ROOM REFLECTIONS ASSISTED SPATIAL SOUND FIELD REPRODUCTION

P. N. Samarasinghe† *T. D. Abhayapala*† *M.A. Poletti*‡ *

† Research School of Engineering, CECS, The Australian National University, Australia.
‡ Callaghan Innovation, PO Box 31-310, Lower Hutt, New Zealand.

ABSTRACT

With recent advances in surround sound technology, an increased interest is shown in the problem of virtual sound reproduction. However, the performance of existing surround sound systems are degraded by factors like room reverberation and listener movements. In this paper, we develop a novel approach to spatial sound reproduction in reverberant environments, where room reverberation is constructively incorporated with the direct source signals to recreate a virtual reality. We also show that the array of monopole loudspeakers required for reproduction can be clustered together in a small spatial region away from the listening area, which in turn enables the array's practical implementation via a single loudspeaker unit with multiple drivers.

1. INTRODUCTION

A prevailing problem relevant to loudspeaker based spatial sound reproduction is the accurate and efficient rendering of virtual soundfields. Existing approaches to virtual source reproduction include *Amplitude panning* [1], various methods based on the synthetic head related transfer function (HRTF) [2], wavefield synthesis (WFS) [3] and higher order ambisonics [4]. Amplitude panning [1] applies a sound signal with different amplitudes to different loudspeakers which collectively recreate a virtual source. Even though this approach is simple and effective, it is limited for virtual sources at the loudspeaker radius. For virtual sources within the loudspeaker radius, complex array processing is required with additional focusing techniques. The HRTF based methods use the science of human sound perception to create virtual soundfields [2] . However, since the HRTF varies with varying listener positions and orientations, it only works when the user stays still in a small zone called the "sweet spot". To improve robustness to listener movements, additional features like head tracking and face tracking needs to be implemented which in turn increases the complexity of the system. Spatial reproduction based on higher order ambisonics and WFS often uses an array of monopole (and dipole for WFS) loudspeakers that enclose the listening region to produce a desired soundfield. However, when the design task involves a significantly large spatial area like a conference room, the implementation of a 3D loudspeaker array which encloses the entire listening region remains less practical. An alternate approach for sound reproduction in reflective environments with plane surfaces was introduced in [5]. This method used a linear array of loudspeakers, and appropriate delay and sum beamforming to utilize wall reflections in generating virtual sound beams. However, due to the limited directivity and the dependence on plane surface reflections, the applicability of this method is limited.

In this paper, we propose a novel approach for room reflection assisted sound reproduction, where we first calibrate the room response, and then use higher order ambisonics related mode matching to exactly create a virtual field. Furthermore, unlike in [4], we distribute the array of loudspeakers within a smaller spatial region that lies completely apart from the listening region. Therefore, it has the potential to be replaced by a single loudspeaker unit with multiple drivers, similar to the higher order loudspeaker concept introduced in [6]. Since the implementation of such a system is substantially simple compared to a larger array, and since reverberation is present in most enclosed environments, this solution is suitable for a plethora of applications like home theater systems, teleconferencing, film and TV production, environmental noise control etc.

Initially, we define a continuous spatial region where the listener is allowed to move around and a secondary spatial region within which an array of loudspeakers lie. Then, we characterize the room response between the two regions using the Room Transfer Function (RTF) parameterization introduced in [7, 8]. Next, we derive the RTF between each loudspeaker of the array and each point in the receiver region, and later add them together to predict the total room response at the listening region. Finally, we use the mode matching approach [4] to calculate loudspeaker weights that will produce the desired spatial soundfield coming from a virtual source position. For simplicity, we restrict our design to 2D reproduction, but all the theories developed can be readily extended for 3D reproduction.

*This work was supported by the Australian Research Council Discovery Grant DP 110103369.

2. PROBLEM FORMULATION

Our objective is to utilize the direct and reverberant fields caused by an arbitrary array of L loudspeakers lying inside a sizeable spatial region ζ to reproduce a desired soundfield at a secondary non-overlapping spatial region η, where the listener(s) lie. From here onwards, these two spatial regions will be referred to as the *source region* and the *receiver region* respectively. For the purpose of simplifying the analysis of this paper, we choose the receiver region η to be a circle of radius R_r centered about the origin $\boldsymbol{O}$ and the source region ζ to be another circle of radius R_s centered about $\boldsymbol{O_s}$. An arbitrary receiver point within η is denoted by $\boldsymbol{x}$ and an arbitrary source point within ζ is denoted by $\boldsymbol{y}$ where $\boldsymbol{y} = \boldsymbol{y}^{(s)} + \boldsymbol{R_{sr}}$ with $\boldsymbol{y}^{(s)}$ representing the same source location with respect to $\boldsymbol{O_s}$ and $\boldsymbol{R_{sr}}$ representing the vector connecting $\boldsymbol{O}$ to $\boldsymbol{O_s}$. The separation between η and ζ has a strong impact on the reproduction outcome and the preferred minimum separation of $\boldsymbol{R_{sr}}$ is later discussed in detail.

Consider an arbitrarily distributed loudspeaker array within ζ. If the ℓ^{th} ($\ell = 0 \cdots L$) loudspeaker located at $\boldsymbol{y}_\ell$ transmits a unit amplitude sound wave, the resulting room response at an arbitrary receiver point $\boldsymbol{x}$ is defined by the acoustic room transfer function $H(\boldsymbol{x}, \boldsymbol{y}_\ell, k)$, where $k = 2\pi f/c$ is the wave number, f is the frequency, and c is the speed of sound propagation. Therefore, for a weighted sum of L number of loudspeaker signals, the corresponding total soundfield incident at $\boldsymbol{x}$ will be

$$P(\boldsymbol{x}, k) = \sum_{\ell=1}^{L} W_\ell(k) H(\boldsymbol{x}, \boldsymbol{y}_\ell, k) \tag{1}$$

where $W_\ell(k)$ denotes the weight applied to the ℓ^{th} loudspeaker.

The design task of soundfield reproduction is to choose loudspeaker weights that produce a desired spatial soundfield within η, which can be represented in terms of a cylindrical harmonic decomposition of the form

$$P_d(\boldsymbol{x}, k) = \sum_{m=-M}^{M} \beta_m(k) J_m(k \|\boldsymbol{x}\|) e^{im\phi_x} \tag{2}$$

where ϕ_x denotes the angular orientation of $\boldsymbol{x}$, $\beta_m(k)$ denotes the modal coefficients of the desired field incident at η, $J_m(\cdot)$ represents the cylindrical Bessel function of order m and $M = \lceil keR_r/2 \rceil$ denotes the interior field truncation limit [6].

When addressing the above problem, we first decompose the room transfer function $H(\boldsymbol{x}, \boldsymbol{y}_\ell, k)$ according to a modal based parameterization introduced in [7, 8], and summarized in sec. 3.2. This enables us to rewrite (1) in terms of a cylindrical harmonic decomposition of the form (2). Next, we follow a mode matching approach [4] to find the loudspeaker weights that will minimize the normalized reproduction error over η. This step-by-step approach is descriptively explained in the following sections.

3. SOUNDFIELD REPRODUCTION

3.1. Source-receiver separation

The initial step in soundfield reproduction is determining an appropriate source region location for a given receiver region, so that both the direct and reverberant fields from the loudspeaker array can be fully utilized. We base this determination on the *reverberation radius* of a given room, which is the distance at which the sound pressure from the direct soundfield caused by a monopole loudspeaker is the same as that of the resulting reverberant field. When a receiver point lies within the reverberation radius from a particular source, the incident direct field is dominant than the reverberant field. Similarly, if η is fully or partially inside the reverberation radius from any loudspeaker location within ζ, the direct field overpowers the reverberant field at η. Therefore, when the desired field at η is from a direction other than that of ζ, it becomes difficult for the loudspeaker array to suppress the direct field and amplify the reverberation from the desired direction. For this reason, it is important to make sure that all points within η lie outside of the reverberation radius from all points within ζ. For a given room, the reverberation radius is defined as [9]

$$R_D = 0.057\sqrt{V/RT_{60}} \tag{3}$$

where V denotes the room volume and $RT_{60} = 0.1611\frac{V}{Sa}$ denotes the reverberation time with S representing the total surface area of the room and a representing the average absorption coefficient of the room surfaces. Since R_D is known for a given room, the aforementioned preference for source-receiver separation can be expressed by the condition

$$\|\boldsymbol{R_{sr}}\| > (R_D + R_s + R_r). \tag{4}$$

Once ζ is located according to (4), the next step is to decompose $H(\boldsymbol{x}, \boldsymbol{y}, k)$.

3.2. Parameterization of the room transfer function

In this section, we present a modal based parameterization for the RTF between two points $\boldsymbol{x}$ and $\boldsymbol{y}$. This parameterization was first introduced in [7] and [8] for 2D and 3D soundfields respectively. It is valid for any two arbitrary points from a continuous source region similar to ζ and a continuous receiver region similar to η that lie completely apart from each other. For an N^{th} order source region and an M^{th} order receiver region, the above parameterization is given by

$$H(\boldsymbol{x}, \boldsymbol{y}, k) = \sum_{n=-N}^{N} \sum_{m=-M}^{M} \alpha_m^n(k) J_n(k\|\boldsymbol{y}^{(s)}\|) J_m(k\|\boldsymbol{x}\|) e^{im\phi_{y_s}} e^{im\phi_x} \tag{5}$$

where ϕ_{ys} denotes the angular orientation of $\boldsymbol{y}^{(s)}$, $N = \lceil keR_s/2 \rceil$ denotes the truncation limit of the outgoing wavefield from ζ, $\alpha_m^n(k)$ denotes the RTF coefficients related to the two regions, and $\boldsymbol{y}^{(s)} = \boldsymbol{y} - \boldsymbol{R_{sr}}$. Based on (5) a finite set of $(2N+1)(2M+1)$ RTF coefficients are capable of calculating the room response between any two points $\boldsymbol{x}$ and $\boldsymbol{y}$ and therefore, the accurate acquisition of $\alpha_m^n(k)$ enables us to characterize the room response over the entire loudspeaker region ζ and the entire listening region η.

According to [7], $\alpha_m^n(k)$ is defined as the m^{th} order coefficient of the incident soundfield at η, caused by a unit amplitude single mode outgoing wavefield of mode n transmitted from ζ. To generalize an N^{th} order source field ζ, at least $(2N+1)$ number of distinct outgoing modes has to be considered, and each such single mode outgoing wavefield results in an M^{th} order receiver field comprising of $(2M+1)$ incident modes. Therefore, the acquisition of $\alpha_m^n(k)$ for all n and all m requires a total of at least $(2N+1)(2M+1)$ number of measurements to avoid spatial aliasing. It is also shown in [7] that given the room characteristics remain stationary over time, these measurements only require a single loudspeaker unit and a single microphone unit. The reader is encouraged to refer to [7, 8] in order to obtain a full understanding of the derivation of (5) and the acquisition of $\alpha_m^n(k)$.

3.3. Loudspeaker array processing

Once all the RTF coefficients related to ζ and η are recorded, (5) can be substituted in (1) giving a modal decomposition of the soundfield $P(\boldsymbol{x}, k)$ as

$$P(\boldsymbol{x}, k) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} \sum_{\ell=1}^{L} W_\ell(k) \alpha_m^n(k) J_n(k \|\boldsymbol{y}_\ell^{(s)}\|) e^{-in\phi_{ys\ell}} J_m(k\,\|\boldsymbol{x}\|) e^{im\phi_x}. \quad (6)$$

where $\phi_{ys\ell}$ denotes the angular orientation of $\boldsymbol{y}_\ell^{(s)}$. The above decomposition can be compared with (2) to obtain the incident field coefficients within η as

$$\beta_m(k) = \sum_{\ell=1}^{L} W_\ell(k) \underbrace{\sum_{n=-N}^{N} \alpha_m^n(k) J_n(k \|\boldsymbol{y}_\ell^{(s)}\|) e^{-in\phi_{ys\ell}}}_{t_{m\ell}} \quad (7)$$

For an M^{th} order desired field, this will result in $(2M+1)$ number of linear equations which can be represented in matrix form as

$$\boldsymbol{T}\boldsymbol{w} = \boldsymbol{\beta} \quad (8)$$

where $\boldsymbol{T}$ is a $(2M+1) \times L$ translation matrix with $t_{m\ell}$ of (7) being its elements, $\boldsymbol{w}$ is an $L \times 1$ vector of loudspeaker weights and $\boldsymbol{\beta}$ is a $(2M+1) \times 1$ vector of desired field coefficients. Since $\boldsymbol{T}$ and $\boldsymbol{\beta}$ are known, the required loudspeaker weights can be solved using

$$\boldsymbol{w} = \boldsymbol{T}^\dagger \boldsymbol{\beta} \quad (9)$$

where $\boldsymbol{T}^\dagger$ denotes the pseudoinverse. To avoid spatial aliasing, $L \geq (2M+1)$ has to be satisfied with (9) yielding the minimum energy weight solution [4]. The maximum soundfield order that can be controlled by an array of L number of loudspeakers is $M = (L-1)/2$ and since $M = \lceil keR_r/2 \rceil$, the array's maximum achievable frequency is limited to

$$f_{\text{max}} = c(L-1)/2\pi e R_r \quad (10)$$

3.4. Approximate reproduction error

In this section, we quantify the reproduction accuracy of the proposed method. For computational simplicity, we define an approximate relative error averaged over a finite number of design points as

$$\epsilon = \frac{\sum_{v=1}^{V} |P_d(\boldsymbol{x}, k) - P(\boldsymbol{x}, k)|}{\sum_{v=1}^{V} |P_d(\boldsymbol{x}, k)|} \quad (11)$$

where V denotes the number of design points considered from the reproduction region η.

4. SIMULATIONS

Simulation examples are presented to illustrate the accuracy of the proposed method. A $6 \times 5 \times 2.5$ m rectangular room was considered as the height invariant reverberant environment (no reflections from the floor and the ceiling) with its center defined as the origin $\boldsymbol{O}$. The wall reflection coefficients were assumed to be $[0.9\ 0.9\ 0.9\ 0.9\ 0\ 0]$ producing a highly reverberant environment. Since the reflection coefficient b of a given wall is related to its absorption coefficient a through $a = 1 - b^2$, the corresponding wall absorption coefficients were derived as $[0.19\ 0.19\ 0.19\ 0.19\ 1\ 1]$. The desired reproduction region η was assumed to be a circle of radius $R_r = 0.2$ m centered about $\boldsymbol{O}$ and the desired maximum frequency was fixed at $f_{\text{max}} = 1$ kHz.

From (3), the room's reverberation radius was calculated to be $R_D = 1.19$ m, and based on (4), the source region was decided to be located at $\boldsymbol{O_s} = (2, 2)$ with $\|\boldsymbol{R_{sr}}\| = 2.83$ m. The source region geometry was assumed to be a circle of radius $R_s = 0.5$ m centered about $\boldsymbol{O_s}$ and the corresponding RTF coefficients $\alpha_m^n(k)$ were calculated following the procedure given in [7]. Since the maximum reproduction frequency $f_{\text{max}} = 1$ kHz resulted in a receiver region of order $M = 5$ and a source region of order $N = 13$, the acquisition of $\alpha_m^n(k)$ required a minimum of $(2M+1)(2N+1) = 297$ measurements. When simulating the room reflections, a 2D adaptation of the image-source method [10] was used where the RTF between any two arbitrary points $\boldsymbol{x}$ and $\boldsymbol{y}$ is given in terms of

$$H(\boldsymbol{x}, \boldsymbol{y}, k) = H_0(k\,\|\boldsymbol{x} - \boldsymbol{y}\|) + \sum_{i=1}^{I} \zeta_i H_0(k\,\|\boldsymbol{x} - \boldsymbol{y_i}\|) \quad (12)$$

where I denotes the total number of image sources and $\boldsymbol{y_i}$ and ζ_i denote the position and accumulated wall reflection coefficient of the i^{th} image source respectively. In the following simulations, images up to the third order were considered resulting in a total of $I = 24$ images for each loudspeaker.

Once the room response was characterized through the acquisition of $\alpha_m^n(k)$, the next step was to arrange a suitable loudspeaker array within ζ. The minimum number of loudspeakers required to reproduce a desired field up to 1 kHz was calculated to be $L = 2M + 1 = 11$. While our design permits complete freedom to choose any arbitrary geometry for the loudspeaker array, we opted for a cylindrical shell/annulus that was recently used in [7] with a proven increase in array robustness. To obtain this geometry, the distance to each loudspeaker $\left\|\boldsymbol{y}_\ell^{(s)}\right\|$ was randomly varied (with a uniform distribution) within a virtual cylindrical shell of outer radius $R_s = 0.5$ m and an inner radius $R_s' = 0.4$ m, while their angular distribution remained uniform. The objective of varying $\left\|\boldsymbol{y}_\ell^{(s)}\right\|$ was to avoid Bessel zeros in $\boldsymbol{T}$ of (8) which decreases $\boldsymbol{T}$'s condition number and in turn increases the array's robustness [7].

The desired soundfield at η was assumed to be caused by a line source located at $\boldsymbol{R_0}$ with $\|\boldsymbol{R_0}\| > R_r$. The soundfield at any point $\boldsymbol{x}$ within η due to the above source can be given by [7]

$$\begin{aligned} P_d(\boldsymbol{x}, k) &= H_0(k\,\|\boldsymbol{x} - \boldsymbol{R_0}\|) \\ &= \sum_{m=-M}^{M} \underbrace{H_m(k\,\|\boldsymbol{R_0}\|)e^{-im\phi_0}}_{\beta_m(k)} J_m(k\,\|\boldsymbol{x}\|)e^{im\phi_x}. \end{aligned} \tag{13}$$

where ϕ_0 denotes the angular orientation of $\boldsymbol{R_0}$. The soundfield coefficients of (13) can be substituted for $\boldsymbol{\beta}$ in (9) to find the desired loudspeaker weights $\boldsymbol{w}$.

The first simulation example presents the reproduction of an incoming field caused by a line source located at $\boldsymbol{R_0} = (2, 3)$ m which also happens to be an image location of ζ's origin $\boldsymbol{O_s}$. Figures 1(a) and 1(b) show the desired and reproduced soundfields over η which appear almost identical. Figure 1(c) illustrates the direct field transmitted from the loudspeaker array within ζ in the absence of any room reflections. The bigger circle represents ζ, while the smaller one represents η. It can be observed that the transmitted field's maximum amplitude is only around twice as that of the desired field and therefore, the loudspeaker array's implementation is very much practical. Figure 1(d) shows the actual soundfield present throughout the source and receiver regions (including reverberation) where the smaller circle represents a scaled down version of 1(b). This figure ensures that the maximum amplitude heard elsewhere in the room is only around 4 times as that of the desired field and amplitudes of such scale only occur far from the listening region.

The next example shows a similar reproduction result for a desired field caused by a line source located at $\boldsymbol{R_0} = (-3, 0)$ m and unlike the previous case, this was not an image location of $\boldsymbol{O_s}$. Figures 2(a) and 2(b) show the desired and reproduced soundfields over η which re-assures the accuracy of the proposed method. From fig 2(c) it can be observed that the transmitted field's maximum amplitude is around 4 times as that of the desired field which is larger than the previous case. This could be due to the fact that the desired field's direction of arrival (DOA) ϕ_0, is almost opposite to that of the direct field coming from ζ. Therefore, the loudspeaker array has to amplify the reflected power from appropriate image sources using larger control weights inside ζ. Figure 2(d) shows the actual soundfield present over ζ and η, from which it's evident that the combination of direct and reverberant soundfields have accurately reproduced the desired DOA. Furthermore, according to fig. 2(d), the maximum amplitude heard outside of η is only around 4 times as that of the desired field and amplitudes of such scale do not occur near η. Finally, fig. 3 illustrates the reproduction error (11) plotted against varying DOAs of the desired soundfield for a source distance of $\|\boldsymbol{R_0}\| = 3.6$ m. It is important to note that the error curves obtained for different values of $\|\boldsymbol{R_0}\|$ turned out to be almost the same as fig. 3 and therefore, the effect of $\|\boldsymbol{R_0}\|$ on the reproduction error can be considered to be minimal. While each plot refers to a different frequency, it can be observed that there is an obvious increase in the error for $f_{\text{max}} = 1$ kHz which can be expected due to the array approaching it's maximum functionality. However, for frequencies below f_{max}, the error variation appears random yet acceptable for many DOAs. For example, the error curve representing $f = 800$ Hz has only 2 regions of ϕ_0 from around $1.12 - 1.81$ radians and $4.4 - 5.2$ radians where the relative error is fairly high. Both examples given in fig. 1(b) and fig. 2(b) are outside of the above regions thus, producing accurate reproduction. The reason behind the error peaks can be assumed to be triggered by the limited number of control weights, and can be easily improved by introducing a secondary source region ζ'.

5. CONCLUSION

We have presented a novel technique for spatial sound reproduction in reverberant environments. We used an array of loudspeakers distributed away from the listener, which in practice, can be replaced by a single loudspeaker unit with multiple drivers. The accuracy of the proposed method was supported by appropriate simulation examples up to 1 kHz. The introduction of a secondary loudspeaker unit of the above nature will not only improve the reproduction accuracy, but will also enable sound reproduction up to $3-4$ kHz. The practical implementation of such a setup has the potential to replace the stereo sound systems while delivering full surround sound and increased robustness to listener movements.

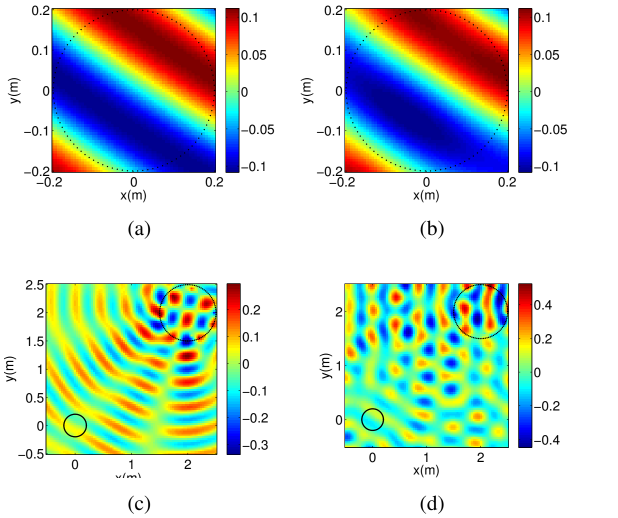

Fig. 1. For a desired line source at $\boldsymbol{R_0} = (2, 3)$ m and $f = 800$ Hz, the (a) desired and (b) reproduced soundfields over η and the (c) transmitted (without reflections) and (d) reproduced soundfields over ζ and η.

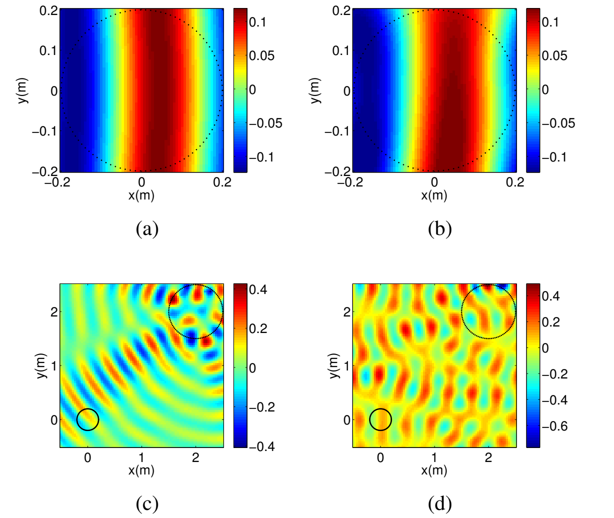

Fig. 2. For a desired line source at $\boldsymbol{R_0} = (-3, 0)$ m and $f = 800$ Hz, the (a) desired and (b) reproduced soundfields over η and the (c) transmitted (without reflections) and (d) reproduced soundfields over ζ and η.

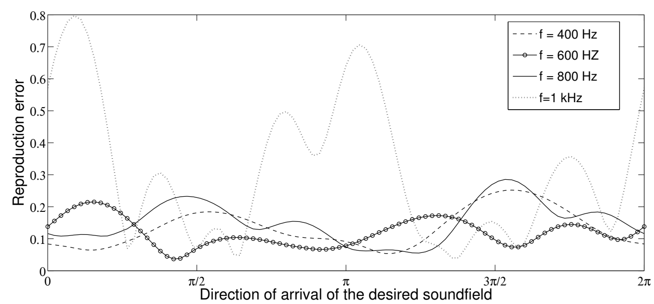

Fig. 3. Reproduction error (11) against DOA of the desired soundfield (ϕ_0) for a source distance of $\|\boldsymbol{R_0}\| = 3.6$ m over a set of frequencies $f = 400, 600, 800$ and 1000 Hz.

6. REFERENCES

[1] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[2] A. Mouchtaris, J.S. Lim, T. Holman, and C. Kyriakakis, "Head-related transfer function synthesis for immersive audio," in *IEEE Workshop on Multimedia Signal Processing*, pp. 155–160, 1998.

[3] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[4] T. Betlehem and T.D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the Acoustical Society of America*, vol. 117, pp. 2100, 2005.

[5] I.A. Bienek, A. Hooley, and P.R. Windle, "Compact surround-sound effects system," US Patent App. 11/988,667, 2006.

[6] M.A. Poletti, T.D. Abhayapala, and P. Samarasinghe, "Interior and exterior sound field control using two dimensional higher-order variable-directivity sources," *Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 3814–3823, 2012.

[7] P.N. Samarasinghe, T.D. Abhayapala, and M.A. Poletti, "On room impulse response between arbitrary points: An efficient parameterization," in *IEEE International Symposium on Communications, Control, and Signal Processing (ISCCSP)*, May 2014.

[8] P.N. Samarasinghe, T.D. Abhayapala, M.A. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE Transactions on Audio Speech and Language Processing*, Under review.

[9] H. Kuttruff, *Room acoustics*, Applied Science Publishers (London), London, UK, 1979.

[10] J.B Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943, 1979.