SOUND FIELD ESTIMATION USING TWO SPHERICAL MICROPHONE ARRAYS

Satoru Emura

NTT Media Intelligence Laboratories, NTT Corporation

ABSTRACT

We propose a method of estimating a sound field with two spherical microphone arrays (SMAs). This method estimates plane-wave expansion coefficients of the sound field by using sparse representation modeling in the frequency domain. The dictionary matrix we propose for this modeling achieves the integration of the measurements of two SMAs in a straightforward manner. The effectiveness of the proposed method was evaluated in simulations with computer-generated and measured impulse responses.

Index Terms— Spherical microphone array, plane wave expansion, frequency domain, sparse representation modeling

1. INTRODUCTION

A spherical microphone array (SMA) and so-called binaural synthesis are combined for spatial sound reproduction of an actual 3D auditory scene according to a listener's head movements [1] [2] [3]. A sound field is captured using a SMA and the plane-wave expansion coefficients of the sound field is obtained. From these coefficients and head-related transfer functions (HRTFs), binaural signals are synthesized and reproduced for the listener's left and right ear through a headphone.

SMAs have been studied in recent years because of their ability to analyze sound fields in three dimensions [4] [5]. A rigid SMA, in which microphones are mounted on a rigid baffle, is often preferred to an open SMA because it improves the numerical stability of many processing algorithms [6][7] and its scattering effects are calculable [5] [8] [9]. Later Wu et al. [10] showed that sparse representation modeling [11] [12] is effective for estimating the plane-wave expansion of a sound field when the number of possible plane-wave incident angles is much larger than that of microphones on an SMA.

In binaural synthesis, the coefficients of the plane-wave expansion are manipulated according to the listener's head movements and filtered using HRTFs. By superimposing the filtered coefficients, binaural signals are obtained. Li and Duraiswami proposed a method of manipulating the coefficients according to the rotation of the listener's head [2]. Later Schultz and Spors proposed a method of manipulating the coefficients according to the translatory movements of the listener's head [3].

For synthesizing binaural signals corresponding to large translatory movements, it is desirable that a larger sound field be captured using an SMA. However, the effective area of the sound field captured using a single SMA is actually limited.

We propose a method that involves two rigid SMAs to estimate a larger area of a sound field. This method estimates plane-wave expansion coefficients of the sound field by using sparse representation modeling. In this modeling, the frequency-domain dictionary matrix integrates the measurements by two SMAs in a straightforward manner. The proposed method differs from [13] and[14] in the following points. First, rigid SMAs are used in the proposed method, instead of open SMAs. Second, the proposed method works in the frequency domain, not in the spherical harmonic domain. In Section 2, we review conventional methods for an single SMA. In Section 3 we discuss our proposed method that involves two SMAs. In Section 4 we discuss the evaluation of the proposed method.

2. SINGLE SPHERICAL MICROPHONE ARRAY

Let us consider a unit-magnitude plane wave and an SMA of radius r_a at frequency ω as shown in Fig. 1, where the center of the SMA is at the origin. The incident angle of the plane wave is given as Ω_s in a spherical coordinate. Ω_s denotes a pair of elevation θ_s and azimuth ϕ_s. The sound pressure by this plane wave is expressed at $\mathbf{R} = (x, y, z)$ in the Cartesian coordinate as

$$p(\omega, \mathbf{R}) = e^{i\mathbf{k}_s \bullet \mathbf{R}}, \tag{1}$$

$$\mathbf{k}_s = k \left[\begin{array}{ccc} \sin\theta_s \cos\phi_s, & \sin\theta_s \sin\phi_s, & \cos\theta_s \end{array} \right]^T. \tag{2}$$

The term $k = \omega/c$ is the wave number for frequency ω and speed of sound c.

The sound pressure at (r, Ω) due to this unit-magnitude

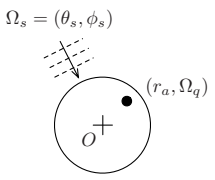

Fig. 1. Plane wave of incident angle Ω_s and sphere.

plane wave and the sphere is expressed as [5] [6][15],

$$p(\omega, r, \Omega) = 4\pi \sum_{l=0}^{\infty} i^l \sum_{m=-l}^{l} b_l(kr) Y_l^{m*}(\Omega_s) Y_l^m(\Omega) \quad (3)$$

$$= \sum_{l=0}^{\infty} i^l \, b_l(kr) \, (2l+1) P_l(\cos \Theta_{\Omega_s,\Omega}), \quad (4)$$

where $i = \sqrt{-1}$ and $Y_l^m()$ is the spherical harmonic function of order l and degree m. $b_l(kr)$ is the so-called mode strength expressed as follows.

$$b_l(kr) = \begin{cases} j_l(kr) & \text{open sphere} \\ j_l(kr) - \frac{j_l'(kr_a)}{h_l^{(1)'}(kr_a)} h_l^{(1)}(kr) & \text{rigid sphere} \end{cases} \quad (5)$$

Here $j_l()$ is the spherical Bessel function of order l. $h_l^{(1)}()$ is the spherical Hankel function of the first kind and of order l. $j_l'()$ and $h_l^{(1)'}()$ denote the their first derivatives. According to the spherical harmonic addition theorem [16, (12.197)], (3) can be rewritten as (4). The $P_l()$ is the Legendre polynomial of degree l. $\Theta(\Omega_s, \Omega)$ is the angle between Ω_s and Ω.

The sound pressure on the sphere of radius r_a is expressed in the spherical harmonic domain (SH domain) as

$$\widetilde{p}_{l,m} = \int_{\Omega \in S^2} p(\omega, r_a, \Omega) \, Y_l^{m*}(\Omega) \, d\Omega, \quad (6)$$

$$= 4\pi i^l \, b_l(kr_a) Y_l^{m*}(\Omega_s). \quad (7)$$

Since, in an actual situation, the sound field is measured by a limited number of microphones on the SMA, (6) is approximated by a summation. The order of $\widetilde{p}_{l,m}$ is truncated to L that satisfies $(L+1)^2 \leq Q$, where Q is the number of microphones. A measurement vector in the frequency domain

$$\mathbf{p}(\omega) = \left[\; p(\omega, r_a, \Omega_1) \quad \cdots \quad p(\omega, r_a, \Omega_Q) \; \right]^T \quad (8)$$

is transformed to a vector of $(L+1)^2$ elements in the SH domain

$$\widetilde{\mathbf{p}}(\omega) = \left[\; \widetilde{p}_{0,0} \quad \widetilde{p}_{1,-1} \quad \widetilde{p}_{1,0} \quad \widetilde{p}_{1,1} \quad \cdots \quad \widetilde{p}_{L,L} \; \right]^T. \quad (9)$$

Let us consider obtaining plane-wave expansion coefficients from $\widetilde{\mathbf{p}}(\omega)$. This problem can be formulated with the compressed sensing (CS) approach [10][11], where $N_D (\gg Q)$ possible plane-wave incident angles $\Omega_1 \cdots \Omega_{N_D}$ are assumed beforehand. This approach attempts to solve an underdetermined problem

$$\widetilde{\mathbf{p}}(\omega) = \widetilde{D}(\omega) \mathbf{a}(\omega), \quad (10)$$

where

$$\widetilde{D}(\omega) = \left[\; \widetilde{\mathbf{u}}(\Omega_1) \quad \cdots \quad \widetilde{\mathbf{u}}(\Omega_{N_D}) \; \right], \quad (11)$$

$$\widetilde{\mathbf{u}}(\Omega) = 4\pi \begin{bmatrix} i^0 \, b_0(kr_a) \, Y_0^{0*}(\Omega) \\ i^1 \, b_1(kr_a) \, Y_1^{-1*}(\Omega) \\ \vdots \\ i^L \, b_L(kr_a) \, Y_L^{L*}(\Omega) \end{bmatrix}. \quad (12)$$

Here $\widetilde{D}(\omega)$ is a dictionary matrix of size $(L+1)^2 \times N_D$. $\mathbf{a}(\omega)$ is a $N_D \times 1$ vector of plane-wave expansion coefficients, where a few elements have non-zero values. This sparse $\mathbf{a}(\omega)$ is obtained solving an ℓ_1 optimization problem.

3. TWO SPHERICAL MICROPHONE ARRAYS

We propose a method of estimating a sound field by using two SMAs (SMA A and SMA B), with which the coefficients of the plane-wave expansion of the sound field are estimated using a dictionary in the frequency domain. The proposed method is based on the assumption that the scattering from one SMA to another SMA can be negligible. The testing of this assumption is discussed in the next section. For this testing, an approximate model of this scattering was also derived.

3.1. Proposed method

For constructing a dictionary matrix for two SMAs, we have to know how the same plane wave is represented on both SMAs. Since considering this representation in the SH domain is not straightforward, we propose to obtain the representation in the frequency domain.

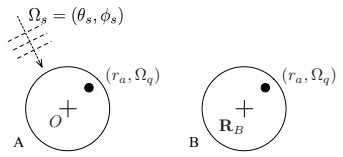

Fig. 2. Plane wave of incident angle Ω_s and two spheres.

Let us assume a plane wave of incident angle $\Omega_s = (\theta_s, \phi_s)$ and SMA A at the origin, as shown in Fig. 2, where the positions of the microphones on SMA A are given by $\mathbf{r}_q = (r_a, \Omega_q)$ $(1 \leq q \leq Q)$. Then, the output of its qth microphone to the incident plane wave of Ω_s is expressed as

$$p_q(\omega, \Omega_s) = 4\pi \sum_{l=0}^{\infty} i^l \sum_{m=-l}^{l} b_l(kr_a)(2l+1) P_l(\cos \Theta(\Omega_s, \Omega_q)). \quad (13)$$

The dictionary matrix in the frequency domain

$$D(\omega) = \begin{bmatrix} p_1(\omega, \Omega_1) & \cdots & p_1(\omega, \Omega_{N_D}) \\ \vdots & & \vdots \\ p_Q(\omega, \Omega_1) & \cdots & p_Q(\omega, \Omega_{N_D}) \end{bmatrix} \quad (14)$$

relates the output of SMA A

$$\mathbf{p}_A(\omega) = \left[\; p_A(\omega, 1) \quad \cdots \quad p_A(\omega, Q) \; \right]^T \quad (15)$$

and plane-wave expansion coefficients $\mathbf{a}(\omega)$ as $\mathbf{p}_A(\omega) = D(\omega)\mathbf{a}(\omega)$.

Next, consider the outputs of SMA B whose center is given as $\mathbf{R_B}$ in the Cartesian coordinate. According to (1), the sound pressure at $\mathbf{R}$ due to the incident plane wave of Ω_s is given as $e^{i\mathbf{k}_s \bullet \mathbf{R}}$. Hence, the sound pressure $p(\omega)$ is 1 at the origin and $p(\omega) = e^{i\mathbf{k}_s \bullet \mathbf{R_B}}$ at $\mathbf{R_B}$. This means that considering the scattered sound field at $\mathbf{R_B}$ is equivalent to considering the scattered sound field at the origin with the phase shift of $e^{i\mathbf{k}_s \bullet \mathbf{R}_B}$. Hence, the output of the qth microphone on SMA B is expressed as

$$p\left(\omega, \Omega_s, \Omega_q\right) = e^{i\mathbf{k}_s \bullet \mathbf{R}_B} p_q(\Omega_s), \tag{16}$$

where $\mathbf{k}_s$ is given by (2). Therefore, the dictionary matrix for SMA B is expressed as

$$D(\omega, \mathbf{R}_B) = \left[\begin{array}{ccc} \mathbf{u}_1 & \cdots & \mathbf{u}_{N_D} \end{array} \right], \tag{17}$$

$$\mathbf{u}_n = e^{i\mathbf{k}_n \bullet \mathbf{R}_B} \left[\begin{array}{c} p_1(\Omega_n) \\ \vdots \\ p_Q(\Omega_n) \end{array} \right] \quad (1 \leq n \leq N_D). \tag{18}$$

The relation between the plane-wave expansion coefficients $\mathbf{a}(\omega)$ and the outputs of SMAs A and B is expressed as

$$\mathbf{p}(\omega) = \left[\begin{array}{c} \mathbf{p}_A(\omega) \\ \mathbf{p}_B(\omega) \end{array} \right] = \left[\begin{array}{c} D(\omega) \\ D(\omega, \mathbf{R}_B) \end{array} \right] \mathbf{a}(\omega). \tag{19}$$

Thus, considering the dictionary matrix in the frequency domain instead of the SH domain, has two advantages. First, $D(\omega, \mathbf{R}_B)$ is obtained easily by using the phase shift $e^{i\mathbf{k}_s \bullet \mathbf{R}_B}$. Second, the order of spherical harmonics of $p_q(\Omega_n)$ in (13) can be chosen freely for constructing the dictionary matrix. This order is not limited by the number of microphones on the SMAs.

With this dictionary matrix, $\mathbf{a}(\omega)$ is obtained by solving the ℓ_1-constraint problem

$$\mathbf{a}(\omega) = \arg\min \left| \left[\begin{array}{c} D(\omega) \\ D(\omega, \mathbf{R}_B) \end{array} \right] \mathbf{a}(\omega) - \left[\begin{array}{c} \mathbf{p}_A(\omega) \\ \mathbf{p}_B(\omega) \end{array} \right] \right|^2 \quad \text{subject to } |\mathbf{a}(\omega)|_1 \leq \gamma. \tag{20}$$

The reason for using this formulation [17] is that the norm of $\mathbf{a}(\omega)$ can be directly controlled. We use γ as the upper limit in the above constraint equation.

When an rigid SMA with radius r_a is set virtually at $\mathbf{R}$, the output from the SMA is synthesized as

$$\mathbf{p}(\omega) = D(\omega, \mathbf{R})\mathbf{a}(\omega). \tag{21}$$

3.2. Scattering between SMAs

We analyzed the scattering from rigid SMA A to rigid SMA B by using a point sound source instead of a plane wave. The reason is that actual sound fields are considered generated by a set of point sound sources.

Consider a unit amplitude point source at $\mathbf{r}_s = (r_s, \Omega_s)$ and the wave from this source to SMA A at the origin. The sound field scattered by rigid SMA A of radius r_a is expressed [15, (8.22)] as

$$p(k, \mathbf{r}, \mathbf{r}_s) = \frac{ik}{4\pi} \sum_{l=0}^{\infty} \frac{j_l'(kr_a)}{h_l'^{(1)}(kr_a)} h_l^{(1)}(kr) \times h_l^{(1)}(kr_s)\,(2l+1)\,P_l\left(\cos\Theta_{\mathbf{r},\mathbf{r_s}}\right). \tag{22}$$

According to [15, (6.68)]

$$h_l^{(1)}(kr) \approx (-i)^{l+1} \frac{e^{ikr}}{kr}, \tag{23}$$

(22) can be approximated as

$$p(k, \mathbf{r}, \mathbf{r_s}) \approx \frac{e^{ikr}}{4\pi r} \frac{4\pi}{k} T(k, r_a, \mathbf{r}, \mathbf{r}_s), \tag{24}$$

$$T(k, r_a, \mathbf{r}, \mathbf{r}_s) = \left\{ \frac{ik}{4\pi} \sum_{l=0}^{\infty} \frac{j_l'(kr_a)}{h_l'^{(1)}(kr_a)} (-i)^{l+1} \times h_l^{(1)}(kr_s)(2l+1) P_l\left(\cos\Theta_{\mathbf{r},\mathbf{r_s}}\right) \right\}. \tag{25}$$

Let $\mathbf{r}$ be the center of SMA B. Then (24) means that the sound field due to the scattering from SMA A to SMA B is approximated by the sound field generated by a point sound source at A with intensity of $4\pi\, T(k, r_a, \mathbf{r}, \mathbf{r}_s)/k$. Hence, the sound pressure on the qth microphone on SMA B due to this scattering wave from SMA A is expressed as

$$p_{sc}(k, \tilde{\mathbf{r}}_q, \tilde{\mathbf{r}}_A) = \frac{ik}{4\pi} \sum_{l=0}^{\infty} b_l(k\tilde{r}_q) h_l^{(1)}(k\tilde{r}_A)(2l+1) P_l\left(\cos\Theta_{\tilde{\mathbf{r}}_q, \tilde{\mathbf{r}}_A}\right) \times \frac{4\pi}{k}\, T(k, r_a, \mathbf{r}, \mathbf{r}_s), \tag{26}$$

where $\tilde{\mathbf{r}}_A$ is the positions of the center of SMA A in the coordinate whose origin is at the center of SMA B, and $\tilde{\mathbf{r}}_q$ is that of the qth microphone on SMA B.

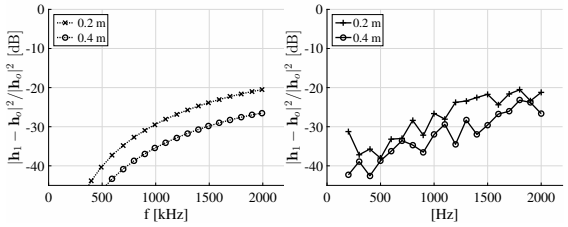

Fig. 3. Difference of sound field captured by SMA A with and without rigid sphere corresponding to SMA B: (a) simulated data (b) measured data.

4. EVALUATION

First, the assumption that the interaction between two SMAs is negligible was confirmed in an anechoic room. We set an SMA of radius 0.042 m and 32 microphones at (x, y, z) =(0 m, 0 m, 1 m) and a loudspeaker at (0 m, 2 m, 1 m). We also set a rigid sphere with the same radius at (x_1 m, 0 m, 1 m). Let $\mathbf{h}_o(\omega)$ and $\mathbf{h}_1(\omega)$ be the vectors of the transfer functions between the loudspeaker and all microphones without and with the rigid sphere. By changing x_1 to 0.2 and 0.4 m, we investigated the interaction between two rigid spheres.

Fig. 3 shows the $|\mathbf{h}_1(\omega) - \mathbf{h}_o(\omega)|^2/|\mathbf{h}_o(\omega)|^2$ of (a) simulated data and (b) measured data. For the simulated data, $\mathbf{h}_o$ was obtained using *SMIR generator* [8] [9]. $\mathbf{h}_1$ was obtained by adding the scattered sound field due to the rigid sphere expressed by (26). For the measured data, we used EigenMike as the SMA. It can be seen that the relative difference was below -20 dB. Hence, the assumption can be considered effective in this setting.

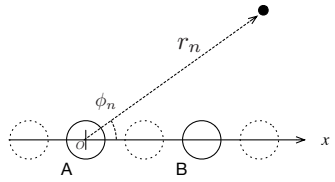

Fig. 4. Arrangement of sound source and rigid SMAs.

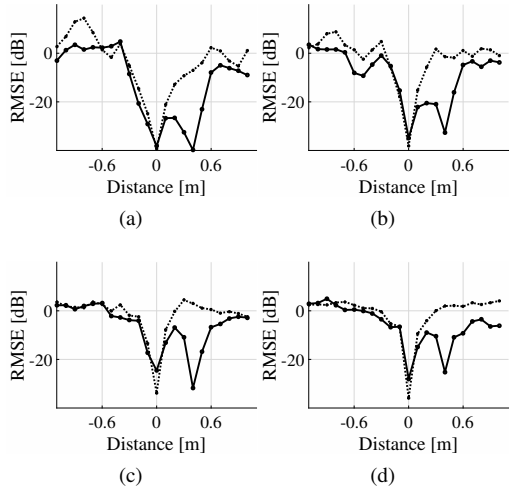

Fig. 5. RMSE of sound field estimation with computer-generated impulse responses by using single SMA (dotted) and two SMAs (solid): (a) 500 Hz, (b) 1000 Hz, (c) 1500 kHz, and (d) 2000 Hz.

Second, we compared the estimation by a single SMA and that by two SMAs. We focused below 2 kHz because phase information in this range is used as a cue to the direction of the sound in human auditory system[18]. The impulse responses were generated as in the first evaluation. The room size was $11 \times 10 \times 3$ m (length×width×height) and its reverberation time was 0.2 s. Signal-to-noise ratio was set to 35 dB. Fig. 4 shows the setting. The positions of sound sources were (r, ϕ)=(3.5 m, 1/6π) and (4 m, 2/3π). The same SMAs as the first evaluation were used. The distance between SMA A and B was set to 0.4 m. We investigated the difference between the output of a single rigid SMA $\mathbf{p}(x, \omega)$ and the estimate from the proposed method $\hat{\mathbf{p}}(x, \omega)$ using root-mean square error (RMSE) $|\mathbf{p}(x, \omega) - \hat{\mathbf{p}}(x, \omega)|^2$ along the $x-$axis. The dictionary matrix was obtained from 642 pre-determined directions. The CVX [19][20] was used for solving ℓ_1 optimization (20). γ was set to $1.8\, max(abs([\mathbf{p}_A(\omega)^T \mathbf{p}_B(\omega)^T]^T))$. Fig. 5 shows the results for f =500, 1000, 1500, and 2000 Hz. At all frequencies, the estimation by using two SMAs was better than that by using a single SMA.

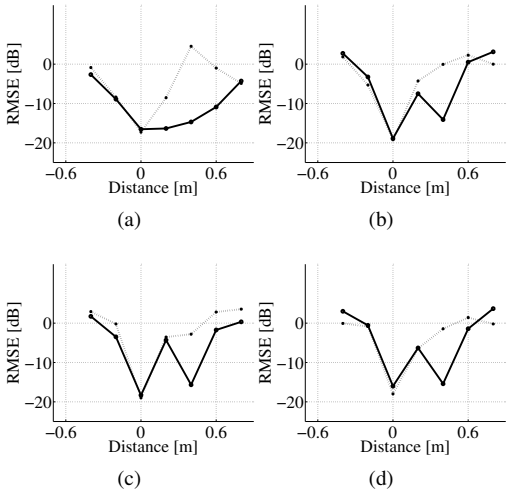

Fig. 6. RMSE of sound field estimation with measured impulse responses by using single SMA (dotted) and two SMAs (solid): (a) 500 Hz, (b) 1000 Hz, (c) 1500 kHz, and (d) 2000 Hz.

Next, we evaluated the proposed method with measured impulse responses by using the EigenMike. The room size was $9 \times 8 \times 3$ m. Its reverberation time was 0.3 s. The positions of sound sources were $(r, \phi) = (2.5\text{ m}, 1/4\pi)$ and $(2.1\text{ m}, 2/3\pi)$. γ was set as in the previous evaluation. Fig. 6 shows the results for f =500, 1000, 1500, 2000 Hz. At all frequencies, the areas of RMSE below 0 dB were larger for the proposed method.

5. CONCLUSION

We proposed a method of estimating a sound field with two rigid SMAs. We show that constructing a dictionary matrix in the frequency domain achieves a straightforward integration of the measurements of two SMAs. Simulations with computer-generated and measured impulse responses showed that the estimation by using two SMAs was better than that by using a single SMA.

6. REFERENCES

[1] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoust. (WASPAA2005)*, 2005.

[2] Z. Li and R. Duraiswami, "Headphone-based reproduction of 3d auditory scenes captured by spherical/hemispherical microphone arrays," in *Proc. ICASSP2006*, 2006, pp. v337–v340.

[3] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head movements," in *Proc. AES 52nd International Conference*, Sept. 2012.

[4] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. ICASSP2002*, 2002, pp. 1949–1952.

[5] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. ICASSP2002*, 2002, pp. 1781–1784.

[6] B. Rafaely, "Analysis and design of spherical microphone arrays,," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[7] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acous. Soc. Am*, vol. 118, no. 5, pp. 3094–3103, 2005.

[8] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, ," in *Proc. ICASSP 2011*, 2011, pp. 128–132.

[9] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acous. Soc. Am.*, vol. 132, pp. 1462–1472, 2012.

[10] P. K. T. Wu, N. Epain, and C. Jin, "A dereverberation algorithm for spherical microphone arrays using compressed sensing techniques," in *Proc. ICASSP2012*, 2012, pp. 4053–4056.

[11] E. J. Cande's and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[12] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[13] P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "3d spatial soundfield recording over large regions," in *Proc. IWAENC2012*, 2012.

[14] F. Wang and X. Pan, "Acoustic sources localization in 3d using multiple shperical arrays," *J. Electr. Eng. Technol.*, pp. 759–768, 2016.

[15] E. G. Williams, *Fourier Acoustics*, Academic, New York, 2000.

[16] G. Arfken, *Mathematical methods for physicists third ed.*, Academic Press Inc., Orlando, 1985.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, 2001.

[18] J. Blauert, *Spatial Hearing: The Psychophysics of human sound localisation*, MIT Press, Cambridge, 1997.

[19] Inc. CVX Research, "CVX: Matlab software for disciplined convex programming," http://cvx.com/cvx, Aug. 2012.

[20] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex program," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Lecture Notes in Control and Information Sciences, pp. 95–100. Springer-Verlag, 2008.