A COMPARATIVE STUDY OF PERCEPTIONAL QUALITY BETWEEN WAVEFIELD SYNTHESIS AND MULTIPOLE–MATCHED RENDERING FOR SPATIAL AUDIO

J. Hannemann, C.A. Leedy, and K.D. Donohue

University of Kentucky
Center for Visualization and Virtual Environments
Lexington, KY, USA

S. Spors and A. Raake

Deutsche Telekom Laboratories
Berlin, Germany

ABSTRACT

This paper introduces a new algorithm to render virtual sound sources with spatial properties in immersive environments. The algorithm, referred to as Multipole–Matched Rendering, uses the Method–of–Moments and Singular–Value Decomposition to optimally match a spherical–multipole expansion of the virtual source to the field resulting from a spatially distributed speaker set. The flexibility of this method over other approaches, such as Wavefield Synthesis, allows for complex speaker geometries, and requires a smaller number of speakers to achieve a similar spatial rendering performance for listeners in immersive environments. The trade–off for the enhanced performance is a smaller area of faithful reproduction. This limited area, however, can be focused around listener locations for a *sweet–spot* solution. Experimental results are presented from perceptual tests comparing Multipole–Matched Rendering to both Wavefield Synthesis and stereo rendering using a linear speaker array. The experiments included 13 subjects and demonstrated that the perceived direction of a virtual sound source for the new method is comparable to that of Wavefield Synthesis (no significant difference). The results demonstrate the potential of Multipole–Matched Rendering as an efficient technique for rendering virtual sound sources in immersive environments.

Index Terms— Acoustic arrays, Moment methods, Singular value decomposition

1. INTRODUCTION

Rendering sound spatially to enhance the sense of immersion in virtual environments presents unique challenges [1, 2]. The traditional approach of Wavefield Synthesis (WFS) [3, 4] tries to faithfully recreate an original wavefield within an entire domain of interest and can require a relatively large number of speakers. This can be prohibitive for immersive virtual environments, especially if they need to be portable or set up in many smaller rooms. In addition, the reproduction of the sound field at every point in an immersive enironment with only a few listeners is often not necessary.

This paper presents a new method called Multipole–Matched Rendering (MMR) especially designed for immersive enviroments, and has the potential to require fewer speakers to spatially render sound, since it focuses on a small area around a single listener's head. The new method is based on matching multipole expansions [5, 6] of the sound fields emanated by the virtual source to the speakers using the Method of Moments (MoM) [7]. The number of higher order modes of the multipole expansions is selected to result in an overdetermined linear system of equations that are solved using the Singular–Value Decomposition (SVD). The ranking of singular value magnitudes is used for selecting the optimal subset of the Spherical Harmonics for the particular speaker–listener geometry and solving for the complex–valued weights on each speaker.

The algorithm is derived in section 2. To illustrate its implementation and validate its rendering performance, an experiment was designed to compare the MMR performance with that of WFS and classical stereo panning. The exeriment used a linear array of 8 speakers, which is typical for WFS applications. While the algorithm derived in section 2 suggests that MMR can handle more elaborate speaker–listener geometries, a linear array was used primarily for the sake of comparison with other methods. Detials of the experiment are described in section 3, and the results are presented in section 4.

2. THEORY

Consider an array of N speakers located at $\boldsymbol{r}_i'$ and emanating spherical waves. The composite field from the array can approximate the pressure field of a virtual monopole source located at $\boldsymbol{r}_s'$. For a listener at $\boldsymbol{r}_l$, the matching of rendered field to the virtual source can be expressed as:

$$\frac{e^{-jk|\boldsymbol{r}_l-\boldsymbol{r}_s'|}}{4\pi|\boldsymbol{r}_l-\boldsymbol{r}_s'|}=\sum_{i=1}^{N}A_i\frac{e^{-jk|\boldsymbol{r}_l-\boldsymbol{r}_i'|}}{4\pi|\boldsymbol{r}_l-\boldsymbol{r}_i'|}+e(\boldsymbol{r}_l,\boldsymbol{r}_1',\ldots,\boldsymbol{r}_N'), \quad (1)$$

where A_i is the unknown complex speaker weight and $e(\boldsymbol{r}_l,\boldsymbol{r}_1',\ldots,\boldsymbol{r}_N')$ is an error term stemming from the limited degrees of freedom from the finite speaker set. The

spherical wave terms in equation (1) are then replaced by their corresponding multipole expansions [5, p. 259, eq. 8.22]

$$\frac{e^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} = -jk\sum_{n=0}^{\infty}\mathrm{j}_n(kr_<)\mathrm{h}_n^{(2)}(kr_>) \sum_{m=-n}^{+n}\mathrm{Y}_{n,m}(\vartheta,\varphi)\mathrm{Y}_{n,m}^*(\vartheta',\varphi')\,, \quad (2)$$

where:

$$r_> = \begin{cases} r & , r > r' \\ r' & , r < r' \end{cases} \quad \text{and} \quad r_< = \begin{cases} r & , r < r' \\ r' & , r > r' \end{cases}. \quad (3)$$

For the following derivation it is assumed that a single listener is located at the coordinate system origin. Assuming that all real and virtual sources are outside the listener's head of radius r_l implies that $r_{<(l,s)} = r_l$ and $r_{>(l,s)} = r'_s$ for the source and $r_{<(l,i)} = r_l$ and $r_{>(l,i)} = r'_i$ for the speakers. This results in

$$\sum_{n=0}^{\infty}\mathrm{j}_n(kr_l)\mathrm{h}_n^{(2)}(kr'_s)\sum_{m=-n}^{+n}\mathrm{Y}_{n,m}(\vartheta_l,\varphi_l)\mathrm{Y}_{n,m}^*(\vartheta'_s,\varphi'_s) = \sum_{i=1}^{N}A_i\sum_{n=0}^{\infty}\mathrm{j}_n(kr_l)\mathrm{h}_n^{(2)}(kr'_i) \sum_{m=-n}^{+n}\mathrm{Y}_{n,m}(\vartheta_l,\varphi_l)\mathrm{Y}_{n,m}^*(\vartheta'_i,\varphi'_i) + e(\boldsymbol{r}_l,\boldsymbol{r}'_1,\ldots,\boldsymbol{r}'_N)\,. \quad (4)$$

We now apply MoM and mandate that the average error vanishes on a sphere of radius r_l around the listener's head:

$$\int_0^{\pi}\int_0^{2\pi} e(\boldsymbol{r}_l,\boldsymbol{r}'_1,\ldots,\boldsymbol{r}'_N)\mathrm{Y}_{n,m}^*(\vartheta_l,\varphi_l)r_l\sin\vartheta_l\,d\vartheta_l\,d\varphi_l \stackrel{!}{=} 0\,. \quad (5)$$

This allows us to exploit the orthogonality relation of the Spherical Harmonics, thus filtering a single term out of the summation over m. The final result is

$$\mathrm{j}_n(kr_l)\mathrm{h}_n^{(2)}(kr'_s)Y_{n,m}^*(\vartheta',\varphi') = \sum_{i=1}^{N}A_i\mathrm{j}_n(kr_l)\mathrm{h}_n^{(2)}(kr'_i)\mathrm{Y}_{n,m}^*(\vartheta'_i,\varphi'_i)\,. \quad (6)$$

This can be truncated and rewritten in matrix form as

$$[C_{j,i}]\,[A_i] = [B_j]\,. \quad (7)$$

The index j is related to n and m by $j = n^2 + n + m$, where $n = 0\ldots,N_R$ and $m = -n,\ldots,+n$, and N_R is the number of radial modes. Let $N = N_R^2$, and S be the number of speakers. Then C is an $N \times S$ matrix, $\boldsymbol{A}$ is an $S \times 1$ column vector and $\boldsymbol{B}$ is an $N \times 1$ column vector. The matrix entries are

$$C_{j,i} = \mathrm{j}_n(kr_l)\mathrm{h}_n^{(2)}(kr'_i)\mathrm{Y}_{n,m}(\vartheta'_i,\varphi'_i)\,, \quad (8)$$

the A_i are the unknown speaker weights and the right–hand side elements are

$$B_j = \mathrm{j}_n(kr_l)\mathrm{h}_n^{(2)}(kr'_s)\mathrm{Y}_{n,m}(\vartheta'_s,\varphi'_s)\,. \quad (9)$$

The number of speakers is fixed for a given setup but the maximum number of radial modes N_R is a degree of freedom. N_R should at least be large enough to ensure convergence of the method. For a typical setup with a low number of speakers (10–20), this usually results in an overdetermined linear system of equations. SVD can be used to compute the pseudo–inverse C^+ of the system [8], with the singular values representing the optimal subset of the Spherical Harmonics [9, sec. 2.5]. The system can then be repeatedly solved for multiple virtual source positions by setting up the right–hand side $\boldsymbol{B}$ according to (9) and a matrix–vector multiplication

$$\boldsymbol{A} = C^+\boldsymbol{B}\,. \quad (10)$$

3. EXPERIMENTAL SETUP

Perception tests were performed in the anechoic chamber of Deutsche Telekom Laboratories in Berlin, Germany. A linear speaker array was constructed consisting of 8 speakers with a spacing of 21.4cm between the speaker cones. The subjects were placed facing the speakers 130cm away from the center point of array. The speaker was placed at a height close to the inter–aural axis of each listener. A schematic image of the room layout can be seen in figure 1. This created an ideal setup for the stereo rendering method when using the far left and far right speakers as the stereo channels. So while this method is very simple it should perform reasonably well in this experiment. A scale representing angular measurements from $-40°$ to $40°$ with negative and positive directions corresponding to the listener's left and right, respectively, was placed on the wall behind the speakers for use as reference by the subjects.

To avoid bias created by the stimuli, three sound files were used with different spectral content. The sounds consisted of a pink noise burst, a music clip, and a speech clip. The MMR algorithm, however, was derived for sources composed of a single frequency. So the speaker weights for MMR were computed for each narrowband range of the source signal. To achieve this the source signal was decomposed into separate frequency bins from 25Hz to 22.05kHz in 25 Hz increments (corresponding to an 882 point FFT). The complex speaker weights were then calculated for each frequency bin using equation (10). An Inverse Fourier Transform was then performed on each channel of speaker amplitudes to compute the impulse response for the rendering operation. This impulse response was then convolved with a broadband source file using the open–source convolution engine BruteFIR.

Audio files were rendered using three methods: Multipole–Matched Rendering, Wave Field Synthesis, and Stereo Amplitude Panning. The purpose of the experiment was to

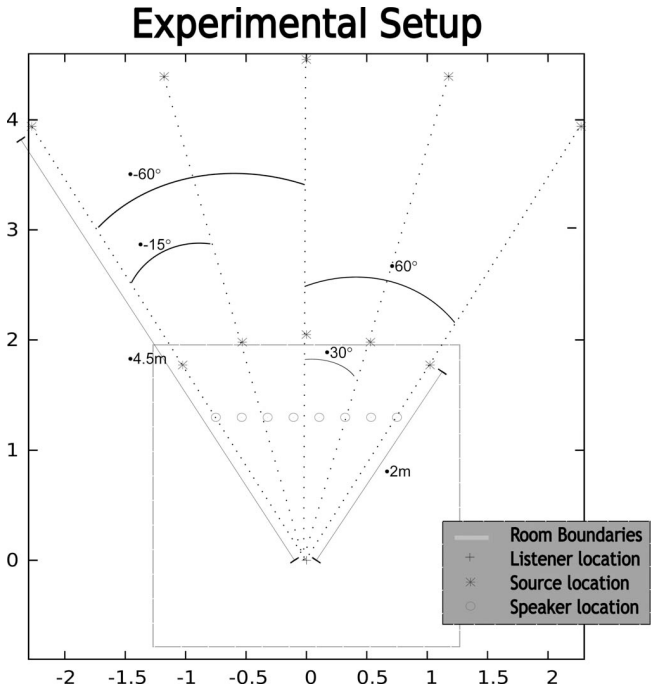


Fig. 1: Diagram of Experimental layout.

validate the performance of MMR for broadband signals by comparing its performance to WFS. Stereo amplitude panning was also included in the experiment to serve as a reference, since the setup was most conducive for this approach. For the spatial rendering methods two distances were rendered at five angular directions resulting in thirty samples for WFS and thirty samples for MMR. Two distances were used to determine if the different theoretical distances would result in a change in perceived direction for WFS and MMR. Because stereo rendering cannot recreate distance, only one distance was used at the same angular directions as WFS and MMR. This resulted in 15 samples for the stereo method and a total of 75 samples.

Thirteen subjects participated in the study. All participants were research staff and students working at Deutsche Telekom Labs characterized as non-expert listeners with average hearing. Each subject was presented with all seventy-five samples in random order and instructed not to move from the specified listening position. The listener indicated the perceived direction using a touch screen and a graphical user interface designed specifically for the localization tests. The GUI consisted of a slider and an image of the reference scale on the wall. In this way, a listener could determine a direction in reference to the primary scale and match the slider direction to the corresponding scale on the GUI. Average time for completion of the experiment was approximately 20 minutes per subject.

4. RESULTS

Statistical data for the three different stimuli can be seen in figure 2. The mean perceived angle as well as the 95% confidence intervals for all subjects and each method are represented in the figures. Also on the plots are the theoretical locations of rendered sound sources represented by horizontal black lines. To rule out influence of user error during the experiment, one outlier was removed from each set of data. Outlier removal was based on a mean squared error calculation of each response with respect to the theoretical source direction. The single response with the largest error was then removed from the data.

The results suggest that the Multipole–Matched Rendering method yields comparable localization of sound sources to wave field synthesis. In most cases, the theoretical source directon falls within the 95% confidence intervals of the collected data in both MMR and WFS. Results also show that MMR far exceeds localization of sources rendered with the stereo rendering method.

5. CONCLUSION

The perception experiment compared a newly developed algorithm for spatial audio rendering based on matched multipoles with the well established method known as wave field synthesis. As can be clearly seen in the resulting analysis, there is no statistical distinction between localization of virtual sources rendered with MMR and WFS. Matched–Multipole Rendering is able to achieve comparable results to wave field synthesis even in the case of a linear array which is not optimal for the MMR method. The advantage to MMR is the flexability of using arbitrary arrays with the trade–off being a reduction in the size of the sweet spot.

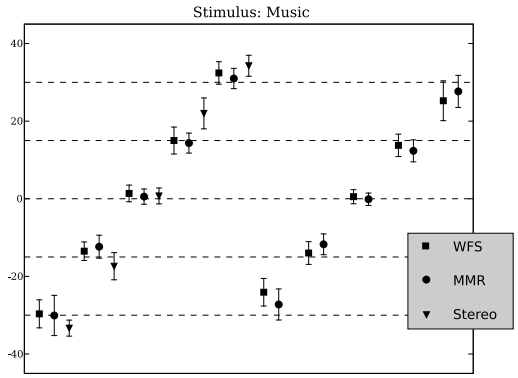
6. REFERENCES

[1] P. Larsson, D. Vastfj All, and M. Kleiner, "Better presence and performance in virtual environments by improved binaural sound rendering," in *Proceedings of the AES 22nd Intl. Conf. on virtual, synthetic and entertainment audio*, 2002, pp. 31–38.

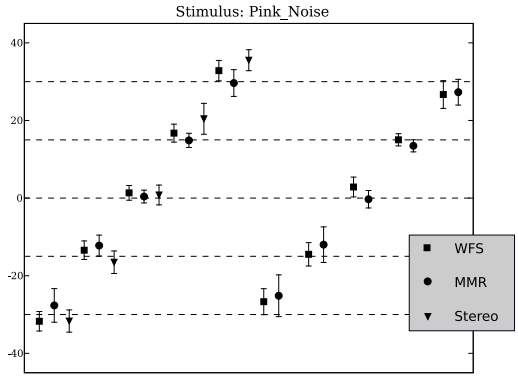
[2] N. I. Durlach and A. S. Mavor, "Virtual reality scientific and technological challenges," National research council report, National Academy Press, 1995.

[3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave–field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.

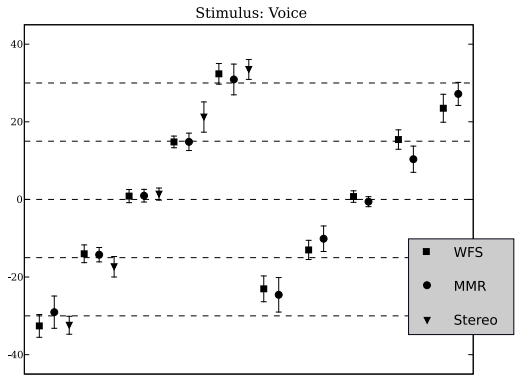
[4] Marinus M. Boone, Edwin N. G. Verheijen, and Peter F. van Tol, "Spatial sound–field reproduction by wave–field synthesis," *J. Audio Eng. Soc.*, vol. 43, no. 12, pp. 1003–1012, Dec. 1995.



(a)



(b)



(c)

Fig. 2: Plots of the statistical data of music (a), Pink Noise (b), and Speech (c) stimuli. Dashed horizontal lines represent theoretical source locations. Averages for each method at each location are shown as well as 95% confidence intervals.

[5] Earl George Williams, *Fourier Acoustics*, Academic Press, 1999.

[6] Heinz Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, Springer Verlag, 2007.

[7] Roger F. Harrington, *Field Computation by Moment Methods*, IEEE Press, 1993.

[8] Gene H. Golub and Charles F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[9] P. A. Nelson and Y. Kahana, "Spherical harmonics, singular–value decomposition and the head–related transfer function," *Journal of Sound and Vibration*, vol. 239, no. 4, pp. 607–637, Jan. 2001.