JOINT OPTIMIZATION OF LOUDSPEAKER PLACEMENT AND RADIATION PATTERNS FOR SOUND FIELD REPRODUCTION

Hanieh Khalilian, Ivan V. Bajić, and Rodney G. Vaughan

School of Engineering Science, Simon Fraser University

ABSTRACT

A new method is presented for optimizing loudspeaker placement, radiation patterns, and excitations for Sound Field Reproduction (SFR). A power constraint is included which is to help control the sound increase in the regions away from listening volume. For a known primary source, the loudspeaker locations and patterns are jointly optimized by Constrained Matching Pursuit (CMP), and this can be undertaken offline, i.e., before system operation. The excitations of the designed loudspeakers are then determined by conventional Lagrangian optimization. Simulations for free space conditions show that the new method yields a lower reproduction error under a power constraint than other SFR methods.

***Index Terms*—** Sound field reproduction, Constrained Matching Pursuit, higher order loudspeakers.

1. INTRODUCTION

Sound Field Reproduction (SFR) is the process of creating a desired sound field in a listening area using an array of loudspeakers. The source of the desired field is referred to as the primary source while the loudspeakers, which create the sound field, are known as secondary sources. SFR systems have static Degrees of Freedom (DOFs), which are the parameters optimized for the system design, and dynamic DOFs which are optimized during system operation and determine the time complexity. For example in [1, 2], the locations and radiation patterns of loudspeakers are determined before the system operation, and the excitations are optimized during the system operation. In [3–5], the locations of loudspeakers are fixed while the excitation and radiation patterns of loudspeakers change during system operation. The method in this paper *jointly* optimizes the locations and radiation patterns of loudspeakers before the system operation, and updates only the excitations during the system operation.

For a given location of the primary source, the loudspeaker locations can be optimized by methods developed in [6–10]. The placement algorithms from [6, 7, 10] minimize the weighted combination of the ℓ_2 norm of the sound field reproduction error and the ℓ_1 norm of the loudspeaker excitation vector. This results in a sparse representation of the desired field in terms of the Acoustic Transfer Functions (ATFs) of the loudspeakers. In [8], the locations of the loudspeakers are determined to approximate an "ideal" ATF matrix, which uses singular value decomposition of an idealized "benchmark" ATF matrix. The method in [9] optimizes loudspeaker locations using a CMP algorithm.

For array analysis, loudspeaker radiation patterns are often considered to be omni-directional despite this seldom being the case. Patterns are classically characterized by their peak and distributed directivities and parameters such as sidelobe levels, front-to-back ratio (and for antennas, polarization purity), etc., and the spherical mode expansion is a standard tool in their analysis and synthesis. Recently, for SFR, the directive nature of a loudspeaker pattern has been expressed as its highest mode without further direct reference to the modal makeup. Examples include "dipole" patterns ($\cos\theta$) for 2D SFR [11], fixed higher-order modal patterns for 3D SFR [2, 12], higher-order patterns for 2D SFR [3] and 3D SFR [4]; and with a power constraint with just the first two modes, viz., omni and dipole [5], and with higher orders [13].

For jointly optimizing the loudspeaker locations and radiation patterns, a constrained version [9] of the matching pursuit algorithm [14] selects samples of ATFs from a dictionary of the loudspeakers at different locations with various patterns. The excitation is then determined using conventional Lagrangian optimization. The contribution is that we generalize the approach of [9] where all patterns were omnidirectional and fixed, and instead optimize both the loudspeaker locations and their higher-order patterns together. The numerical experiments below include the cases of fixed location loudspeakers and omnidirectional loudspeakers, enabling performance comparisons for free-space conditions. Our approach also highlights the use of the power constraint as a tool to trade-off the accuracy of the reproduction within the listening volume against the unwanted synthesis of sound outside of this volume.

2. CONSTRAINED MATCHING PURSUIT

The CMP algorithm [9] is suitable for power-constrained SFR because the cost function is non-convex in terms of the loudspeaker locations, and finding the global optimum is not feasible owing to the size of the problem. With $\mathbb{C}^Q$ denoting a complex vector space, $\mathfrak{D} \subset \mathbb{C}^Q$ a dictionary of unit-norm

vectors and $\mathbf{a} \in \mathbb{C}^Q$ a Q-dimensional vector, the CMP is an iterative procedure for the linear expansion of $\mathbf{a}$ in terms of vectors in $\mathfrak{D}$ so that the coefficients are upper-bounded by p:

$$\mathbf{a} = \sum_{n=1}^{N} \alpha_n \mathbf{b}^{(n)} + R^{N+1}\mathbf{a} \qquad \text{s.t.} \quad \sum_{n=1}^{N} |\alpha_n|^2 \leq p, \quad (1)$$

where $\mathbf{b}^{(n)}$ is the dictionary member selected at the n-th iteration and $||\mathbf{b}_n||_2^2 = 1$, α_n is the corresponding coefficient, and $R^{N+1}\mathbf{a}$ is the approximation error vector after N iterations. The constraint limits the power of the excitations, elaborated below. Given N (which is also the number of loudspeakers, see below) the expansion is obtained by:

1) Initialization: Let $n = 1$ and $R^n\mathbf{a} = \mathbf{a}$.

2) Dictionary member selection: Find (brute force search) the dictionary member with largest inner product with $R^n\mathbf{a}$:

$$\mathbf{b}^{(n)} = \underset{\mathbf{b}\in\mathfrak{D}}{\operatorname{argmax}} \, |\mathbf{b}^H R^n \mathbf{a}|, \quad (2)$$

where $(\cdot)^H$ denotes complex conjugate. The corresponding expansion coefficient is calculated as:

$$\alpha_n = \underset{|\alpha|^2 \leq p_n}{\operatorname{argmin}} \, ||\alpha \mathbf{b}^{(n)} - R^n\mathbf{a}||_2^2, \quad (3)$$

where $p_n \geq 0$ is the upper bound on $|\alpha_n|^2$ chosen such that $\sum_{n=1}^{N} p_n = p$. We set $p_n = p/N$, i.e., a uniform distribution, to ensure that power is allocated to each iteration, which turns out to suit a constrained power situation. Equation (3) can be solved by Lagrangian methods, which leads to

$$\alpha_n = \begin{cases} \sqrt{p_n}\frac{(\mathbf{b}^{(n)})^H R^n\mathbf{a}}{|(\mathbf{b}^{(n)})^H R^n\mathbf{a}|} & \text{if } \sqrt{p_n} \leq |(\mathbf{b}^{(n)})^H R^n\mathbf{a}|, \\ (\mathbf{b}^{(n)})^H R^n\mathbf{a} & \text{otherwise.} \end{cases} \quad (4)$$

3) Desired vector at the next iteration: Calculate the approximation error as $R^{n+1}\mathbf{a} = R^n\mathbf{a} - \alpha_n\mathbf{b}^{(n)}$, and increase n by 1. If $n \leq N$ go to step *2)*, otherwise stop.

3. METHOD

The goal is to find locations, patterns, and excitations for N loudspeakers in order to reproduce the desired sound field as accurately as possible. Let $\mathbf{s} = [s_1, s_2, .., s_N]^T$ be the loudspeaker excitation vector where s_n is the excitation of the n-th loudspeaker at a single frequency, M be the number of sampling points in the listening area, $\mathbf{G}$ be the $M \times N$ ATF matrix whose (m, n)-th element is the ATF at m-th sampling point $\mathbf{y}_m$ of the n-th loudspeaker located at $\mathbf{x}_n$. The M samples of the desired and reproduced fields are arranged in vectors $\mathbf{p}^{des}$ and $\mathbf{p}$ respectively, where $\mathbf{p} = \mathbf{G}\mathbf{s}$. The optimal excitation vector $\mathbf{s}$ can be computed again using a Lagrangian approach:

$$\mathbf{s}_{opt} = \operatorname{argmin}(||\mathbf{p}^{des} - \mathbf{G}\mathbf{s}||_2^2 + \gamma||\mathbf{s}||_2^2) \quad (5)$$

where γ is the Lagrange multiplier chosen as in [15] such that $||\mathbf{s}||_2^2 \leq p_{\max}$, and $p_{\max}$ is the maximum normalized power. The solution $\mathbf{s}_{opt}$ and the resulting reproduction error depend on the ATF matrix $\mathbf{G}$ [8, 15], whose entries in turn depend on the loudspeaker radiation patterns and their locations relative to the sampling points.

An L-th order loudspeaker refers to the highest term in the expansion of its outgoing waves, e.g., [16]:

$$G_{m,n}(r, \theta, \phi) = \sum_{l=0}^{L} \sum_{u=-l}^{l} C_{u,l}^L h_l(kr) Y_u^l(\theta, \phi), \quad (6)$$

where $r = ||\mathbf{x}_n - \mathbf{y}_m||_2$, (θ, ϕ) give the direction to the m-th sampling point $\mathbf{y}_m$, the origin (phase centre) is at the n-th loudspeaker $\mathbf{x}_n$, $k = 2\pi/\lambda$ is the wave number with λ the wavelength, $h_l(\cdot)$ is the l-th order outward traveling spherical Hankel function, $Y_u^l(\theta, \phi)$ is the spherical harmonic function of degree u and order l, and $C_{u,l}^L$ are the coefficients. Note the pattern relates just to the function $Y_u^l(\theta, \phi)$.

The physical realization of a higher order loudspeaker pattern could be through a micro-electronic array, with element weighting integrated into the same single device. The element weights do not not relate directly to the harmonics, rather the desired complex pattern is taken as a function to be matched by the far field of a finite array of weighted directional elements, and this is the pattern synthesis problem well known in antenna design. This aspect is beyond the scope of this paper which is confined to the modeled systems aspect. An omni-directional radiation pattern, which is often used in SFR and other array processing, is a zeroth-order loudspeaker ($L = 0$):

$$G_{m,n}(r, \theta, \phi) = C_{0,0}^0 \frac{e^{jkr}}{\sqrt{4\pi}kr}, \quad (7)$$

where $C_{0,0}^0 = k/\sqrt{4\pi}$ in free space [7].

To make sure that power-constrained comparisons of SFR systems with various loudspeaker radiation patterns are fair, the coefficients are scaled by $k/\sqrt{4\pi}$. Then the coefficients are constrained such that $\sum_{l=0}^{L} \sum_{u=-l}^{l} |C_{u,l}^L|^2 \leq 1$ in Step 5 of Algorithm 1, below.

3.1. Dictionary construction

Each candidate loudspeaker location is identified with a sub-dictionary whose members are samples of $h_l(kr)Y_u^l(\theta, \phi)$ of various radiation patterns placed at that location. So the i-th candidate location gives rise to an $M \times (L + 1)^2$ matrix $\mathbf{D}_i$ whose $(L + 1)^2$ columns represent samples at the M sampling points of the terms $h_l(kr)Y_u^l(\theta, \phi)$ from the summation in (6). Such sub-dictionaries are generated for $N_v \gg N$ candidate locations and arranged into a large dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_{N_v}]$, which is a $M \times (N_v \cdot (L + 1)^2)$ matrix.

3.2. Optimization of loudspeaker locations and patterns

The optimization consists of finding, at each iteration, the term $h_l(kr)Y_u^l(\theta, \phi)$ (across all remaining candidate locations

Algorithm 1 Loudspeaker pattern and location optimization

Input: $\mathbf{D}$ ▷ dictionary
Input: $\mathbf{p}^{des}$ ▷ desired vector
Input: $\mathbf{x}_i^v$ ▷ coordinate of i-th candidate location
Input: N ▷ number of loudspeakers
Output: A ▷ selected loudspeaker locations
Output: $\{\mathbf{c}^n\}$ ▷ inner coefficients of loudspeakers

1: Set $R^1\mathbf{p}^{des} = \mathbf{p}^{des}$ and $A = \varnothing$
2: **for** $n = 1$ to N **do**
3: Find a column of $\mathbf{D}$ that is most correlated with $R^n\mathbf{p}^{des}$, denoted by $\mathbf{d}^{(n)}$.
4: Find sub-dictionary to which $\mathbf{d}^{(n)}$ belongs. Suppose $\mathbf{D}_i$ is that sub-dictionary, then the i-th candidate location is selected at this iteration, $A = A \cup \{\mathbf{x}_i^v\}$.
5: To find the radiation pattern for the selected location, apply the CMP algorithm on the sub-dictionary $\mathbf{D}_i$ by considering $R^n\mathbf{p}^{des}$ as the desired vector, and using $p = 1$ in (1). The resulting expansion coefficients give the normalized inner coefficients of the pattern, and the corresponding linear combination of the columns of $\mathbf{D}_i$ gives the loudspeaker ATF at the M sampling points, denoted $\mathbf{g}^{(n)}$.
6: The output will be the coefficients ($C_{u,l}^L$) of the higher order loudspeaker, which is arranged in a $1 \times (L+1)^2$ vector $\mathbf{c}^n$.
7: Compute α_n in (4) using $p_n = p_{\max}/N$, $\mathbf{b}^{(n)} = \mathbf{g}^{(n)}$ and $R^n\mathbf{a} = R^n\mathbf{p}^{des}$.
8: Update the approximation error as $R^{n+1}\mathbf{p}^{des} = R^n\mathbf{p}^{des} - \alpha_n\mathbf{g}^{(n)}$. Remove $\mathbf{D}_i$ from $\mathbf{D}$ so that the same location cannot be selected again.
9: **end for**
10: **return** $\{\mathbf{c}^1, .., \mathbf{c}^n\}$ and A.

and all patterns) that is most correlated with the current approximation error, and running CMP over the corresponding sub-dictionary. The details are shown in Algorithm 1, which optimizes the loudspeaker locations and patterns when the frequency and location of the primary source are fixed.

For a range of frequency, the primary source is sampled at $\{f_1, f_2, ..., f_W\}$, and the region in which the primary source may be located is also sampled at locations $\{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_Y\}$. Then, Algorithm 1 is applied in order to find the locations and radiation patterns of N/WZ loudspeakers when the location of primary source is at $\mathbf{z}_y$ and its frequency is f_w for $1 \leq y \leq Y$ and $1 \leq f \leq W$. The optimized locations and radiation patterns of loudspeakers remain fixed during the system operation, and only their excitations are updated to minimize the reproduction error.

After the design phase, during the system operation, the ATFs of individual optimized loudspeakers are computed and arranged into the ATF matrix $\mathbf{G}$. The excitation vector $\mathbf{s}_{opt}$ for the loudspeaker array is found and updated by solving (5) to minimize the reproduction error during the system operation based on the variable parameters of the primary source while the positions and patterns of loudspeakers are fixed.

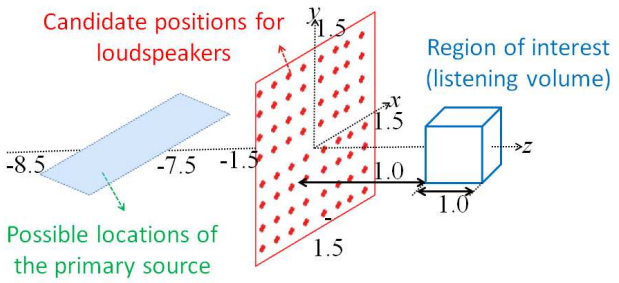

Fig. 1. SFR system configuration.

4. SIMULATION RESULTS

The SFR system considered in the simulations is shown in Fig. 1. The loudspeakers are placed on a 3 m×3 m square centered at the origin, and the listening volume is a 1 m ×1 m ×1 m cube located 1 m away. The primary source is omni-directional and its excitation is arbitrarily set to $8e^{j\angle 0}$. The numbers of loudspeakers, candidate locations, and sampling points are $N = 25$, $N_v = 400$, and $M = 98$, respectively. The sampling points are distributed uniformly on the surface of the cubic volume. The frequency range is 100 Hz to 1500 Hz, which is typical in SFR system simulations [1, 7]. The maximum normalized power is in the range 0.1 to 1. Higher allowed power generally leads to better reconstruction within the listening cube, but less control over the resulting field outside the cube.

The performance of the SFR system designed by the proposed method, *System 4 (S4)*, is compared against that of the following three systems through simulations. *System 1 (S1)* is a benchmark system with N omni-directional loudspeakers placed uniformly on the 3 m×3 m square region. *System 2 (S2)* consists of N higher-order loudspeakers placed uniformly on the 3 m×3 m square region, where only the radiation patterns are designed by the proposed algorithm. *System 3 (S3)* consists of N omni-directional loudspeakers whose locations on the 3 m×3 m square are determined using the CMP-based algorithm from [9]. In the proposed system and *S2*, the loudspeaker order is set to $L = 5$. These systems are compared in terms of the normalized reproduction error as in [7, 8, 13] at $M_2 = 125000$ points which are distributed uniformly in the listening cube.

S2 involves optimized loudspeaker patterns, but no optimization of loudspeaker locations whereas in *S3*, the locations of loudspeakers are optimized, but not their patterns. The comparison of these two systems with the proposed one illustrates the benefits of joint optimization of loudspeaker patterns and locations. In addition, in all systems, the positions and patterns of loudspeakers are designed before the

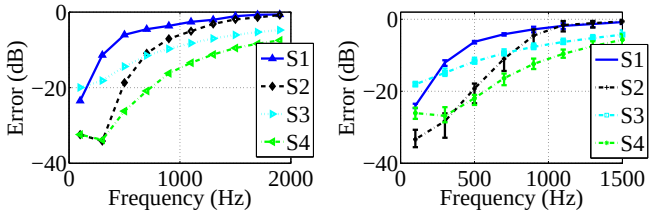

Fig. 2. Reproduction error versus frequency for (a) fixed primary source location, (b) variable primary source location.

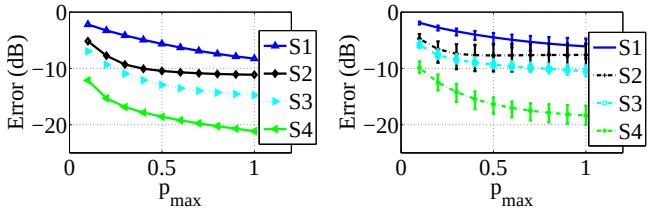

Fig. 3. Reproduction error versus $p_{\max}$ for (a) fixed primary source location, (b) variable primary source location.

system operation (during the design phase), so the number of dynamic DOFs (time complexity) of all systems is equal. The required processing time for Algorithm 1 is 0.5 sec on a 3GHz Intel Core 2 Quad Q9650 processor, but it is not a crucial parameter since this algorithm works offline.

First, the reproduction error is examined at different frequencies with $p_{\max} = 0.3$ in Fig. 2 for two cases. In the first case, the location of the primary source is fixed at $(0, 0, -8)$, and the frequency for which the system parameters are optimized during the design phase is sampled at $\{100, 300, 600, 1000, 1500\}$ Hz. Therefore, the patterns and (or) locations of 5 loudspeakers are optimized for each frequency. After the design phase, the frequency of the primary source is changed between 100 and 1500 Hz, and the reproduction error is illustrated in Fig. 2(a). The performance degrades with increasing frequency because all systems have increasingly spatial undersampling.*S4* outperforms other systems across the whole frequency range, with error performance gains between 3–10 dB relative to the next best system at mid and high frequencies.

For the second case, the locations of the primary source range between $(-4, \pm 0.5, -8)$ and $(+4, \pm 0.5, -8)$, as shown in Fig. 1. In this experiment, the frequency for which the system parameters are optimized during the design phase is sampled only at $\{100, 1500\}$ and the possible locations of the primary source are sampled at $(0, 0, -8)$, $(0, 0, -4)$, $(0, 0, +4)$, $(0, 0, -2)$, and $(0, 0, +2)$. Therefore two or three loudspeakers are designed for each frequency bin and sampled location. Again, after the design phase, the frequency and location of the primary source are changed and the reproduction error is shown in Fig. 2(b). In this figure, each curve shows the average reproduction error at the corresponding frequency when the location of primary source is moving across 30 positions between $(-4, \pm 0.5, -8)$ and $(+4, \pm 0.5, -8)$, and the error bars show the maximum and minimum errors at each frequency. As shown in this figure, the reproduction error is less than -10 dB for frequencies less than 1000 Hz, 600 Hz, 600 Hz, 300 Hz in *S4*, *S3*, *S2*, and *S1*, respectively, which implies that the average performance of the new system is better than that of the other systems.

The reproduction error as a function of the maximum normalized power is depicted in Fig. 3 for $f = 800$ Hz for the first case and second case of the previous experiment. As shown in this figure, the performance of all methods improves as the power increases, meaning that the desired field is more accurately reproduced within the listening cube. However, outside the cube, the deviation from the desired field may become larger as the power increases. Considering this point, one should choose the maximum power that is large enough to meet the error objective within the listening cube, but no larger. There is no formal design criteria for the choice of $p_{\max}$, this must be done on a case-by-case SFR configuration, but this parameter is clearly useful for controlling the tradeoff between the SFR accuracy and the unwanted sound generated outside a listening volume. According to the figure, for the first case, the error of *S4* for $p_{\max} = 0.1$ is equal to that of *S3* for $p_{\max} = 0.4$, and *S2* and *S1* for $p_{\max} > 1$. Hence, for the same reproduction error, controlling the sound field outside the listening cube is much easier in the proposed system compared to the other three systems.

5. CONCLUSION

A new method for SFR allows the locations and the patterns, in terms of their spherical harmonic excitations, to be optimized together for minimum reproduction error. The method lends itself to any configuration, but a specific situation is investigated comprising a square aperture for the loudspeakers and a close-proximity cubic listening volume. For other configurations, the error gain depends on the geometry as well as the possible locations of the primary source. Simulations for the chosen configuration are undertaken for free-space conditions. This means that to interpret the results for an indoor situation the walls of the room would have to be reasonably non-reflective, and the loudspeakers would have to have configurable patterns. In the design phase, the patterns and locations are determined based on a dictionary of possible locations of the primary source and its frequency makeup, and then only the excitations of the loudspeakers change during the system operation. The power constraint provides a parameter that allows management of a tradeoff between the accuracy of the sound field reproduction and the suppression of unwanted sound outside of the listening volume.

6. REFERENCES

[1] P. Gauthier and A. Berry, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *J. Acoust. Soc. Am.*, vol. 2, pp. 662–678, Feb. 2005.

[2] M. Poletti, F. M. Fazi, and P. A. Nelson, "Sound-field reproduction systems using fixed-directivity loudspeakers," *The Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3590–3601, 2010.

[3] M. A. Poletti and T. D. Abhayapala, "Spatial sound reproduction systems using higher order loudspeakers," in *Proc. IEEE ICASSP'11*, Prague, May 2011, pp. 57–60.

[4] P. N. Samarasinghe, M.A. Poletti, S. M. A. Salehin, T. D. Abhayapala, and F. M. Fazi, "3D sound field reproduction using higher order loudspeakers," in *Proc. IEEE ICASSP'13*, Vancouver, BC, May 2013, pp. 306–310.

[5] M. A. Poletti, F. M. Fazi, and P. A. Nelson, "Sound reproduction systems using variable-directivity loudspeakers," *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1429–1438, 2011.

[6] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Parsimonious sound field synthesis using compressive sampling," in *Proc. IEEE WASPAA'09.*, 2009, pp. 253–256.

[7] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using lasso," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 1902–1921, Nov. 2010.

[8] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "Towards optimal loudspeaker placement for sound field reproduction," in *Proc. IEEE ICASSP'13*, Vancouver, May 2013, pp. 321–325.

[9] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "Loudspeaker placement for sound field reproduction by constrained matching pursuit," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'13)*, New Paltz, NY, Oct. 2013.

[10] N. Radmanesh and I. Burnett, "Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm," *IEEE Transaction on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 378–387, 2013.

[11] J. Ahrens and S. Spors, "An analytical approach to 2.5D sound field reproduction employing linear distributions of non-omnidirectional loudspeakers," in *Proc. IEEE ICASSP'10*, 2010, pp. 105–108.

[12] J. Ahrens and S. Spors, "An analytical approach to 3D sound field reproduction employing spherical distributions of non-omnidirectional loudspeakers," in *4th Intl. Symp. Communications, Control, Signal Processing (ISCCSP)*, Mar. 2010, pp. 1–5.

[13] H. Khalilian, I. V. Bajić, and R. G. Vaughan, "3D sound field reproduction using diverse loudspeaker patterns," in *Proc. IEEE ICME'13*, San Jose, CA, Jul. 2013.

[14] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.

[15] T. Betlehem and C. Withers, "Sound field reproduction with energy constraint on loudspeaker weights," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 8, pp. 2388–2392, Oct. 2012.

[16] E. G.Williams, *Fourier Acoustics*, Academic Press, 1999.