SOUNDFIELD RENDERING WITH LOUDSPEAKER ARRAYS THROUGH MULTIPLE BEAM SHAPING

F.Antonacci, A.Calatroni, A.Canclini, A.Galbiati, A.Sarti, S.Tubaro.

Politecnico di Milano
Dipartimento di Elettronica ed Informazione
Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

ABSTRACT

This paper proposes a method for the acoustic rendering of a virtual environment based on a geometric decomposition of the wavefield into multiple elementary acoustic beams, all reconstructed with a loudspeaker array. The point of origin, the orientation and the aperture of each beam is computed according to the geometry of the virtual environment that we want to render and to the location of the sources. Space-time filters are computed with a Least Squares approach to render the desired beam. Experimental results show the feasibility as well as the critical issues of the proposed algorithm.

Index Terms— wavefield rendering; acoustic beam; loudspeaker arrays

1. INTRODUCTION

Consider the problem of rendering the acoustics of a virtual environment in a low-reverberation (dry) room. In order to obtain this result, we need to modify the soundfield by adding the contributions of the waves reflected from walls of the virtual environment. In order to do so, we think of the soundfield as the superposition of beams originating from a number of virtual sources. Decompositions of the wavefield into elementary components are quite common in the literature: Ambisonics decomposes the soundfield through spherical harmonic functions ([1]), while WaveField Synthesis (WFS) ([2],[3]) exploits the Huygens principle, which states that any wavefront can be decomposed into a superposition of elementary spherical wavefronts emitted from secondary sources. Each loudspeaker, therefore, is independently controlled to operate as a secondary source. In this paper we focus on the rendering of soundfields through a superposition of beams originated from multiple image sources. In order to do so, we need to be able to render an acoustic beam and to compute the parameters of all beams according to the acoustics of the virtual environment we want to render.

As far as the rendering of a single beam is concerned, in [4] the authors propose an algorithm that controls the direction of the beam. However, this methodology is not able to accurately control the shape of the beampattern but just its direction. Moreover, the authors focus on the farfield case. Generalized Sidelobe Cancelling (GSC) [5] introduces multiple constraints on the beampattern but the number of constraints is limited by the number of loudspeakers, which places some limits on the accuracy of the beamshaping. An interesting solution for the shaping of an arbitrary beampattern can be found in [6]. Here the authors design the near field beamshaper from a farfield one. The validity of the beampattern in the broadband makes this algorithm very interesting. However, the computational cost prevents us from using this methodology for our purposes.

In this paper we propose an alternative technique to simulate by means of an array of M loudspeakers the arbitrary beampattern of a virtual source. More specifically, we define N test points in a listening area of arbitrary shape. We impose that the wavefield on the test points best approximates the wavefield produced by the virtual source with the specified beampattern. This condition yields a system of N equations whose unknowns are the M loudspeaker weights that we solve through Singular Value Decomposition. We have used beam tracing [7] to compute the configuration of beams given the geometry of the virtual environment we want to render and the position of the real source. The extension of the algorithm to multiple image sources is done by summing up the loudspeaker weights for each virtual source. The technique is extended to wideband signals by finding a filter for each loudspeaker through a frequency sampling approach.

The paper is structured as follows: Section 2 illustrates the problem and gives an overview of the background. Section 3 describes the proposed solution and illustrates the extension to broadband signals and to multiple virtual sources. Section 4 provides some experimental results to show the feasibility of the proposed approach. Finally, Section 5 draws some conclusions.

2. PROBLEM STATEMENT AND BACKGROUND

As said above, we conceive the soundfield as a superposition of acoustic beams. In this Section we formulate the problem of rendering a single acoustic beam and we provide an overview on the techniques available in the literature, which address the same problem.

Consider the problem of rendering within a pre-determined area of interest (*listening area*) the presence of a source that emits the signal $s(t)$. We do so with a uniform linear array of $m = 1, \dots, M$ loudspeakers placed in $\mathbf{p}_1, \dots, \mathbf{p}_M$. The listening area is sampled in points $\mathbf{a}_1, \dots, \mathbf{a}_N$ that are arbitrarily located.

We first consider a narrowband source. In Section 3 we extend our method to wideband sources. The sound pressure $p_n(t)$ in the test-points $\mathbf{a}_1, \dots, \mathbf{a}_N$ is described by the equation:

$$p_n(t) = \mathbf{g}_n^T \mathbf{h} s(t) = \Psi_n s(t) \tag{1}$$

where $s(t)$ is the source signal, $\mathbf{h} = [H_1 \quad H_2 \quad \dots \quad H_M]^T$ is the vector of complex coefficients applied to the loudspeakers

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 226007

and $\mathbf{g}_n = [g(\mathbf{p}_1, \mathbf{a}_n), \ldots, g(\mathbf{p}_M, \mathbf{a}_n)]$ is the juxtaposition of the Green's functions from each loudspeaker to the listening point $\mathbf{a}_n$ [8]:

$$g(\mathbf{p}_m, \mathbf{a}_n) = \frac{1}{4\pi||\mathbf{p}_m - \mathbf{a}_n||} e^{-j\omega \frac{||\mathbf{p}_m - \mathbf{a}_n||}{c}} , \qquad (2)$$

where c is the sound speed and ω is the frequency of $s(t)$. The term Ψ_n in eq.(1) is the spatial response of the loudspeaker array in $\mathbf{a}_n$. The goal of traditional beamshaping techniques is to leave undistorted the signal in $\mathbf{a}_{\overline{n}}$ ($\Psi_{\overline{n}} = 1$) while attenuating the field in the $N-1$ remaining points.

Our goal, however, is quite different as we aim at controlling the shape of the beampattern instead of minimizing the energy emitted by the loudspeaker array.

Generalized Sideobe Cancelling (GSC) [5] allows us to impose multiple constraints at the same time on the emitting directions. In GSC the maximum number of constraints is limited by the number of loudspeakers M. In order to obtain a smooth spatial response, however, one has to impose a number of constraints that is much higher. Moreover, traditional techniques fail when considering nearfield beams, as the juxtaposition of the propagation vectors for all the test points produces an ill-conditioned propagation matrix that yields an unstable filter.

In the next Section we will show how our technique addresses both problems. We will also extend the algorithm to wideband signals and multiple virtual sources to generate the desired soundfield as superposition of elementary beams.

3. PROPOSED SOLUTION

As shown in the previous Section, we need to use different design criteria with respect to the state of the art in order to achieve the desired beampattern: instead of minimizing the energy of the beamshaper, we are interested in controlling the wavefield over the listening area.

The location of the test points $\mathbf{a}_n,\ n = 1, \ldots, N$ and of the loudspeaker $\mathbf{p}_m,\ m = 1, \ldots, M$ in the listening area can be chosen at random. Figure 1 shows the notation we will use throughout the rest of the paper. As stated above, the algorithm we present

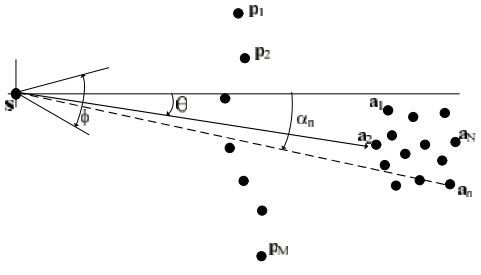

Figure 1: Notation proposed method for near-field beamshaping. The virtual source is located in $\mathbf{s}$ and emits a beam towards the direction θ with angular aperture ϕ; the points $\mathbf{p}_1, \ldots, \mathbf{p}_M$ denote the loudspeaker array; $\mathbf{a}_1, \ldots, \mathbf{a}_N$ are the test points.

in this paper is conceived for wideband signals. More specifically, we find the spatial filter for each loudspeaker through a frequency sampling approach: complex weights are found for a set of frequencies in the range of interest. Complex weights for intermediate frequencies are found through interpolation. We will give further details at the end of the section. In a first stage, for the sake of simplicity in notation, we omit the dependency of all the variables from the frequency. As already written in eq.(1), the spatial response of the loudspeaker array in $\mathbf{a}_n$ is

$$\Psi_n = \mathbf{g}_n^T \mathbf{h} ,$$

where $\mathbf{h}$ is the spatial filter of the loudspeaker array and $\mathbf{g}_n$ is the propagation vector from $\mathbf{p}_1, \ldots, \mathbf{p}_M$ to $\mathbf{a}_n$. As stated above, our goal is to render the acoustic beam emitted by a virtual source placed in $\mathbf{s}$ that emits towards the direction θ and with an angular aperture ϕ. The desired response in $\mathbf{a}_n$ is

$$\overline{\Psi}_n = g(\mathbf{s}, \mathbf{a}_n)\Theta(\theta, \phi, \alpha_n) ,$$

where $\Theta(\theta, \phi, \alpha_n)$ is the radiation pattern of the virtual source and α_n is the angle under which the n-th listening point is seen from $\mathbf{s}$, as depicted in Figure 1. Although in Section 4 we use a Gaussian beampattern, we remark that this is only a design choice that does not prevent us from using a custom function. Our goal is to approximate the desired beampattern through the spatial response of the loudspeaker array, which means that

$$\mathbf{g}_n^T \mathbf{h} = g(\mathbf{s}, \mathbf{a}_n)\Theta(\theta, \phi, \alpha_n) , \qquad (3)$$

where: $\mathbf{h} = [H_1 \quad H_2 \quad \ldots \quad H_M]^T$ is the coefficient vector and $\mathbf{g}_n = [g(\mathbf{p}_1, \mathbf{a}_n) \quad g(\mathbf{p}_2, \mathbf{a}_n) \quad \ldots \quad g(\mathbf{p}_M, \mathbf{a}_n)]^T$ is the juxtaposition of the Green's functions from the m-th emitter to the considered listening point.

If we consider all the listening points at once, we obtain the following matrix-formulation:

$$\mathbf{G}\mathbf{h} = \mathbf{r_d} , \qquad (4)$$

where $\mathbf{r_d} = [g(\mathbf{s}, \mathbf{a}_1)\Theta(\theta, \phi, \alpha_1), \ldots, g(\mathbf{s}, \mathbf{a}_N)\Theta(\theta, \phi, \alpha_N)]^T$ is the desired response; and $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_N]^T$ is the $N \times M$ propagation matrix from each loudspeaker to each test point. We observe that, in order to obtain a smooth beampattern, we use $N \gg M$, thus we have to use Least Squares-like techniques to obtain $\mathbf{h}$. The methodology we adopt to solve eq.(4) is related to the inverse problems theory [9]. The system in eq.(4) is overdetermined and it admits no exact solution. However, an estimation $\hat{\mathbf{h}}$ of the vector $\mathbf{h}$ can be calculated by introducing the pseudo-inverse operation on the matrix $\mathbf{G}$:

$$\mathbf{G}^+ = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H .$$

The loudspeakers weight vector is approximated by:

$$\hat{\mathbf{h}} = \mathbf{G}^+\mathbf{r_d} = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H\mathbf{r_d} . \qquad (5)$$

In general $\mathbf{G}\hat{\mathbf{h}} \neq \mathbf{r_d}$; however $\hat{\mathbf{h}}$ represents the best solution to the problem in the least squares sense.

The matrix $(\mathbf{G}^H\mathbf{G})$ is positive definite and, therefore, invertible. However, nothing guarantees that it will be a well-conditioned matrix. In order to avoid instability problems a reconditioning of $(\mathbf{G}^H\mathbf{G})$ is needed. We do so through an SVD decomposition:

$$\mathbf{G}^H\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H , \qquad (6)$$

where $\mathbf{U}$ and $\mathbf{V}$ are, respectively, the left and right singular vectors and $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_M)$ is the singular value diagonal matrix and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_M$. In order to perform the reconditioning, we seek for the greatest index k which guarantees that $\sigma_k/\sigma_1 \geq$

0.01. We retain the first k columns and rows of matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$. The approximate inverse matrix is therefore

$$(\mathbf{G}^H\mathbf{G})^{-1} \approx \mathbf{V}_k\mathbf{\Sigma}_k^{-1}\mathbf{U}_k^H \ . \quad (7)$$

The SVD inversion of the matrix $\mathbf{G}^H\mathbf{G}$ is a costly operation. However, we observe that a change in either the radial beampattern of the virtual source or its position correspond only to a change in the vector $\mathbf{r}_d$ of the desired response, as the matrix $\mathbf{G}$ is composed by the Green's functions from each loudspeaker to each test point. As a consequence, the SVD inverse of $\mathbf{G}^H\mathbf{G}$ may be easily precomputed once the positions of loudspeakers and test points are known.

3.1. Extension to multiple image sources and wideband signals

As stated in Section 1, our final goal is the rendering of a complex soundfield represented as a superposition of beams generating from a set of image sources. Since the system described by eq.(4) is linear, its extension to the multiple image sources turns out to be quite straightforward, using the superposition principle. Consider Z image sources characterized by their positions and radiation patterns. According to eq.(4), let $\mathbf{G}_z$, $\mathbf{h}_z$ and $\mathbf{r}_{\mathbf{d}z}$ be respectively the propagation matrix, the coefficient vector and the radiation pattern for the z-th image source, leading to the system $\mathbf{G}_z\mathbf{h}_z = \mathbf{r}_{\mathbf{d}z}$. Analogously, $\hat{\mathbf{h}}_z$ is the approximated solution for each of the systems in eq.(4). We obtain the global coefficient vector as the superposition of the individual coefficient vectors:

$$\hat{\mathbf{h}}_{TOT} = \sum_{z=1}^{Z} \mathbf{G}_z^{+}\mathbf{r}_{\mathbf{d}z} = \sum_{z=1}^{Z} (\mathbf{G}_z^H\mathbf{G}_z)^{-1}\mathbf{G}_z^H\mathbf{r}_{\mathbf{d}z} \ . \quad (8)$$

From the filter computed according to eq.(8) we find the complex weight vectors $\hat{\mathbf{h}}_{TOT}^{(k)}$ for the frequencies $f_1, \ldots, f_K$. From $\hat{\mathbf{h}}_{TOT}^{(k)}, k = 1, \ldots, K$ we obtain a set of filters whose frequency responses are $\mathbf{F}_1(f_l), \ldots, \mathbf{F}_M(f_l), l = 1, \ldots, L$ for each loudspeaker through parabolic interpolation of the amplitude and cubic interpolation of the phase. Even if we work with wideband signals, we work preserving the spatial Nyquist criterion, which means that the maximum operating frequency is limited by the loudspeaker reciprocal distance: $f_{max} < \frac{c}{2d}$, where d is the distance between emitters.

4. EXPERIMENTAL RESULTS

In this Section we show some simulations in order to illustrate the accuracy of the presented rendering method. We consider the experimental setup depicted in Fig. 2, that consists of:

- a circular array with radius of 1 m composed by $M = 32$ equally spaced omnidirectional emitters;
- a listening area including $N = 5000$ points uniformly distributed in a circular region concentric with the array with a radius of 0.9 m;
- a virtual room having dimensions $w \times w$;

The position and configuration of the Z beams is computed with fast beam tracing [7] starting from the position $\mathbf{s}_0$ of a real source inside the listening area. The radiance pattern function $\Theta(\theta, \phi, \alpha_n)$ used for the beamshaping is a gaussian function. In particular, the variance and the center of the gaussian windows

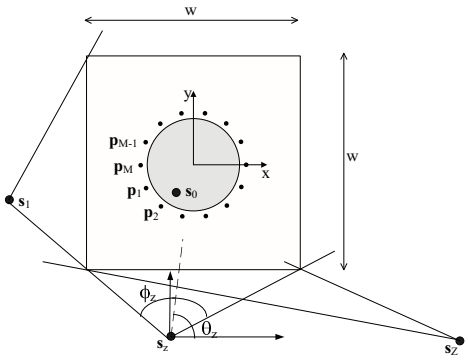

Figure 2: Experimental setup for the simulations: a circular array is composed by loudspeakers placed in $\mathbf{p}_1, \ldots, \mathbf{p}_M$; the gray shaded region within the array is the listening area; the $w \times w$ rectangle depicts the virtual environment; the real source is placed in $\mathbf{s}_0$; with reference to Fig.1 points $\mathbf{s}_1, \ldots, \mathbf{s}_Z$ denote a set of virtual image sources emitting beams towards the direction θ_z and with angular aperture ϕ_z.

are determined according to the direction θ and aperture ϕ of the beams to be rendered.

For each simulation two pictures will be shown: the absolute value $S(\mathbf{q})$ of the theoretical soundfield (where $\mathbf{q}$ is a point inside the listening area) as it were generated by the virtual image sources, and the absolute value $\hat{S}(\mathbf{q})$ of the rendered soundfield. The metric we use to evaluate the accuracy of the rendered soundfield is the *Root Mean Square Error* (*RMSE*), computed as follows:

$$E_{RMSE} = \sqrt{\frac{\sum_{i=1}^{Q}[S(\mathbf{q}_i) - \hat{S}(\mathbf{q}_i)]^2}{Q}} \ ,$$

where Q is the number of points constituting the soundfield images.

The first experiment we conduct aims at finding the optimum number K of frequencies to be used for the extension to wideband signals. A monochromatic signal at the frequency of $550Hz$ is emitted. This frequency is intermediate between two adjacent constrained frequencies f_k and f_{k+1}. As a consequence, we expect that, increasing the frequency step we obtain a lower accuracy of the rendered soundfield. The frequency step $\Delta f = f_{k+1} - f_k$ ranges from $10Hz$ to $100Hz$. Figure 3 shows the result: we observe that for $\Delta f \leq 40Hz$ the error is almost constant. For

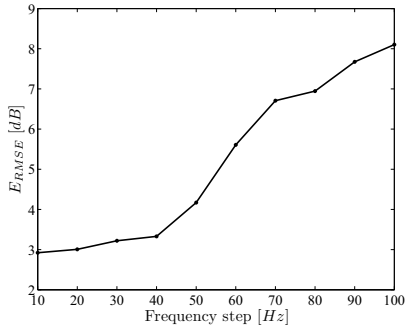

Figure 3: Plot of E_{RMSE} against the frequency step averaged for different positions of the real source $\mathbf{s}_0$.

the following tests we keep the real source position fixed in $\mathbf{s}_0 = (0.7\ m, 0.7\ m)$. The reflection coefficient of the room's walls is

0.7. We consider wall reflections up to the second order, that yields $Z = 16$. We use a frequency step of 40 Hz that gives $K = 77$ constrained frequencies in the range $[330\ Hz, 3370\ Hz]$.

We first analyze the behavior of the system at a constrained frequency (530 Hz) in a 10 $m \times 10$ m room. The resulting *RMSE* is 4.58 dB. Figure 4.(a) shows $S(\mathbf{q})$; $\hat{S}(\mathbf{q})$ is shown in

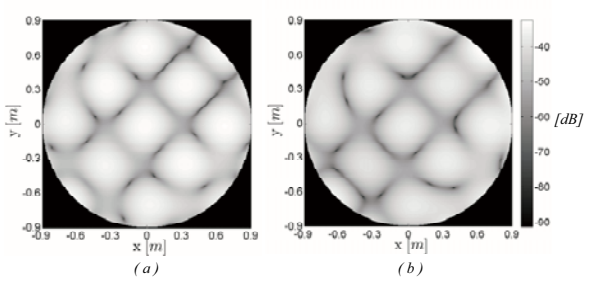

Figure 4: Plot of $S(\mathbf{q})$ and $\hat{S}(\mathbf{q})$ for a room 10 $m \times 10$ m wide at the constrained frequency of 530 Hz.

Figure 4.(b). We will see later that for a different configuration of the virtual room, the error introduced by some virtual sources $\mathbf{s}_z$ is higher than the *RMSE* of the soundfield. This is due to the fact that for some configurations the beam is oriented towards directions that make it difficult to obtain the desired soundfield, since the mainlobe falls outside the listening area.

We now repeat the previous experiment in a 3 $m \times 3$ m room. Fig.5.(a) depicts $S(\mathbf{q})$ while Fig.5.(b) shows $\hat{S}(\mathbf{q})$; the resulting *RMSE* is 3.13 dB. The comparison between Fig.5 and Fig.4 and their *RMSE* values suggests that, reducing the size of the virtual environment the performances of the system improve. Experimentally we confirm this trend.

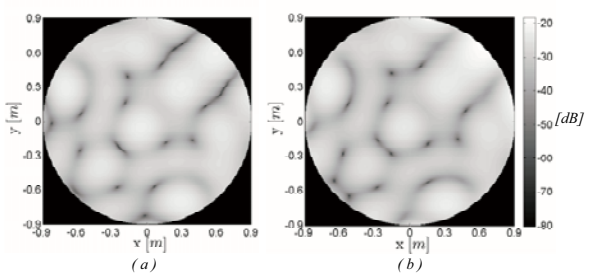

Figure 5: Plot of $S(\mathbf{q})$ and $\hat{S}(\mathbf{q})$ for a room 3 $m \times 3$ m wide at the constrained frequency of 530 Hz.

We now test the performances of the rendering system in a 3 $m \times 3$ m room at a non-constrained frequency. We choose a test frequency of $550Hz$, that is intermediate between the constrained frequencies of $530Hz$ and $570Hz$. Figures 6.(a) shows $S(\mathbf{q})$, while Figure 6.(b) shows $\hat{S}(\mathbf{q})$. For this experiment we obtain $E_{RMSE} = 3.38\ dB$: even if the working frequency is not constrained, no significant distortion appears due to the interpolation process.

Finally, Table 4 separates the contribution to the soundfield error due to each virtual source for the same configuration of Fig.5. The main lobes of the beam generated by the sources $\mathbf{s}_5$ and $\mathbf{s}_{16}$ fall outside the listening area, therefore they are not perceptually relevant. This observation motivates us, as a future work, in finding a heuristic that prunes the tree of virtual sources retaining only those that are significant.

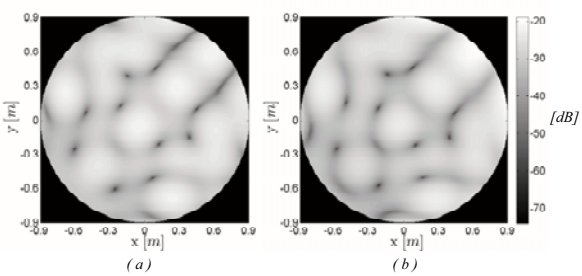

Figure 6: Plot of $S(\mathbf{q})$ and $\hat{S}(\mathbf{q})$ for a room 3 $m \times 3$ m at the unconstrained frequency of 550 Hz.

Table 1: Error of $\hat{S}(\mathbf{q})$ due to the rendering of all the virtual sources for the same configuration of Fig.6. The beams generated by the sources $\mathbf{s}_5$ and $\mathbf{s}_{16}$ are not relevant for the perception of the soundfield, hence the relative error values are indicated with the symbol X.

	$\mathbf{s}_1$	$\mathbf{s}_2$	$\mathbf{s}_3$	$\mathbf{s}_4$
$E_{RMSE}[dB]$	2,22	1,92	1,9	2,23
	$\mathbf{s}_5$	$\mathbf{s}_6$	$\mathbf{s}_7$	$\mathbf{s}_8$
$E_{RMSE}[dB]$	X	2,46	7,99	2,42
	$\mathbf{s}_9$	$\mathbf{s}_{10}$	$\mathbf{s}_{11}$	$\mathbf{s}_{12}$
$E_{RMSE}[dB]$	9,41	2,32	2,42	10,08
	$\mathbf{s}_{13}$	$\mathbf{s}_{14}$	$\mathbf{s}_{15}$	$\mathbf{s}_{16}$
$E_{RMSE}[dB]$	2,45	7,74	2,44	X

5. CONCLUSIONS

In this paper we have proposed a technique for rendering the acoustics of a virtual environment in a dry room. In particular, we have used a decomposition of the wavefield into elementary beams. Experimental results have shown the feasibility and some critical issues of the algorithm. We are currently working on a real-time demonstrator and on new solutions that work on a room with reverberations.

6. REFERENCES

[1] P.Fellgett, "Ambisonics. part one: General system description," *Studio Sound*, vol. 40, no. 1, pp. 20–22, Aug. 1975.

[2] A.J.Berkhout, "A holographic approach to acoustic control," *J.Audio Eng.Soc.*, vol. 36, pp. 977–995, Dec. 1988.

[3] S. Spors, H. Teutsch, and R. Rabenstein, "High-quality acoustic rendering with wave field synthesis," in *Vision, Modeling, and Visualization*, Nov. 2002, pp. 101–108.

[4] B. Pueo, J. Escolano, and M. Roma, "Precise control of beam direction and beamwidth of linear loudspeaker arrays," in *Proceedings of Sensor Array and Multichannel Signal Processing Workshop Proceedings*, June 2004, pp. 538–541.

[5] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. AP, pp. 27–34, Jan 1982.

[6] R.A.Kennedy, T.D.Abhayapala, and D.B.Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Transactions on Signal Processing*, vol. 46, no. 8, August 1998.

[7] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 4, pp. 812–824, May 2008.

[8] S.Spors, "Spatial aliasing artifacts produced by linear loudspeaker arrays used for wave field synthesis," in *Proc. of IEEE ISCCSP 2006*, Marrakech, Morocco, March 2006.

[9] A.Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2004.