SUPER-RESOLUTION ACOUSTIC IMAGING USING SPARSE RECOVERY WITH SPATIAL PRIMING

Tahereh Noohi, Nicolas Epain, Craig T. Jin

Computing and Audio Research Laboratory (CARLab)
School of Electrical and Information Engineering
The University of Sydney, NSW 2006, Australia

ABSTRACT

In this paper, we propose a new strategy to obtain super-resolution maps of the sound field recorded by a spherical microphone array. In recent works, we have demonstrated that sparse recovery (SR) algorithms based on the minimisation of the l_p norm with $0 < p \leq 1$ can effectively produce super-resolution acoustic maps. The issue with the minimisation of the l_p norm when $p < 1$ is that it is a non-convex optimisation problem, thus it is likely that the algorithm converges to a local minimum. In this paper we show that we can improve the convergence of our SR acoustic imaging methods by providing, to the SR solver, priming information relating to the spatial location of the sound sources. This information can be acquired with a pre-processing, coarse analysis using standard blind source separation or direction-of-arrival techniques. Simulation results indicate that this approach can provide accurate estimates of the positions of multiple, simultaneous sound sources in the presence of noise or reverberation and even in an under-determined situation.

***Index Terms*—** Acoustic imaging, Sparse recovery, Spherical microphone arrays

1. INTRODUCTION

Spherical microphone arrays (SMAs) have attracted considerable interest over the past decade. The main characteristic of spherical microphone arrays is that they provide a *panoramic* point of view on the sound field, which makes them particularly well suited for applications such as beamforming or direction-of-arrival (DOA) estimation. Examples of recent works in this field of research are the implementation of the MUSIC and ESPRIT in the spherical harmonic domain, which are referred to as EB-MUSIC [1] and EB-ESPRIT [2]. Spherical microphone arrays also provide an interesting framework for blind source separation using independent component analysis (ICA) [3].

In a series of recent works [4, 5, 6, 7, 8], we have shown that sparse recovery (SR) techniques can be used to analyse the sound field acquired by SMAs at a resolution beyond that associated with normal beamforming. Sparse recovery techniques are based on the idea that the observed sound field can be explained by only a few dominant sound sources. In other words, the signals recorded by the SMA are decomposed over a dictionary of sound sources and we assume that this decomposition should be sparse across space. The most common approach for SR is the minimisation of the l_p norm of the decomposition, with $0 < p \leq 1$. Minimising the l_1 norm, which is the classic method employed in compressed sensing applications, presents the advantage of being a convex optimisation problem. Minimising the l_p norm with $p < 1$, on the other hand, does not constitute a convex optimisation problem, which means that SR algorithms can then converge to a local minimum. Nevertheless, multiple studies [9, 10] have shown that minimising norms of this kind can be more effective than minimising the l_1 norm because they are closer to a direct measure of sparsity (l_0 norm). As well, convergence issues due to the presence of local minima can be alleviated when prior information is available [11, 12].

In a sound field analysis scenario, prior information regarding the source positions is usually unavailable. However, it is possible to perform a coarse spatial analysis of the sound field and use these data to assist the convergence of the SR sound field analysis. In this paper we investigate the use of ICA and MUSIC to acquire coarse sound field energy maps that are then used to prime an IRLS solver.

The remainder of the paper is organised as follows. Section 2 briefly reviews sound field imaging using SR in the SMA framework and then describes our approach to spatial priming. Section 3 presents the results of the numerical simulation. Lastly, we conclude in Section 4.

2. METHODS

2.1. Acoustic imaging in the SMA framework

We begin by introducing the concept of acoustic imaging in the SMA framework.The acoustic pressure, $p(r, \theta, \phi)$, measured on the surface of a spherical array (open or rigid) with spherical coordinates (r, θ, ϕ) at the frequency f can be modelled as a sum of L spherical harmonic modes [3, 13]:

$$p(r,\theta,\phi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \psi_l(kr) Y_l^m(\theta,\phi) b_{lm}(f) \ , \qquad (1)$$

where k is the wave number given by $k = 2\pi f/c$ where c denotes the speed of sound; $Y_l^m(\theta,\phi)$ denotes the value of the order-l, degree-m real-valued spherical harmonic function for direction (θ,ϕ); $b_{lm}(f)$ is the spherical harmonic expansion coefficient for order l and degree m; $\psi_l(kr)$ for an open sphere and a rigid sphere is given by

$$\psi_l(kr) = \begin{cases} i^l j_l(kr) & \text{open sphere} \\ i^l \left(j_l(kr) - \frac{j_l'(kr_0)}{h_l^{(2)\prime}(kr_0)} h_l^{(2)}(kr) \right) & \text{rigid sphere} \end{cases}$$

where i is the imaginary unit. j_l and $h_l^{(2)}$ denote the order-l spherical Bessel function and spherical Hankel function of the second kind, respectively. j_l', h_l' are their derivatives, and $r_0 \leq r$ is the radius of the rigid sphere.

Equation (1) shows that a sound field can be represented by a set of frequency-domain coefficients $b_{lm}(f)$. The corresponding time-domain signals, $b_{lm}(t)$, are referred to as order-L Higher Order Ambisonic (HOA) signals. In order to produce a map of the incoming sound field we perform a plane-wave decomposition of the HOA signals by solving the equation:

$$\mathbf{b}(n,f) = \mathbf{D}\,\mathbf{x}(n,f) \ , \ \text{where:} \qquad (2)$$

- $\mathbf{b}(n,f)$ is the short-time Fourier transform (STFT) of the vector of HOA signals with time index n and frequency index f:

  $$\mathbf{b}(n,f) = [b_{00}(n,f),\, b_{1-1}(n,f),\, ...,\, b_{LL}(n,f)]^\mathsf{T} \ ,$$

- $\mathbf{D}$ is a matrix representing an arbitrarily chosen set of Q plane-wave directions in the spherical harmonic domain:

  $$\mathbf{D} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_Q] \ ,$$
  $$\mathbf{y}_i = \left[Y_0^0(\theta_i,\phi_i), Y_1^{-1}(\theta_i,\phi_i), ..., Y_L^L(\theta_i,\phi_i)\right]^\mathsf{T} \ ,$$

  where $\mathbf{y}_i$ is a vector giving the spherical harmonic components of a plane-wave for direction (θ_i,ϕ_i). We refer to $\mathbf{D}$ as the *dictionary matrix* which we use to decompose the sound field.

- $\mathbf{x}(n,f)$ is the STFT of the vector of Q plane-wave signals:

  $$\mathbf{x}(n,f) = [x_1(n,f), x_2(n,f), ..., x_Q(n,f)]^\mathsf{T} \ .$$

Once the plane-wave signal decomposition has been obtained, we can form an acoustic energy map which shows the incoming acoustic energy for each direction in space by calculating the energy, $e(\theta_i,\phi_i)$, for each plane-wave direction in the dictionary:

$$e(\theta_i,\phi_i) = \sum_{n=1}^{N} \sum_{f=1}^{F} |x_i(n,f)|^2. \qquad (3)$$

Clearly, the resolution of the acoustic energy map depends on the resolution of the plane-wave dictionary.

2.2. Sparse plane-wave decomposition

Our approach to the plane-wave decomposition consists of looking for the *sparsest* plane-wave decomposition that solves Eq. (2). This sparse plane-wave decomposition can be determined by solving the following optimization problem:

$$\text{minimise } \|\mathbf{x}(n,f)\|_p \ \text{ subject to } \ \mathbf{b}(n,f) = \mathbf{D}\mathbf{x}(n,f) \ , \ (4)$$

where $\|\cdot\|_p$ denotes the l_p norm of a vector, defined by:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{Q} |x_i|^p \right)^{\frac{1}{p}} \ , \qquad (5)$$

and $0 < p \leq 1$. Problem (4) with $p = 1$ is the classic basis pursuit problem found in compressed sensing [11]. The advantage of this formulation is that it is a convex problem, *i.e.*, convergence to a global minimum is guaranteed if an appropriate solver is used. Choosing $p < 1$ may lead to faster convergence and/or make it possible to recover signals with less measurements compared to the $p = 1$ case [9, 11]. However, when $p < 1$ Problem (4) is non-convex and convergence to a global minimum is no longer guaranteed.

In the following we solve Problem (4) for $p < 1$ using the iteratively reweighed least squares (IRLS) algorithm [9]. More specifically, the two following steps are repeated until convergence:

1. Compute: $\mathbf{x}(n,f) = \mathbf{\Omega}\mathbf{D}^\mathsf{T}\left(\mathbf{D}\mathbf{\Omega}\mathbf{D}^\mathsf{T} + \alpha\mathbf{I}\right)^{-1}\mathbf{b}(n,f)$,
2. Compute: $\omega_i = \left(|x_i(n,f)|^2 + \mu\right)^{\frac{2-p}{2}}$, (6)

where the weights ω_i are initialised to 1, $\mathbf{\Omega}$ is the diagonal matrix with entries ω_i, α is a regularization factor and μ is a small-valued parameter ensuring that $\mathbf{\Omega}$ is defined when one of the x_i is equal to 0 [14].

2.3. Sparse recovery with priming

In [11] and [12], the authors demonstrate that the accuracy of sparse recovery can be improved when prior information regarding the support of the solution is available. More specifically, knowing which one or more element(s) of the solution vector are non-zero makes it possible to better reconstruct the signal vector with less observations. In the method presented in [11], the weights ω_i in the IRLS algorithm (Eq. (6)) corresponding to the *a priori* non-zero elements of the solution vector are multiplied by a small constant. In this paper we investigate the use of a similar priming method for sound field imaging. In order to improve the convergence of the IRLS algorithm in the context of sound field imaging, we replace Problem (4) by the following optimisation problem:

$$\text{minimise } \|\mathbf{x}'(n,f)\|_p \ \text{subject to } \ \mathbf{b}(n,f) = \mathbf{D}\mathbf{W}\mathbf{x}'(n,f) \ , \qquad (7)$$

where $\mathbf{W}$ is a weighting matrix, $\mathbf{W}=\text{diag}\left([w_1, w_2, \ldots, w_Q]\right)$, $w_i = \bar{e}(\theta_i, \phi_i)$ where $\bar{e}$ is a low-resolution energy map. The weight w_i expresses the likelihood or probability of finding an active plane wave source in the corresponding direction: the more likely the existence of the plane-wave source, the larger is the corresponding weight. In the typical sound field imaging scenario, no prior information regarding the positions of the sources is available. In order to set the values of the priming weights, w_i, we first obtain a low-resolution acoustic energy map of the sound field. The normalised energy values, $\bar{e}$, corresponding to the low-resolution energy map are obtained by dividing each energy value in the map by the maximum energy value. Once the solution to Problem (7) is found, the plane-wave solution is obtained as: $\mathbf{x}(n, f) = \mathbf{W}\mathbf{x}'(n, f)$.

We now briefly describe two methods for calculating low-resolution acoustic energy maps. For both methods, the HOA signals are processed in the time domain. A sound field energy map is calculated for each time frame and the energy maps corresponding to each time frame are added together to obtain the final low-resolution energy map.

2.3.1. Spatial priming using ICA

We briefly review the linear ICA technique described in [3] for obtaining a low-resolution acoustic energy map. In the first step, one applies linear ICA to the $(L+1)^2$ time-domain HOA signals in a given time frame, to obtain a set of $(L+1)^2$ separated signals, $s_j(t)$, and a mixing matrix, $\mathbf{A}$. In the second step, the mixing matrix $\mathbf{A}$ is analysed to determine the directions corresponding to each separated signal. For the j-th signal, the estimated source direction is calculated as $(\hat{\theta}_j, \hat{\phi}_j) = (\theta_{q(j)}, \phi_{q(j)})$ where

$$q(j) = \underset{i}{\text{argmax}}\, C_{ij} \;\; \text{with} \;\; C_{ij} = \frac{\mathbf{y}_i^{\mathsf{T}} \mathbf{a}_j}{\|\mathbf{y}_i\| \, \|\mathbf{a}_j\|} \;, \qquad (8)$$

and $\mathbf{a}_j$ denotes the j-th column of matrix $\mathbf{A}$ and C_{ij} is the correlation between vectors $\mathbf{y}_i$ and $\mathbf{a}_j$. Because real sources should show a large correlation with at least one direction in space, source directions are only considered valid when the maximum correlation value is greater than a given threshold, *e.g.*, 0.9. The ICA acoustic energy map is determined as

$$e_{\text{ICA}}(\theta_i, \phi_i) = \begin{cases} \sum_t s_j(t)^2 & i = q(j), \;\; C_{ij} > 0.9 \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

2.3.2. Spatial priming using MUSIC

Another possible method to obtain the priming weights is to use the MUSIC algorithm. The MUSIC energy value for direction i is given by:

$$e_{\text{MUSIC}}(\theta_i, \phi_i) = \frac{1}{\mathbf{y}_i^{\mathsf{T}} \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{\mathsf{T}} \mathbf{y}_i} \;, \qquad (10)$$

where $\mathbf{V}$ is the matrix of the eigenvectors of the HOA signal correlation matrix which are sorted in descending order based on their corresponding eigenvalues, and $\mathbf{\Lambda}$ is the diagonal matrix given by:

$$\mathbf{\Lambda} = \text{diag}([\lambda_1, \lambda_2, ..., \lambda_{(L+1)^2}]) \;, \;\; \lambda_i = \begin{cases} 0 & \text{for } i \leq K \\ 1 & \text{otherwise} \end{cases} . \qquad (11)$$

The MUSIC algorithm requires that the estimated number of sources, K, be set beforehand. We experimentally set $K = 3$ as it provided reasonable results.

3. NUMERICAL SIMULATIONS

This section presents simulation results illustrating the proposed approach. In our simulations, the sound field is measured using an SMA providing HOA signals up to order 2. The SMA consists of two concentric spherical arrays of omnidirectional microphones: 12 microphones are located on the surface of a rigid sphere with a radius of 3 cm; another 12 microphones are located on the surface of an open sphere with a radius of 15 cm. The dictionary $\mathbf{D}$ consists of 642 directions obtained by successively subdividing the faces of an icosahedron. The parameters used for the IRLS algorithm were set to $p = 0.7$ and $\alpha = \frac{1}{642}\frac{r}{1-r}\text{tr}(\mathbf{D}\mathbf{\Omega}\mathbf{D}^{\mathsf{T}})$ with $r = 0.01$. Lastly the ICA is performed using the FastICA package for MATLAB [15].

3.1. Anechoic Simulation

We first consider an anechoic scenario. In this scenario, the sound field recorded by the microphones originates from twelve spherical sources located at a distance of 2 m. The source directions relative to the SMA centre are indicated in Figure 1. The signals emitted by the sources are speech signals, approximately 3 s long. The presence of measurement noise in the HOA signals is modelled by adding uncorrelated Gaussian white noise (signal-to-noise ratio (SNR)= 20 dB). Note that this scenario is an under-determined problem because there are 12 sources and only 9 HOA signals.

Figure 1 shows the energy maps obtained with the following methods: a) Sparse recovery; b) SR with MUSIC-based spatial priming and c) SR with ICA-based spatial priming. Clearly, the map obtained using SR alone highlights the general location of the source positions, but does not allow to locate the sources accurately. This is due to the presence of measurement noise and to the large number of sources, which make the sound field non sparse. Compared to the map obtained with SR alone, the maps obtained using both spatial priming methods are very accurate, showing energy peaks almost only in the source directions. Note that the method using MUSIC is slightly less precise than that based on ICA as the spots corresponding to the sources are slightly larger and there is one spurious spot located at approximately (0°,75°). In Table1, the error in estimated angular position for different SNR values is shown for SR-alone and SR with ICA-based spatial priming algorithms. The SNR values are defined as the

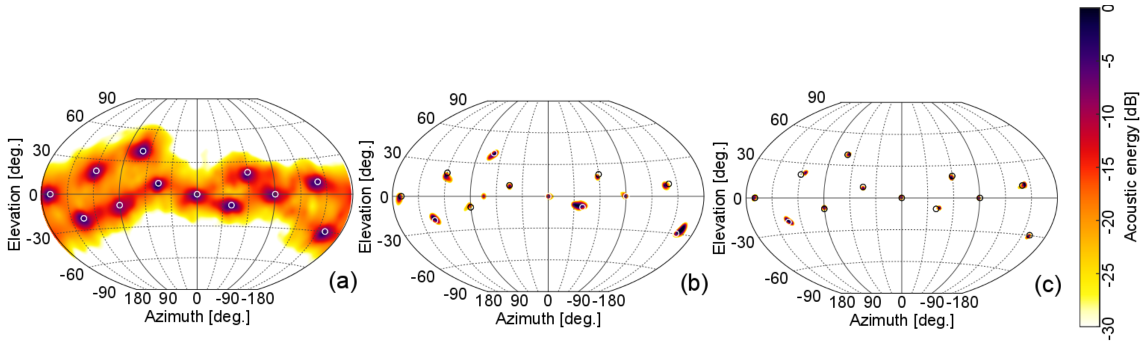

Fig. 1. This figure shows the acoustic energy maps for an anechoic room obtained by: a) SR alone; b) SR with spatial priming using MUSIC; c) SR with spatial priming using ICA. The true source positions are indicated by the circles.

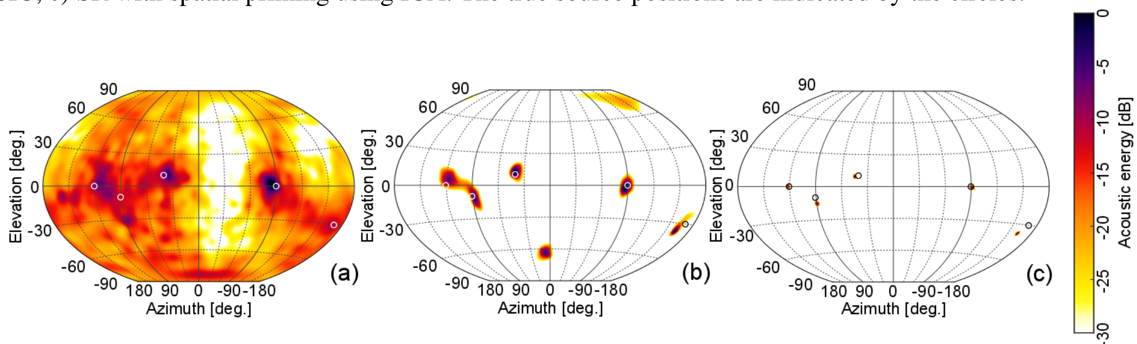

Fig. 2. This figure shows the acoustic energy maps for a reverberant room with T60=250 ms obtained by: a) SR alone; b) SR with spatial priming using MUSIC; c) SR with spatial priming using ICA.

Table 1. This table shows the angular error for different SNR values in the anechoic room (the target source position at (0°,170°)).

SNR [dB]		-16	-13	-10	-8	-5	-2	0
Angular error [deg]	SR alone	missed	missed	12.8	12.4	12.6	9.5	3.9
	SR with spatial priming using ICA	missed	2.1	2.1	2.1	2.1	2.1	2.1

ratio of target source power at position (0°,170°) to power of all interferences (other sources and measurement noise). Note that the value of 2.1 degrees (obtained using spatial priming) corresponds to the angular distance between the target and the closest dictionary direction.

3.2. Reverberant Simulation

We now consider a scenario whereby the SMA and the sources are placed in a room with dimensions 14m×10m×3m. Five spherical sources surround the SMA at a distance of 2 m. The source directions relative to the SMA centre are indicated in Figure 2. The impulse responses between the sources and the microphones were calculated using the MCRoomSim software [16]. The T60 reverberation time of the room is approximately 250 ms and the average signal-to-reverberant ratio (SRR) is 3.7 dB. Similar to the anechoic scenario, the source signals consist of speech and are approximately 3 s long. Lastly, uncorrelated Gaussian white noise is added to the microphone signals (SNR = 40 dB). In order to alleviate the effect of reverberation, we separate the HOA signals into a directive and a diffuse component using a technique described in [8, 17]. The SR plane-wave decomposition is then applied to the directive components.

Figure 2 shows the acoustic energy maps obtained using the different techniques for the reverberant scenario. Similar to the anechoic case, the method employing SR with no spatial priming fails at localising the sources accurately. This is due to the presence of reverberation, which was not totally suppressed by the direct/diffuse separation. The map obtained using MUSIC-based spatial priming is much more accurate, the energy being distributed more precisely around the sources. However, two spurious energy spots are present in this map. Lastly, the map obtained using ICA-based spatial priming is the most precise of the three. In this map only one plane-wave direction is found to be active for each source, this direction being very close to the true source direction.

4. CONCLUSION

In this paper we investigated the use of spatial priming for sound field imaging using sparse-recovery methods. Spatial priming consists in weighting the dictionary used for the SR analysis with a diagonal weighting matrix based on the probability of the existence of a source in the given direction. Simulations results obtained for an anechoic and a reverberant environment show that these techniques improve the accuracy of the acoustic imaging, especially in the presence of reverberation.

5. REFERENCES

[1] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication: Challenges and Perspectives*, Sharon Gannot Israel Cohen, Jacob Benesty, Ed., chapter 11, pp. 281–305. Springer, Berlin, 2010.

[2] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," in *Proceedings of the ICASSP 2011*, Prague, Czech Republic, May 2011, pp. 117–120.

[3] N. Epain and C. T. Jin, "Independent component analysis using spherical microphone arrays," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 91–102, 2012.

[4] A. Wabnitz, N. Epain, A. McEwan, and C.T. Jin, "Upscaling ambisonic sound scenes using compressed sensing techniques," in *Proceedings of WASPAA*, New Paltz, NY, USA, October 2011.

[5] T. Wu, N. Epain, and C. T. Jin, "A dereverberation algorithm for spherical microphone arrays using compressed sensing techniques," in *Proceedings of the ICASSP 2012*, Kyoto, Japan, March 2012.

[6] P.K.T. Wu, N. Epain, and C. T. Jin, "Super-resolution beamforming algorithm for spherical microphone arrays using a compressed sensing approach," in *Proceedings of the ICASSP 2013*, Vancouver, Canada, May 2013.

[7] T. Noohi, N. Epain, and C. T. Jin, "Direction of arrival estimation for spherical microphone arrays by combination of independent component analysis and sparse recovery," in *Proceedings of the ICASSP 2013*, Vancouver, Canada, May 2013.

[8] N. Epain and C.T. Jin, "Super-resolution sound field imaging with sub-space pre-processing," in *Proceedings of the 2013 ICASSP*, Vancouver, Canada, May 2013.

[9] Rick Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 3869–3872, 2008.

[10] Rick Chartrand, Emil Y. Sidky, and Xiaochuan Pan, "Nonconvex compressive sensing for x-ray CT: an algorithm comparison," in *Proceedings of the 2013 Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, November 2013, pp. 665–669.

[11] C.J. Miosso, R. Von Borries, M. Argaez, L. Velazquez, C. Quintero, and C. M. Potes, "Compressive sensing reconstruction with prior information by iteratively reweighted least-squares," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2424–2431, June 2009.

[12] J. Garcia-Frias and I. Esnaola, "Exploiting prior knowledge in the recovery of signals from noisy random projections," in *Proceedings of the 2007 Data Compression Conf. (DCC '07)*, March 2007.

[13] E.G. Williams, *Fourier Acoustics*, Academic Press, San Diego, 1999.

[14] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[15] H. Gavert, J. Hurri, J. Sarela, and A. Hyvarinen, "The FastICA package for MATLAB," `http://www.cis.hut.fi/projects/ica/fastica/`.

[16] A. Wabnitz, N. Epain, C. T. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *In Proceedings of the International Symposium on Room Acoustics*, Melbourne,Australia, August 2010.

[17] N. Epain, T. Noohi, and C. T. Jin, "Sparse recovery method for dereverberation," in *Proceedings of the REVERB Challenge (available online)*, Florence, Italy, May 2014.