SUPER-RESOLUTION SOUND FIELD IMAGING WITH SUB-SPACE PRE-PROCESSING

N. Epain, C.T. Jin

CARLab, School of Electrical and Information Engineering
The University of Sydney
Sydney, Australia

ABSTRACT

Spherical microphone arrays are a powerful tool for sound field analysis. In previous work, we have shown that sparse recovery can be used to arbitrarily increase the resolution of the sound field recorded by a spherical microphone array. Because these super-resolution techniques rely on the assumption that the sound field results from a few dominant plane waves, they are not robust to the presence of noise or reverberation. In this paper we propose a simple method to separate the sound field into a directional component and a diffuse component prior to applying sparse recovery techniques. Simulations show that this pre-processing could dramatically improve the results of sparse recovery in noisy or reverberant environments.

***Index Terms*—** Acoustic imaging, microphone arrays, sparse recovery, subspace methods.

1. INTRODUCTION

Spherical Microphone Arrays (SMAs) have become increasingly popular during the past decade. One of the unique advantages of SMAs is they provide a uniform panoramic view of the sound field. This capability motivates their use for the spatial analysis of sound fields. In [1] and [2], for instance, standard beamforming methods are employed to provide an energy map of the sound field recorded by an SMA. Recently, more elaborate techniques such as EB-MUSIC [3] or EB-ESPRIT [4] (spherical harmonic domain implementations of MUSIC [5] and ESPRIT [6]) have been applied to sound field imaging.

The issue with the above sound field analysis methods is that they are bounded by the SMA's intrinsic resolution. In practice, SMAs are typically comprised of a few dozen microphones. Thus the sound field images obtained from SMA recordings have limited resolution. In recent work [7, 8, 9], techniques based on sparse recovery (SR) have been proposed that enable an increase in the spatial resolution of the sound field description. These techniques are based on the hypothesis that the sound field can be approximated by a few dominant plane-wave sources. Although this is a fair assumption in simple free-field scenarios, it is clearly not true in the presence of noise or reverberation.

In this paper, we present a simple pre-processing method that improves the robustness of SR-based methods in the presence of noise or reverberation. In Sections 2 and 3 we briefly review sound field imaging using SMAs. In Section 4, we describe the proposed pre-processing method. Lastly, in Section 5, we present the results of numerical simulations demonstrating the performance of the proposed algorithm.

2. SOUND FIELD IMAGING WITH SPHERICAL MICROPHONE ARRAYS

In the SMA framework, the signals recorded by the microphones are transformed into a finite-order spherical harmonic representation of the sound field. We refer to the up-to-order-L spherical harmonic domain signals as the order-L Higher Order Ambisonic (HOA) signals. The HOA signals offer a continuous representation of the sound field, the resolution of which depends on the order: the higher the order, the finer the resolution. We consider the matrix, $\mathbf{B}_L$, of the order-L HOA signals with sample length N:

$$\begin{aligned} \mathbf{B}_L &= \left[\mathbf{b}_{0,0}, \mathbf{b}_{1,-1}, ..., \mathbf{b}_{L,L}\right]^\mathsf{T} , \\ \mathbf{b}_{l,m} &= \left[b_{l,m}(1), b_{l,m}(2), ..., b_{l,m}(N)\right]^\mathsf{T} , \end{aligned} \tag{1}$$

where $b_{l,m}(n)$ denotes the n-th time sample of the order-l, degree-m HOA signal.

Sound field imaging with an SMA consists in representing the acoustic energy as a function of the incoming direction. The traditional approach to estimate the energy corresponding to a particular direction consists in steering a beam towards this direction and calculating the energy of the beamformer's output [1, 2]. We estimate the energy, $e_L(u)$, incoming from direction (θ_u, ϕ_u) as:

$$e_L(u) = \left\| \frac{1}{L+1}\, \mathbf{y}_L(u)^\mathsf{T}\, \mathbf{B}_L \right\|^2 , \tag{2}$$

where $\mathbf{y}_L(u)$ is the vector of the spherical harmonic function values, up to order L, in the direction (θ_u, ϕ_u):

$$\mathbf{y}_L(u) = \left[Y_0^0(\theta_u, \phi_u), Y_1^{-1}(\theta_u, \phi_u), ..., Y_L^L(\theta_u, \phi_u)\right]^\mathsf{T} , \tag{3}$$

and Y_l^m denotes the order-l, degree-m real-valued spherical harmonic function.

978-1-4799-0356-6/13/$31.00 ©2013 IEEE

3. SUPER-RESOLUTION IMAGING

In this section, we describe super-resolution imaging using SMAs. Practically, SMAs are limited in the order of the recorded HOA signals (typically up to order 3 or 4) and this in turn limits their resolution. As shown in previous work [7, 8, 9], the resolution of the sound field can be increased using SR techniques. In the SR approach, we perform a plane-wave decomposition of the HOA signals. In other words a matrix of plane-wave signals, $\mathbf{X}$, is calculated such that:

$$\mathbf{D}_L\mathbf{X} = \mathbf{B}_L \ , \tag{4}$$

where $\mathbf{X}$ is defined similarly to $\mathbf{B}_L$:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_U]^{\mathsf{T}} \ , \tag{5}$$

and $\mathbf{D}_L$ is the matrix expressing the contribution of each plane-wave to the HOA signals:

$$\mathbf{D}_L = [\mathbf{y}_L(1), \mathbf{y}_L(2), ..., \mathbf{y}_L(U)] \ , \tag{6}$$

where U is the number of plane wave directions. We refer to $\mathbf{D}_L$ as the plane-wave dictionary, where the number of entries, U, determines the resolution. As the aim is to obtain a spatially sharp description of the sound field, the dictionary is chosen with a number of directions far greater than the number of HOA signals. In this paper we use a dictionary of 2562 directions derived by repeated subdivisions of the faces of an icosahedron.

The classical method for solving Equation (4) is to select the solution with the least-square norm. The problem with this solution is that it is also the one that distributes the energy the most evenly across directions, which is physically incompatible with the assumption of discrete sound sources, resulting in a blurred image. The SR approach, on the other hand, tries to select the solution with the fewest active plane waves, the *sparsest* solution. In practice we obtain such a solution by solving the following optimization problem:

$$\text{minimize} \quad \|\mathbf{X}\|_{12} \quad \text{subject to} \quad \mathbf{D}_L\mathbf{X} = \mathbf{B}_L \ , \tag{7}$$

where $\|\cdot\|_{12}$ denotes the $L12$-norm, defined by:

$$\|\mathbf{X}\|_{12} = \sum_{u=1}^{U} \sqrt{\sum_{n=1}^{N} x_u(n)^2} \ . \tag{8}$$

In this paper we solve Problem (7) using an Iteratively-Reweighted Least-Square (IRLS) algorithm [10]. Solving Problem (7) results in a sparse set of signals and their corresponding directions selected from the plane-wave dictionary. These signals can then be used to estimate the order-L' HOA signals with $L' > L$. We refer to this operation as *upscaling* the HOA signals. Upscaling the signals to order L' is done by simply multiplying the plane-wave signals by the order-L' dictionary, $\mathbf{D}_{L'}$, defined similarly as $\mathbf{D}_L$. In other words, the matrix of the upscaled HOA signals, $\mathbf{B}_{L\rightarrow L'}$ is given by:

$$\mathbf{B}_{L\rightarrow L'} = \mathbf{D}_{L'}\mathbf{X} \tag{9}$$

A *super-resolution* acoustic energy map can then be estimated from the upscaled HOA signals, using the technique described in Section 2 (Equation 2).

4. SUPER RESOLUTION IMAGING WITH SUBSPACE PREPROCESSING

The SR approach presented in the previous section is based on the assumption that the sound field is sparse when expressed using a plane wave basis. However, recorded HOA signals are always polluted by measurement noise which is *not* sparse in the plane-wave domain. Further, the sound field may consist of a mixture of a few spatially discrete sources, together with a diffuse sound field comprising reverberated waves.

In order to make the super-resolution imaging more robust in the presence of diffuse sound or measurement noise, we propose to separate the HOA signals into two parts, a "directional" and "diffuse" component, prior to the imaging. Our approach for extracting the diffuse component of the sound field is to project the HOA signals onto a subspace that is orthogonal (or mostly orthogonal) to the directional component.

We first consider the correlation matrix of the order-L HOA signals, $\mathbf{C}_L$, given by:

$$\mathbf{C}_L = \mathbf{B}_L\ \mathbf{B}_L^{\mathsf{T}} \ . \tag{10}$$

This matrix can be decomposed in terms of its eigenvalues and eigenvectors:

$$\mathbf{C}_L = \mathbf{V}\ \mathbf{S}\ \mathbf{V}^{\mathsf{T}} \ , \tag{11}$$

where $\mathbf{V}$ is the matrix of the eigenvectors and $\mathbf{S}$ is the diagonal matrix whose diagonal coefficients are the eigenvalues:

$$\begin{aligned} &\mathbf{S} = \mathrm{diag}\left(\left[s_1, s_2, s_3, ..., s_{(L+1)^2}\right]\right) \\ &\text{where} \quad s_1 \geq s_2 \geq s_3 \geq ... \geq s_{(L+1)^2} \ . \end{aligned} \tag{12}$$

We define the "diffuse separation" matrix, $\mathbf{A}$, as:

$$\mathbf{A} = \mathbf{V}\ \boldsymbol{\Gamma}\ \mathbf{V}^{\mathsf{T}} \ , \tag{13}$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix of weights comprised between 0 and 1:

$$\begin{aligned} &\boldsymbol{\Gamma} = \mathrm{diag}\left(\left[\gamma_1, \gamma_2, \gamma_3, ..., \gamma_{(L+1)^2}\right]\right) \\ &\text{where} \ \ \gamma_i = \sin\left(\frac{\pi s_{(L+1)^2}}{2 s_i}\right)^2 \ . \end{aligned} \tag{14}$$

The effect of $\mathbf{A}$ operating on the HOA signals is to: 1) Project the HOA signals onto the eigenvectors; 2) Weight the obtained signals, such that the diffuse components of the sound field are preserved, while the direct components are discarded; 3) Return to the HOA domain. The idea here is to compare the

Table 1. This table shows the position of the sound sources in the numerical simulation.

Source #	1	2	3	4	5	6
azimuth (°)	-170	0	0	20	70	140
elevation (°)	20	0	60	10	0	-10

various eigenvalues against the smallest eigenvalue, $s_{(L+1)^2}$, because the small eigenvalues correspond mostly to diffuse or noise components, whereas large eigenvalues correspond to the directional components. Note that, unlike in the MUSIC algorithm [5], no assumption is made on the number of sound sources present in the directional component of the sound field.

We now describe the method for super-resolution imaging with subspace pre-processing. The first step is to extract the diffuse component, $\mathbf{B}_L^{(\mathrm{dif})}$, and directional component, $\mathbf{B}_L^{(\mathrm{dir})}$, of the HOA signals as shown below:

$$\begin{aligned} \mathbf{B}_L^{(\mathrm{dif})} &= \mathbf{A}\,\mathbf{B}_L \,, \\ \mathbf{B}_L^{(\mathrm{dir})} &= (\mathbf{I}-\mathbf{A})\ \mathbf{B}_L \,. \end{aligned} \tag{15}$$

The signals in $\mathbf{B}_L^{(\mathrm{dir})}$ contain less noise or diffuse sound than the original signals and thus are sparser in the plane wave source domain. We then process the directional and diffuse components separately. The second step is to apply the SR method described in Section 3 to obtain the upscaled signals $\mathbf{B}_{L\to L'}^{(\mathrm{dir})}$. The third step is to form a composite energy map comprised of a super-resolution map for the directional component and a low-resolution map for the diffuse component. The energy of the composite map for direction (θ_u, ϕ_u), $\hat{e}_{L\to L'}(u)$, is given by:

$$\hat{e}_{L\to L'}(u) = e_{L\to L'}^{(\mathrm{dir})}(u) + e_L^{(\mathrm{dif})}(u) \,. \tag{16}$$

where $e_{L\to L'}^{(\mathrm{dir})}(u)$ and $e_L^{(\mathrm{dif})}(u)$ are the energies of the directional (order-L') and diffuse (order-L) maps, obtained as described in Section 2 (Equation 2). Note that the diffuse component can be excluded when it is expected to contain mostly measurement noise. However it may be useful to analyze this component in situations where reverberation or other diffuse sounds are present.

5. NUMERICAL SIMULATIONS

In this section we present numerical simulation results validating the imaging method described above. In the simulation, we consider a sound field consisting of 6 plane waves with a diffuse background. The plane waves are incoming from the directions shown in Table 1. The corresponding waveforms are uncorrelated Gaussian white noise signals with equal energy. The diffuse component of the sound field is composed by a very large number of plane-wave, Gaussian, white noise signals with equal energies, that are evenly distributed across space. As ground truth we use the sound field described by $\mathbf{B}_{40}$, the matrix of order-40 HOA signals:

$$\mathbf{B}_{40} = \mathbf{B}_{40}^{(\mathrm{dir})} + \mathbf{B}_{40}^{(\mathrm{dif})} \,, \tag{17}$$

where $\mathbf{B}_{40}^{(\mathrm{dir})}$ and $\mathbf{B}_{40}^{(\mathrm{dif})}$ denote the matrices of the HOA signals corresponding to the 6 plane waves and the diffuse background, respectively. The HOA signals are 1024-sample long. As well, the energies of the diffuse and directional components of the sound field are equal, which corresponds to a signal-to-reverberant ratio of 0 dB if we assume the diffuse background results from reverberation. The energy map for this sound field, calculated using the method described in Section 2, is shown in Figure 1(a).

We now assume that this sound field is measured using an SMA providing the exact HOA signals up to order 2. Note that, in practice, HOA signals acquired using an SMA are noisy at low and high frequencies. Nevertheless, band-pass filtering can be applied to extract the signals corresponding to the frequency range of operation of the SMA [11]. The order-2 energy map of the sound field is shown in Figure 1(b). This map has limited resolution as expected. Using the method described in Section 3, the order-2 HOA signals were upscaled to order 40 to provide a high resolution map. This map is shown in Figure 1(c). Clearly, the upscaling failed due to the presence of a non-sparse sound field. The IRLS solver found sound sources in directions where there where none. This occurs because spurious plane-waves are used to explain the signals originating from the diffuse background. Lastly, Figure 1(d) shows the composite map obtained using the method described in Section 4. Compared to the map presented in Figure 1(c), this map matches the reference map shown in Figure 1(a) much more precisely. Although sources 2 and 4 (see Table 1) were merged into one source, the other plane-wave sources were localized with a reasonable accuracy. As well, the energy of the identified sources is approximately that of the actual sources. Regarding the diffuse part of the sound field, the effect of the subspace pre-processing is clearly visible in that some of the diffuse background energy is missing around the dominant sources. However, the energy of the diffuse background is approximately that of the actual sound field in the directions where there is no dominant source.

As the method used for separating the diffuse and directional components of the sound field is closely related to MUSIC or, more specifically, EB-MUSIC [3], we compare the above results with the MUSIC spatial spectrum. The MUSIC spatial spectrum corresponding to the order-2 HOA signals is given by:

$$\phi(u) = \frac{1}{\mathbf{y}_2(u)^{\mathsf{T}}\,\mathbf{V}\,\mathbf{\Psi}\,\mathbf{V}^{\mathsf{T}}\,\mathbf{y}_2(u)} \,, \tag{18}$$

where $\mathbf{V}$ is the matrix of the correlation matrix's eigenvectors, as defined in Equation (13), and $\mathbf{\Psi}$ is the diagonal matrix given

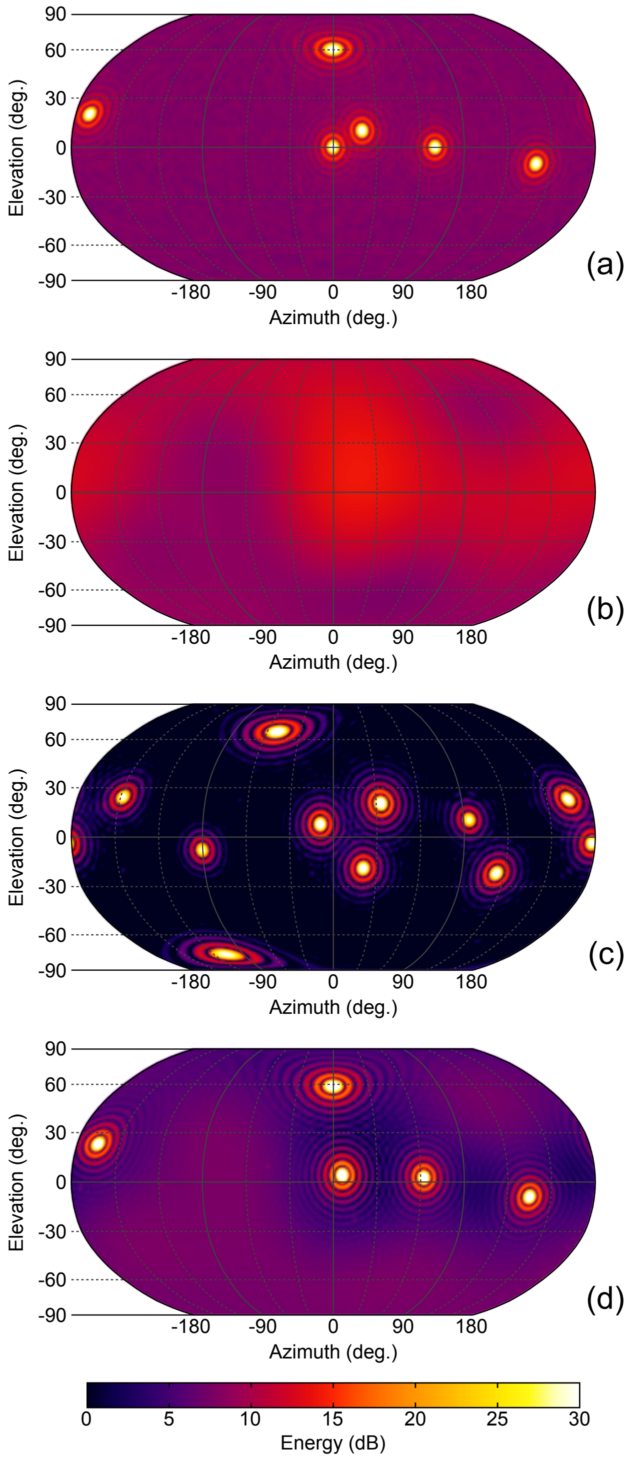

Fig. 1. This figure shows the different energy maps obtained in the numerical simulations: a) the true order-40 map; b) the order-2 map; c) the order-40 map obtained from the upscaled HOA signals without pre-processing; d) the composite map obtained using the method described in Section 4.

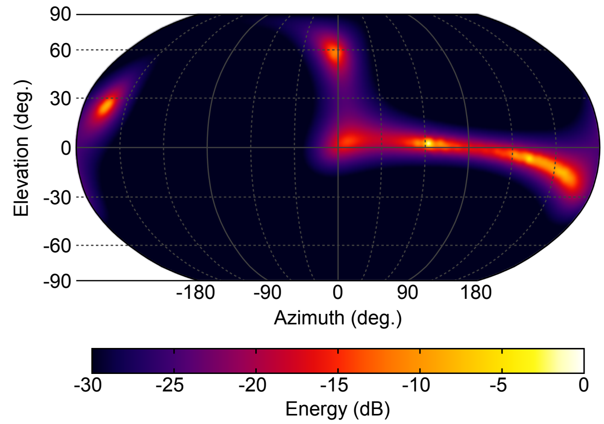

Fig. 2. This figure shows the MUSIC spatial spectrum calculated from the order-2 HOA signals.

by:

$$\mathbf{\Psi} = \mathrm{diag}\left(\left[\psi_1, \psi_2, \psi_3, ..., \psi_{(L+1)^2}\right]\right) \ ,$$
$$\psi_i = \begin{cases} 0 & \text{for } i \leq 6 \\ 1 & \text{otherwise} \end{cases} . \tag{19}$$

The MUSIC spatial spectrum is shown in Figure 2. The spectrum underlines a spatial support for the dominant sources. However, with the exceptions of sources 1 and 3, it is not very clear where the sources are. As well, the MUSIC spectrum does not provide quantitative information about the energies of the sources. Lastly, note that this spectrum was calculated assuming we knew 6 dominant sources were present, whereas no assumption on the number of sources was made to obtain the map presented in Figure 1(d). In summary the main difference between MUSIC and the proposed approach is that we apply SR techniques to obtain a super-resolution image.

6. CONCLUSIONS

In this paper we briefly reviewed sound field imaging using SMAs. Using traditional beamforming methods, sound field images obtained from SMAs have limited resolution. Higher resolution images can be obtained using SR methods, however such methods are likely to fail in the presence of noise or reverberation. In order to increase the robustness of SR-based sound field analysis, we proposed to separate the HOA signals into directional and diffuse components, using a simple subspace method. Our simulations show that this pre-processing has the potential to dramatically improve the accuracy of the SR sound field analysis in the presence of diffuse noise. As well, the pre-processing algorithm does not require the number of sources to be known.

7. REFERENCES

[1] B.N. Gover, J.G. Ryan, and M.R. Stinson, “Microphone array measurement system for analysis of directional and spatial variation of sound fields,” *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 1980–1991, 2002.

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

[3] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, “Spherical microphone array beamforming,” in *Speech Processing in Modern Communication: Challenges and Perspectives*, I. Cohen, J. Benesty, and S. Gannot, Eds. Springer, 2010.

[4] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, “Robust localization of multiple sources in reverberent environments using EB-ESPRIT with spherical microphone arrays,” in *Proceedings of the 2011 ICASSP*, Prague, Czech Republic, May 2011.

[5] R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276–280, 1986.

[6] R. Roy and T. Kailath, “ESPRIT - estimation of signal parameters via rotational invariance techniques,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[7] 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.

[8] A. Wabnitz, N. Epain, and C.T. Jin, “A frequency-domain algorithm to upscale ambisonic sound scenes,” in *Proceedings of the 2012 ICASSP*, Kyoto, Japan, March 2012.

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

[10] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, “Iteratively reweighted least squares minimization for sparse recovery,” *Comm. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[11] 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.