ldots + a_N z^{-n}$$

Writing the desired filter coefficients in equation 18 as the vector

$$\mathrm{x} = \left[h_0, a_1, a_2 \ldots a_N, b_1, b_2 \ldots b_N\right]^T \tag{19}$$

an iterative scheme (equation 20) can be applied to minimise $E(\mathbf{x})$ and hence give filter coefficients approximating towards the desired response.

$$\mathbf{x_{k+1}} = \mathbf{x_k} + \alpha\left(\nabla^2 E + \beta I\right)^{-1} \nabla E \tag{20}$$

Here integer $\mathbf{k}$ denotes the iteration, ∇E and $\nabla^2 E$ matrices of functions of the rates of change of $H(\omega)$, α and β are constants that control rate of descent towards minimum and I is an identity matrix.

EMULATING WAVEFIELD SYNTHESIS SYSTEMS

Although expensive in terms of memory the FDTD technique provides a particularly useful way to emulate wavefields generated by loudspeakers and their detection with multiple microphones. The approach described here has the added capability of modelling the pathological effects of 3D room acoustics with complex geometries and frequency dependant absorption at surfaces. Further the approach facilitates the inclusion of scattering objects (people, furniture, etc) within the hall space.

The WFS emulation specifies

a) arrays of loudspeaker objects with respective x,y,z positions and direction vectors. These are the FDTD sources that set pressures within the grid for each iterative cycle. Typically they'll consist of 64 to 128 sources in line arrays forming a rectangular or octagonal ring.

b) one or more virtual source objects with respective x,y,z positions that may or may not be located within the grid.

c) a matrix of virtual sources / loudspeaker link objects that use relative positions to specify the appropriate frequency dependant amplitudes and phases which in turn control how and when a virtual source's driving function is manifest at each respective emulated loudspeaker.

Note for virtual sources outside the ring only emulated loudspeakers that face away from them will have link object active thus avoiding spurious wavefronts in the listening space that propagate in a direction opposite to the virtual source wavefronts.

For virtual sources within the loudspeaker ring, phases become a function of the ring's dimensions and are effectively reversed in time to render 'focussed sources'. All or selected loudspeakers can be used to emulate focussed sources depending on listener's position. A truly listener agnostic rendering of focused sources can cause the listener to experience concave wavefronts coming from a direction opposite to the desired direction, i.e. that of wavefronts emanating from the focused source.

A crude but effective way to emulate the binaural listener experience of virtual sources is to incorporate a voxellised head and body model within the ring and set the virtual microphones at corresponding ear positions. A Gaussian pulse was used for the virtual source driving function and the resultant impulse responses at virtual left and right ear positions were convolved with anechoically recorded speech to create stereo channels suitable for headphones. Note being able to import and voxellise complex and detailed objects from FBX files has made this approach particularly effect. Laser scanning of real heads and pinnar for incorporation within the software is also possible and described in [6].

Of particular interest to the authors of this paper is how pathological cues effect the perception of virtual sources generated by wavefield synthesis, such cues include

- Spatial aliasing from the separation of discretely located loudspeakers.
- The use of linear as opposed to planar arrays and how the choice of height position and convolution used may effect auralisation.
- The effects of array truncation and choice of windowing to mitigate for this.
- The effects of pre-echoes (i.e. wavefronts reaching the listener before the desired wavefronts emanating from virtual source position). This pertains to focused sources.
- The effect of reflection from walls and other scatterers that form the mediating environment in which wavefield synthesis is deployed.

PRELIMINARY RESULTS

A comprehensive and detailed assessment of the objective and subjective validity of the emulation approach and insights yielded will be presented in a journal publication at a later date. Here is just presented a demonstration of the technique applied for the 3D rendering focussed sources, with and without scattering surfaces.

Auralisations were generated for multiple source positions, each with

1. a conventionally located FDTD source within the ring,
2. a focused source whose respective link matrix object was inactive if the listener was between loudspeaker and virtual source,
3. a focussed source whose generation was listener agnostic.

Figures 4, 5 and 6 shows simulation snap shots during the recording of left and right ear impulse responses from virtual source sources (a Gaussian pulse) left/forward of the listener. For both cases 1 and 2 with the chosen virtual listener position the simulation generated stereo files whose source directions were easily localised for virtual source positions at 0, 30, 45, 60 and 90 degrees left (or right) of the listener. Case 3 gave very poor localisation due mainly to strong behind the ear cues that contribute to the formation of the focussed source. Interestingly weaker but detectible spurious cues also exist for case 2 those these appear to be perceptually rejected in favour of the desired cues for focus source position. This concurs with real experiments using the WFS rig shown in

Figure 3. Clearly the experience of focussed sources will be dependant on the listener's position and forward direction relative to the loudspeaker arrays. A full assessment of focussed sources for different listener positions is a goal of the authors. Also the separation, scattering and truncation of loudspeakers influence source perception; the physical artifacts from these are clearly visible in the FDTD simulation.
Absolute distance localisation for all sources generated within the FDTD wave tank was relative poor owing (in part) to the use of bandwidth limited Gaussian pulses and the limited sample rate (8kHz to 16kHz) of memory expensive FDTD grids. FDTD also introduces errors such as numerical dispersion, anisotropic propagation, potential aliasing and artefacts from the staircase approximations of surfaces.

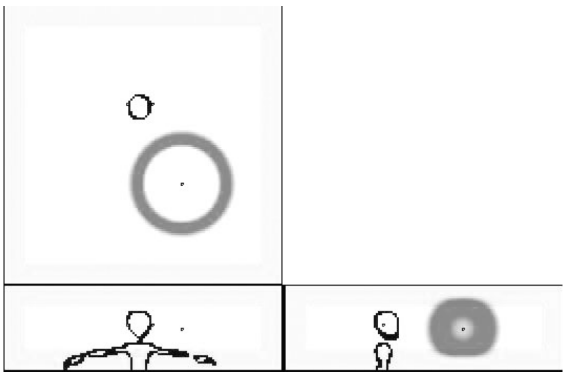

Figure 4. Simulating a simple source with Gaussian pulse. The voxelised figure has detectors at left and right ear positions.

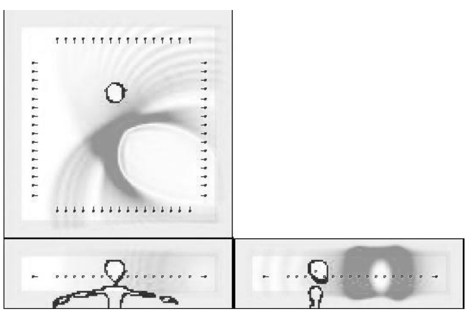

Figure 5. Simulating a focused source with selected source/loudspeaker links active

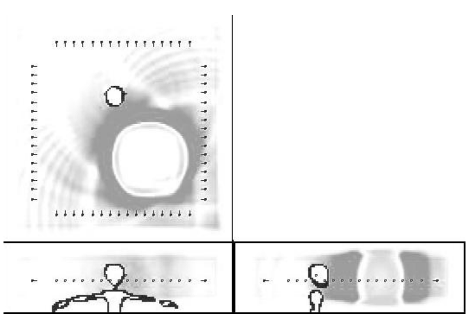

Figure 6. Simulating a focused source with all source/loudspeaker links active

Given the wave tank can import hall data and emulate frequency dependant sound propagation in architectural spaces and aim of the work is to predict and assess the robustness of wfs in application scenarios.

Figure's 7 shows the emulation of focused sources in rooms with acoustically treated walls. Figure 8 shows the emulation of a 128 channel WFS rig situated about eight projector screens thus forming part of a large cave like environment.
Early results show promisingly realistic auralisations of sources and the perceived localisations. Future work will include a comprehensive assessment of the validity of the approach and how it can contribute to better understanding of WFS and how the pathological effects of early reflections, reverberance and room modes can be mitigated for.

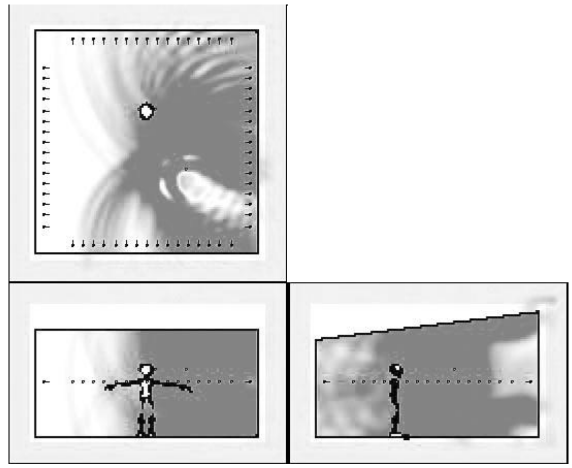

Figure 7. Simulating a focused source with WFS in room environment.

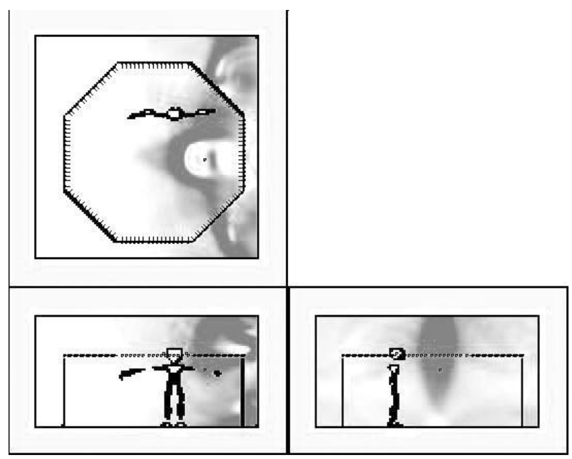

Figure 8. Simulating a focused source with WFS in room environment with eight projection screens.

REFERENCES

1 Berkhout, A. J.; de Vries, D.; Vogel, P: "Acoustic Control by Wave Field Synthesis," J. Acoust. Soc. Am., Vol. 93, pp. 2764–2778, 1993.

2 K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas Propagat. **14**, 302–307 (1966).

3 J.G. Meloney and K.E. Cummings, "Adaption of FDTD techniques to acoustic modelling",11th Annu. Rev. Prog. Applied Computational Electromagnetics, Vol.2, 724 (1995)

4 J.P. Berenger, “A perfectly matched layer for the absorption of electromagnetic waves’. J. Computational Physics, vol 114, 185-200, 1994.

5 I. A. Drumm, Y. W. Lam, 'Development and assessment of a finite difference time domain room acoustic prediction model that uses hall data in popular formats', Proceedings, Internoise-Turkey, (2007)

6 I. A. Drumm, 'Importing arbitrary complex objects with a FDTD based prediction application' in: IOA Spring Conference 2008, 10-11 April 2008, Reading, UK.

7 R.G. Oldfield, I.A. Drumm and J. Hirst, The Perception of Focused Sources in Wave Field Synthesis as a Function of Listener Angle, 128th Conv. Audio Eng. Soc, London, UK, 2010