Compressed Sensing based Multi-zone Sound Field Reproduction

Qipeng Feng, Feiran Yang, and Jun Yang
Key Laboratory of Noise and Vibration Research
Institute of Acoustics, Chinese Academy of Sciences
Beijing 100190, China

***Abstract*—This study proposes a multi-zone 2-D sound field reproduction approach using compressed sensing methods. We assume that the desired multi-zone sound field which originates from a small number of far field sound sources is sparse in the plane wave domain. From the measurements using the random microphone array, the desired sound field is decomposed as a set of plane waves via ℓ_1-norm minimization. The complex multi-zone sound field reproduction problem is then reduced to the plane wave reproduction over a single spatial region. Simulations demonstrate that the proposed method leads to a significantly reduced number of required microphones for accurate multi-zone sound reproduction.**

I. Introduction

The aim of spatial multi-zone sound field reproduction (MZR) is to provide different sound scenes to listeners within different zones of the reproduction space using the same loudspeaker array. The variety of approaches to MZR can be roughly divided into two classes: the first one treats the multi-zone sound field as a whole, and decomposes the desired sound field by means of spatial basis functions, i.e., plane waves or cylindrical harmonics [1-5]. Based on the decomposition, the complex MZR system is reduced to a relatively simple single-zone reproduction problem. This kind of methods usually have special requirements for the arrangement of the loudspeakers and microphones, and the size of the accurate reproduction region depends on the number of microphones.

The second class consists of pressure matching approaches, where the sound pressure is optimized at control points in the least squares (LS) sense [6-10]. This kind of MZR methods improve the flexibility of the loudspeaker setup and express the reproduced multi-zone sound field as a set of weighted transfer functions between the loudspeakers and the control points. If the frequency of the desired sound field is high, more control points are required because the accuracy of the least squares fit depends on the number of control points. For common reproduction scenes in a reverberant room, it seems impractical to utilize too many microphones to capture the transfer functions between the loudspeakers and microphones.

In many cases of interest, the desired sound field which originates from a small number of sound sources, is likely to be sparse in the source domain. Furthermore, due to the propagation characteristic of the spherical and cylindrical waves, the far field sound pressure distribution of a point or line sound source can be approximated as a plane wave. This prompts the idea of assuming that the desired multi-zone sound field which contains a small number of far field sound sources can be sparsely identified by a set of plane waves. Motivated by this idea, we propose a novel MZR approach based on *Compressed Sensing*, namely, *Sparse Plane-wave Decomposition (SPD)*. First, we use a random microphone array to sample the target sound field. The sampled sound field is then decomposed into a set of weighted plane waves via ℓ_1-norm minimization. Existing single-zone sound field reproduction approaches, such as the continuous loudspeaker method [11], can be used in the reproduction step. Simulation results demonstrate that the SPD method reduces the number of required microphones for accurate MZR and yields better performance than traditional least squares approach and the MZR method proposed in [1].

II. Theory

In this study, we mainly focus on the 2-D case in anechoic chamber. As shown in Fig. 1, Q loudspeakers are equally placed on a circle of R, and the radiation sound field of the qth loudspeaker is represented as $\frac{i}{4}H_0^{(1)}\left(k\left\|\mathbf{x}-\mathbf{x}_q\right\|\right)$, where $i=\sqrt{-1}$ is the imaginary unit, $\|\cdot\|$ denotes the Euclidean distance, $\mathbf{x}_q$ represents the cartesian coordinate of the qth loudspeaker, $H_0^{(1)}(\cdot)$ is the 0th-order Hankel function of the first kind, $k=2\pi f/c$ is the wavenumber, f is the frequency, and c is the speed of sound propagation. The weight vector of loudspeaker array is defined as $\mathbf{w}=\left[w_1(k),\cdots,w_Q(k)\right]^T$. Considering two circular reproduction zones of radii r: the bright zone and the quiet zone, N_0 randomly distributed sampling points are selected in each zone. The coordinates of sampling points are $\mathbf{x}_n, n=1,\cdots,N$, where $N=2N_0$.

A. LS approach for MZR problem

The desired sound pressure at sampling points can be expressed as $\mathbf{p}_d=\left[S_d(\mathbf{x}_1,k),\cdots,S_d(\mathbf{x}_{N_0},k),0,\cdots 0\right]^T$. By employing the superposition principle of sound field, the reproduced sound pressure at the nth sampling point can be given as

$$S_r(\mathbf{x}_n,k)=\sum_{q=1}^{Q}w_q(k)\,G(\mathbf{x}_n|\mathbf{x}_q,k), \tag{1}$$

where $G(\mathbf{x}_n|\mathbf{x}_q,k)=\frac{i}{4}H_0^{(1)}\left(k\left\|\mathbf{x}_n-\mathbf{x}_q\right\|\right)$ denotes the transfer function between the qth loudspeaker and the nth sampling

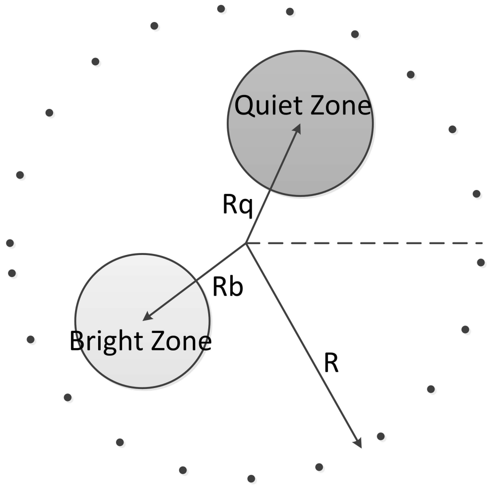

Fig. 1. The 2-D MZR with Q loudspeakers.

point in free field, and $\mathbf{p}_r = [S_r(\mathbf{x}_1, k), \ldots, S_r(\mathbf{x}_N, k)]^T$ is the reproduced sound pressure at sampling points.

The LS-based MZR problem can be mathematically formulated as the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}} \|\mathbf{G}\mathbf{w} - \mathbf{p}_d\|, \\ &\text{s.t.} \ \|\mathbf{w}\| \leq E_0, \end{aligned} \tag{2}$$

where $\mathbf{G}$ is the $N \times Q$ matrix of acoustic transfer functions, the (n, q)th element of $\mathbf{G}$ is $\frac{i}{4} H_0^{(1)}(k \|\mathbf{x}_n - \mathbf{x}_q\|)$, and E_0 is the limit on the loudspeaker weight energy. The total loudspeaker weight energy quantifies the ability of an array to suppress spatially uncorrelated noise in the loudspeaker signals. E_0 should not be large to improve the robustness of the MZR system. Essentially, (2) is a convex problem which can be solved using the CVX toolbox [12].

B. Continuous loudspeaker method for MZR problem

Any arbitrary 2-D desired sound field $\mathbf{S}_d(\mathbf{x}, k)$ in a spatial region with no sources and no scatterers can be represented by the following Fourier-Bessel series expansion [13]:

$$S_d(\mathbf{x}, k) = \sum_{m=-\infty}^{+\infty} \alpha_m^d(k) J_m(k\|\mathbf{x}\|) e^{im\phi_x}, \tag{3}$$

where $J_m(\cdot)$ is the Bessel function of order m, m is referred to as the mode, $\alpha_m^d(k)$ is the mth mode coefficient, and superscript “d” represents the desired sound field. Considering the spacial high-pass property of Bessel functions and the fact that the desired sound field has to be bounded within a limited spatial region, we can truncate this series expansion as follows:

$$S_d(\mathbf{x}, k) = \sum_{m=-M_0}^{+M_0} \alpha_m^d(k) J_m(k\|\mathbf{x}\|) e^{im\phi_x}, \tag{4}$$

where $M_0 = \lceil kR \rceil$, in which $\lceil \cdot \rceil$ denotes the top integral function.

The desired sound field can also be approximated as a sum of weighted plane waves, as follows:

$$S_d(\mathbf{x}, k) = \sum_{p=1}^{P} g(p) e^{ik\mathbf{x}\cdot\hat{\phi}_p}, \tag{5}$$

where $\hat{\phi}_p \equiv (1, \phi_p)$ is the unit vector in the direction of the pth plane wave, $\phi_p = 2\pi p/P$ is the incident angle of the pth plane wave, P is the number of plane waves, $g(p)$ denotes the pth plane wave weight. Using the Jacobi-Anger expression, a plane wave arriving from angle ϕ_{pw} can be written as follows:

$$e^{ik\mathbf{x}\cdot\hat{\phi}_{pw}} = \sum_{m=-M_0}^{M_0} i^m e^{-im\phi_{pw}} J_m(k\|\mathbf{x}\|) e^{im\phi_x}. \tag{6}$$

By substituting (6) into (5), the desired sound field coefficients $\alpha_m^d(k)$ can written as

$$\alpha_m^d(k) = \sum_{p=1}^{P} g(p) i^m e^{-im\phi_p}. \tag{7}$$

For Q discrete loudspeakers uniformly distributed on a circle of R with angle interval $\Delta\phi = \frac{2\pi}{Q}$, the qth loudspeaker weight can be derived as follows [11]:

$$w_q(k) = \sum_{m=-M_0}^{M_0} \frac{2}{i\pi H_m^{(1)}(kR)} \alpha_m^d(k) e^{im\phi_q} \Delta\phi, \tag{8}$$

where $H_m^{(1)}(\cdot)$ is the mth-order Hankel function of the first kind, $\alpha_m^d(k)$ denotes mode coefficient of the desired sound field, and $\phi_q = q\Delta\phi$ is the polar angle of the qth loudspeaker. According to (7) and (8), if we can decompose the desired sound field into a series of weighted plane waves, then we can obtain the loudspeaker weights using the continuous loudspeaker method. For a simple sound field, we use compressive sensing methods to accomplish this decomposition, which leads to a significantly reduced number of required microphones for accurate multi-zone sound reproduction.

III. Method

Compressed sensing describes a set of mathematical methods and theories that enable accurate recovery of original signals from a small number of samplings. Given the MZR problem, if the desired multi-zone sound field can be described by a sparse set of elementary functions, then we can use a small amount of microphone samples to reconstruct the desired sound field over the predefined target regions accurately. Given a single-zone sound field reproduction problem, if a sound field originates from a small number of far field sound sources, then a plane wave basis can be selected as the sparsity domain [14]. For a MZR system, the reflected sound field of a single loudspeaker in a reverberant room can also be sparsely identified in the domain of plane wave decomposition [15].

Based on the aforementioned results, it is assumed that the desired multi-zone sound field that originates from a few far field sound sources is sparse in the plane wave domain. We select a sufficiently large number of plane waves

$\{e^{ik\mathbf{x}\cdot\hat{\phi}_p}\}, p = 1, 2, \ldots, P$, and sample the desired sound field at N randomly distributed positions within the bright and quiet zone. The main task of compressed sensing is to derive the sparsest set of plane waves from $\{e^{ik\mathbf{x}\cdot\hat{\phi}_p}\}$ to describe the desired sound field faithfully. Thus, it amounts to solving the following ℓ_0-norm minimization problem [16] [17]

$$\min_{\mathbf{g}} \|\mathbf{g}\|_0, \quad \text{s.t.} \mathbf{p}_d = \mathbf{\Phi}\mathbf{g}, \tag{9}$$

where $\mathbf{\Phi}$ is the $N \times P$ dictionary matrix whose columns are measurements vectors of the plane wave set $\{e^{ik\mathbf{x}\cdot\hat{\phi}_p}\}$, $\mathbf{p}_d$ is the $N \times 1$ vector that contains the desire sound field pressure at N randomly distributed microphones, $\mathbf{g}$ is the $P \times 1$ plane wave coefficients vector, and $\|\cdot\|_0$ indicates the number of non-zero elements of a vector.

The ℓ_0-norm minimization problem (9) is combinatorial and non-deterministic polynomial-time hard (NP-hard), however, it can be relaxed to a convex problem

$$\min_{\mathbf{g}} \|\mathbf{g}\|_1, \quad \text{s.t.} \mathbf{p}_d = \mathbf{\Phi}\mathbf{g}, \tag{10}$$

if $\mathbf{\Phi}$ satisfied the Restricted Isometry Property (RIP) [18]. Various families of random matrices satisfy the RIP with high probability. In this study, we introduce the randomness of $\mathbf{\Phi}$ by randomly locating the microphones within the desired region. In practice, $\mathbf{p}_d$ and $\mathbf{g}$ are often corrupted by noise and approximation error; thus, a more robust model is to relax the minimization problem (10) to

$$\min_{\mathbf{g}} \|\mathbf{g}\|_1, \quad \text{s.t.} \frac{\|\mathbf{p}_d - \mathbf{\Phi}\mathbf{g}\|}{\|\mathbf{p}_d\|} \leq \varepsilon, \tag{11}$$

where ε is a specified precision that depends on the noise variance. This problem can be solved using CVX [8]. After the decomposition, the desired multi-zone sound field is expressed as a set of weighted plane waves. The loudspeaker weight can be obtained by the continuous loudspeaker method [11].

IV. Simulation

In this simulation, 75 loudspeakers are equally placed on a circle of 1m and two circular reproduction zones with the radius of 0.3m are considered. The polar coordinates of the two regional centers are $(0.6, \pi)$ and $(0.6, 0)$ respectively. The desired sound field in the bright zone is a plane wave arriving from angle ϕ_{pw} at a frequency of $f = 2000$Hz, whereas the desired sound field in the quiet zone is zero. The following parameters are used: $M_0 = 37$, $\varepsilon = 0.01$, and $c = 340m/s$. The size of the dictionary $\mathbf{\Phi}$ in (13) is $2N_0 \times 1000$. Two measurements are adopted to evaluate the performance of the MZR system [5]:

(i). The acoustic brightness contrast between two predefined zones to evaluate the energy leakage, expressed as follows:

$$\varsigma(k) = \frac{\int_{D_b} |\mathbf{S}_b(\mathbf{x}, k)|^2 d\mathbf{x} \Big/ D_b}{\int_{D_q} |\mathbf{S}_q(\mathbf{x}, k)|^2 d\mathbf{x} \Big/ D_q}, \tag{12}$$

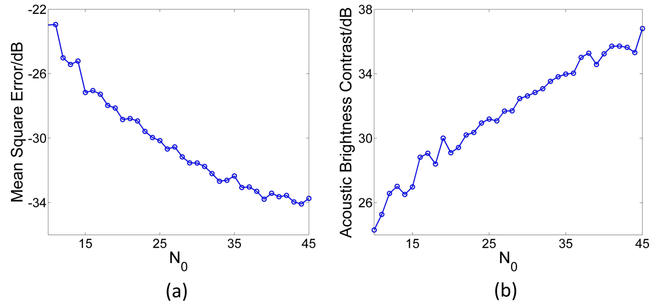

Fig. 2. MZR performance with different number of microphones. Plotted are (a) the NMSE in the bright zone and (b) the acoustic brightness contrast between bright and quiet zones. Each point has been averaged over 50 trial runs.

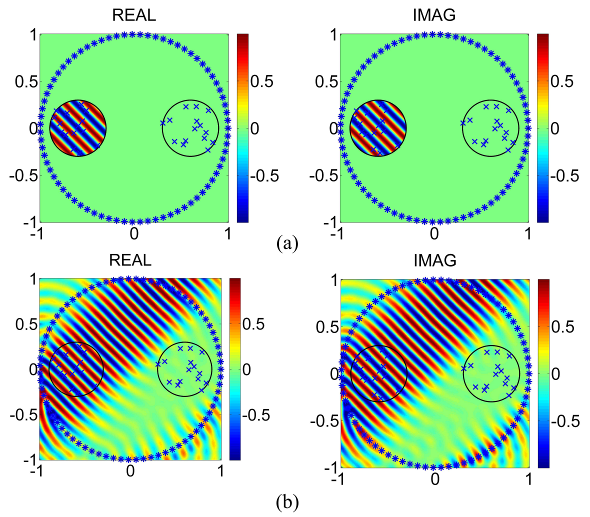

Fig. 3. Reproduction of a 2-D two-zone sound field with 15 randomly distributed sampling points within each zone in free field. The desired sound field in the bright zone is a plane wave arriving from angle $\phi_{pw} = \pi/4$ at a frequency of $f = 2000$Hz, whereas the desired sound field in the quiet zone is 0. (a) Desired sound field, and (b) reproduced sound field.

where $\mathbf{S}_b(\mathbf{x}, k)$ and $\mathbf{S}_q(\mathbf{x}, k)$ denote the reproduced sound field in the bright zone and the quiet zone. D_b and D_q mark the size of the bright zone and the quiet zone.

(ii). The normalized mean square error (NMSE) between the desired sound and the reproduced sound over the bright zone to evaluate the reproduction accuracy, expressed as follows:

$$e(k) = \frac{\int_{D_b} |\mathbf{S}_d(\mathbf{x}, k) - \mathbf{S}_b(\mathbf{x}, k)|^2 d\mathbf{x}}{\int_{D_b} |\mathbf{S}_d(\mathbf{x}, k)|^2 d\mathbf{x}}, \tag{13}$$

where $\mathbf{S}_d(\mathbf{x}, k)$ denotes the desired sound field.

Fig. 2 depicts the reproduction performance at $\phi_{pw} = \pi/4$ with different number of microphones. Notably, the more microphones are used, the better MZR performance is. Using a small number of microphones, e.g. $N_0 = 10$, the acoustic brightness contrast between two zones is still larger than 20 dB and the NMSE of the bright zone is smaller than -20 dB. Fig. 3 (a) illustrates the desired sound field which is a plane wave arriving from $\pi/4$, and Fig. 3 (b) presents the

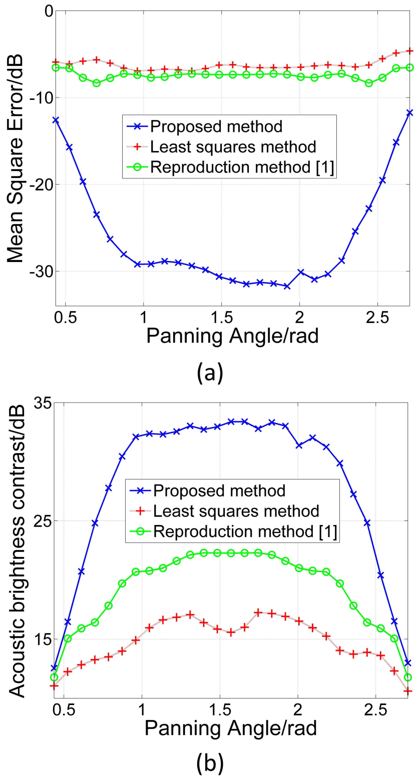

Fig. 4. MZR performance with 15 microphones for each zone when the desired plane wave in the bright zone is panned using different methods. In the reproduction method [1], the regularization parameter is 0.01, and microphones are uniformly distributed on the contour of two zones. Plotted are (a) the NMSE in the bright zone and (b) the acoustic contrast between bright and quiet zones. Each point has been averaged over 50 trial runs.

reproduced sound field using 15 measurements for each zone. The synthesized multi-zone sound field exhibits an acoustic contrast of 28 dB between two zones, whereas the NMSE over the bright zone is -30.4 dB.

Fig. 4 illustrates the MZR performance with 15 microphones in each zone when ϕ_{pw} is panned. We observe that ϕ_{pw} affects the system performance and that the worst performance is achieved when the plane wave is in-line with both zones,while the best performance is obtained when it is perpendicular with the line draw through the centers of the zones. The aforementioned results in Fig. 4 are consistent with the conclusions in [5] [6]. With the same small number of microphones, the proposed method yields up to 20 dB in the NMSE of the bright zone over the least squares approach and the MZR method proposed in [1]. For the acoustic brightness contrast between two zones, the proposed method can provide up to 15 dB improvement over the least squares approach and more than 10 dB over the MZR method in [1]. Fig. 5 (a) illustrates the desired sound field which contains 4 plane waves, and Fig. 5 (b) presents the reproduced sound field using 15 microphones for each zone. The synthesized multi-zone sound field exhibits an acoustic contrast of 25.2 dB between two zones, whereas the NMSE over the bright zone is -24.9 dB.

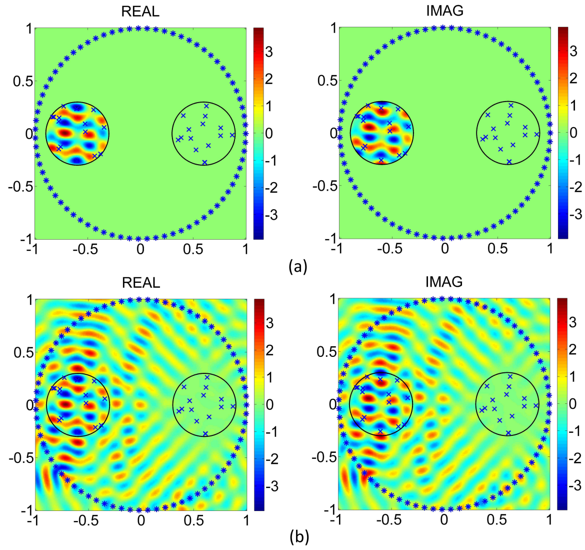

Fig. 5. Reproduction of a 2-D two-zone sound field with 15 randomly distributed microphones for each zone in free field. The desired sound field in the bright zone contains 4 plane waves, which arrives from angle $\pi/4$, $7\pi/18$, $\pi/2$, and $3\pi/4$ respectively, at a frequency of $f = 2000$Hz, whereas the desired sound field in the quiet zone is 0. (a) Desired sound field, and (b) reproduced sound field.

V. Conclusion

In this study, we have proposed a new MZR approach. In the case where only a few far field sound sources are present, the desired sound field can be described as a sparse set of plane waves based on compressed sensing methods. With a small number of microphones, the ℓ_1 norm minimization algorithms enable us to obtain a good MZR performance. For the reproduction of under-sampled sound fields, simulations results demonstrate that the proposed method yields better performance than traditional least-square approach and the MZR method in [1].

Acknowledgment

The authors would like to thank Longshuai Xiao and Zhenhai Yan for the instructive discussions regarding this work. This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDA06040501 and National Natural Science Fund of China under Grant 61501449.

References

[1] Y. J. Wu and T. Abhayapala, “Spatial multizone soundfield reproduction: Theory and design,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1711–1720, August 2011.

[2] M. Poletti and F. Fazi, “An approach to generating two zones of slience with application to personal sound systems,” *Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 598–605, February 2015.

[3] M. Poletti and T. Betlehem, “Creation of a single sound field for multiple listeners,” *INTER-NOISE and INTER-NOISE Congress and Conference Proceedings. Institute of Noise Control Engineering*, 2014.

[4] H. Chen, T. Abhayapala, and W. Zhang, “Enhanced sound field reproduction within prioritized control region,” *INTER-NOISE and INTER-NOISE Congress and Conference Proceedings. Institute of Noise Control Engineering*, 2014.

[5] W. Jin, W. Kleijn, and D. Virette, "Multizone soundfield reproduction using orthogonal basis expansion," *Proc. IEEE Int Conf. Acoust. Speech Signal Process.*, pp. 311–315, 2013.

[6] M. Poletti, "An investigation of 2D multizone surround sound system," *Proc. AES 125th Convention Audio Eng. Society*, 2008.

[7] T. Betlehem and P. Teal, "A constrained optimization approach for multi-zone surround sound," *Proc. IEEE Int Conf. Acoust. Speech Signal Process.*, pp. 437–440, 2011.

[8] P. Teal, T. Betlehem, and M. Poletti, "An algorithm for power constrained holographic reproduction of sound," *Proc. IEEE Int Conf. Acoust. Speech Signal Process.*, pp. 101–104, 2010.

[9] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acostical contrast control constraint," *Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 734–741, February 2014.

[10] P. Coleman, P. Jackson, and M. Olik, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1929–1940, April 2014.

[11] Y. Wu and T. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IIEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 107–116, January 2009.

[12] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming(web page and software),* http://stanford.edu/∼boyd/cvx, June 2009.

[13] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* Academic, New York, 1999.

[14] N. Epain, C. Jin, and A. Schaik, "The appliction of comprehensive sampling to the analysis and synthesis of spatial sound fields," *Proc. AES 127th Convention Audio Eng. Society*, 2009.

[15] W. Jin and W. Kleijn. "Multizone soundfield reproduction in reverberant rooms using compressed sensing techniques," *Proc. IEEE Int Conf. Acoust. Speech Signal Process.*, pp. 4728–4732, 2014.

[16] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[17] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[18] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008