DIRECTION OF ARRIVAL ESTIMATION FOR SPHERICAL MICROPHONE ARRAYS BY COMBINATION OF INDEPENDENT COMPONENT ANALYSIS AND SPARSE RECOVERY

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

Spherical microphone arrays provide a powerful tool for examining source localization and direction of arrival (DOA) estimation in the spherical harmonic domain. In previous work, we have investigated applying instantaneous independent component analysis (ICA) or sparse recovery separately in the spherical harmonic domain for DOA estimation. These algorithms work reasonably well, but rely on different signal characteristics: namely statistical independence or the spatial distribution of sources. In this paper, we describe methods to combine the ICA and sparse recovery algorithms to improve DOA estimation. The simulation results indicate that combining ICA and sparse recovery leads to more robust DOA estimation.

Index Terms— Direction of-arrival estimation, Compressed sensing, Independent component analysis, Spherical microphone arrays

1. INTRODUCTION

Spherical microphone arrays (SMAs) enable array signal processing such as beamforming and DOA estimation to be performed in the spherical harmonic domain [1]. In the SMA signal processing framework, the microphone signals are first decomposed into spherical harmonics or phase modes. We refer to the corresponding signals as Higher-Order Ambisonic (HOA) signals, in keeping with the spatial sound field literature (see [1–4]). There are several advantages to using SMAs and HOA domain signal processing, including a panoramic spatial analysis and wideband beamforming properties [5]. Different source localization algorithms have recently been proposed for spherical microphone arrays [4, 6–11]. Most of these are steered beamforming-based and subspace localization-based techniques such as eigenbeam minimum variance distortionless response (EB-MVDR) [6, 7], eigenbeam multiple signal classification (EB-MUSIC) [8] and eigenbeam estimation of signal parameters via rotational invariance techniques (EB-ESPRIT) [9, 10].

In [4], we demonstrate that applying a linear ICA model in the HOA domain yields a mixing matrix that can be compared with a spherical harmonic matrix to provide DOA estimation that is more robust than eigenbeam MUSIC analysis. In [11], we describe a sparse recovery (SR) method for DOA estimation of the early echoes in a reverberant speech signal. We have generally found that DOA estimation using either the ICA or SR algorithm out performs the eigenbeam MUSIC algorithm for most sound conditions. Still, these algorithms can fail. The ICA-based algorithm fails to localize sources when they are not statistically independent and the SR-based algorithm fails to estimate source directions that are very close to each other. In this paper, we investigate the benefits of combining the ICA and SR approaches to provide a more robust source localization algorithm.

This paper is organized as follows. Section 2 describes the sound field decomposition into spherical harmonics using a SMA. Sections 3 and 4 detail two different methods to combine the ICA and SR approach to DOA estimation. The simulation results are described in Section 5. Conclusions are drawn in Section 6.

2. SOUND FIELD ANALYSES IN THE HOA DOMAIN

This section briefly describes microphone array signal processing in the HOA domain. In the frequency domain, any sound field consisting of incident sound waves can be modelled as a sum of L spherical harmonic modes, i.e., we have [4]:

$$p(r \leq \hat{r}, \theta, \phi) \approx \sum_{l=0}^{L} \sum_{m=-l}^{l} i^l j_l(kr) Y_l^m(\theta, \phi) b_{lm}(f)$$
$$\text{with} \quad \hat{r} = \frac{2L}{ke} , \tag{1}$$

where $p(r, \theta, \phi)$ is the acoustic pressure corresponding to the frequency f and at the point with spherical coordinates (r, θ, ϕ); k is the wave number given by $k = 2\pi f/c$ where c denotes the speed of sound; i is the imaginary unit; j_l is the spherical Bessel function of degree l; Y_l^m is the spherical harmonic function of order l and degree m; $b_{lm}(f)$ is the spherical harmonic expansion coefficient for order l and degree m; and e is the mathematical constant known as Euler's number. 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 HOA signals. In the case where the sound field consists of V plane waves, the resulting HOA signals are given by:

$$\mathbf{b}(t) = \mathbf{Y}\mathbf{s}(t) , \tag{2}$$

where

- $\mathbf{b}(t)$ is the vector of the order-L HOA signals:
$$\mathbf{b}(t) = [b_{00}(t), b_{1-1}(t), b_{10}(t), ..., b_{LL}(t)]^T ,$$
- $\mathbf{Y}$ is the mixing matrix, given by:
$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_V]^T ,$$
- $\mathbf{y}_v$ is the vector of the spherical harmonic function values :
$$\mathbf{y}_v = [Y_0^0(\theta_v, \phi_v), Y_1^{-1}(\theta_v, \phi_v), Y_1^0(\theta_v, \phi_v), Y_1^1(\theta_v, \phi_v), ..., Y_L^L(\theta_v, \phi_v)]^T ,$$

- $\mathbf{s}(t)$ is the vector of the plane-wave source signal, $s_v(t)$, incoming from the angular direction (θ_v, ϕ_v):

$$\mathbf{s}(t) = [s_1(t), s_2(t), ..., s_V(t)]^T \ .$$

The HOA signals form a linear, instantaneous mixture of the plane-wave signals. SMAs easily allow the determination of the HOA signals corresponding to a recorded sound field. There is a transformation between the microphone array signal domain and the HOA domain. Letting $\mathbf{d}(f)$ be the microphone signals and $\mathbf{b}(f)$ the HOA signals, we have:

$$\begin{aligned} \mathbf{d}(f) &= \mathbf{\Omega}(f)\, \mathbf{b}(f)\,, \\ \mathbf{b}(f) &= \text{pinv}\left(\mathbf{\Omega}(f)\right) \mathbf{d}(f)\,, \end{aligned} \tag{3}$$

where $\mathbf{\Omega}(f)$ is the transfer matrix between the HOA components of the sound field and the microphone signals at frequency f, and pinv$[\cdot]$ is the Moore-Penrose pseudo-inverse matrix operator. In practice, the transformation is implemented in the time domain using a set of filters, referred to as HOA encoding filters. For further details regarding the transformation between the microphone array signal domain and the HOA domain, please refer to [4].

3. ROBUST DOA ESTIMATION: ICA FOLLOWED BY SPARSE RECOVERY

The HOA domain signals form a linear and instantaneous mixture of the incoming plane-wave signals as shown in equation (2). We previously proposed to localize the source signals using a standard, linear ICA model [4]. A difficulty with this method is that ICA's underlying hypothesis that the incoming plane-wave signals are statistically independent may not be entirely valid, e.g., in a reverberant situation in which two echoes arrive from different directions simultaneously. In this section, we briefly review the ICA approach to DOA estimation and then propose a new method, which we refer to as the ICA-SR method, to improve the results using sparse recovery.

In [4], we apply the standard linear ICA model to the order-L HOA signals. ICA estimates a vector of separated signals, $\hat{\mathbf{s}}(t)$, and a mixing matrix, $\hat{\mathbf{H}}$, such that:

$$\mathbf{b}(t) = \hat{\mathbf{H}}\hat{\mathbf{s}}(t)\ . \tag{4}$$

Assuming that the ICA algorithm separated the source signals perfectly, the separated signals are proportional to the actual source signals. In other words, each column of $\hat{\mathbf{H}}$, $\hat{\mathbf{h}}_n$, is proportional to a plane-wave direction vector of spherical harmonics $\mathbf{y}_v$. We generally form a large matrix, $\mathbf{Y}_{\text{dict}}$, of plane-wave direction vectors. The v-th column of $\mathbf{Y}_{\text{dict}}$ provides the spherical harmonic expansion for a possible plane-wave source located in the direction (θ_v, ϕ_v). Thus, V is the number of entries in the dictionary of possible plane-wave source directions and is chosen much larger than possible number of sources. Putative source directions can be estimated as the columns of $\mathbf{Y}_{\text{dict}}$ that are maximally correlated with the columns of $\hat{\mathbf{H}}$. One approach for DOA estimation, which we refer to as the thresholding approach, is to apply a threshold (typically 0.95) to the correlation values between the columns of $\hat{\mathbf{H}}$ and the columns of $\mathbf{Y}_{\text{dict}}$. Assuming that the separated signals have a large correlation with the correct direction in space, this method works.

When the ICA method fails, however, intuitively each ICA output, s_n, is a combination of a few correlated sources and the corresponding column, $\hat{\mathbf{h}}_n$, of the mixing matrix is actually a mixture of more than one column of the matrix $\mathbf{Y}_{\text{dict}}$. Mathematically this can be expressed as:

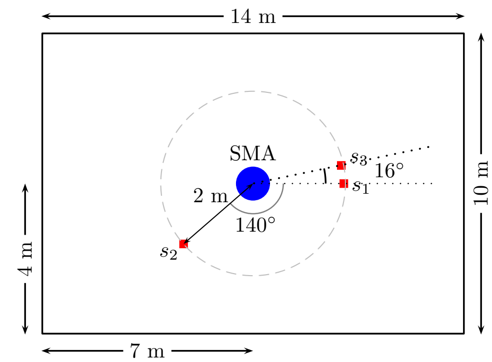

Fig. 1. The geometry of the SMA and sources for the acoustic simulation is shown.

$$\begin{aligned} \hat{\mathbf{h}}_n &= \sum_v a_{vn}\mathbf{y}_v \\ &= \mathbf{Y}_{\text{dict}}\mathbf{a}_n \qquad\qquad n = 1, .., (L+1)^2\,, \end{aligned} \tag{5}$$

where $\mathbf{a}_n$ is a $V \times 1$ vector. By solving equation (5) for $\mathbf{a}_n$, we can determine the source directions that comprise $\hat{\mathbf{h}}_n$. The difficulty is that equation (5) is an under-determined system. There are an infinite number of solutions and the inverse problem is ill-posed. Assuming that $\hat{\mathbf{h}}_n$ is a sum of only a few directions, the vector $\mathbf{a}_n$ will be sparse with a few non-zero elements (i.e., the a_{vn} is non-zero if there is a signal in the direction (θ_v, ϕ_v)). Thus in the new ICA-SR method, we propose to impose a sparsity constraint on the solution $\mathbf{a}_n$. Mathematically we formulate the sparse recovery problem as:

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{a}_n\|_1 \\ &\text{subject to} \quad \hat{\mathbf{h}}_n = \mathbf{Y}_{\text{dict}}\mathbf{a}_n\ , \end{aligned} \tag{6}$$

where $\|\mathbf{a}_n\|_1$ is the L_1-norm of $\mathbf{a}_n$. The iteratively re-weighted least square (IRLS) minimization [12] is used for solving equation (6). Each $\hat{\mathbf{h}}_n$ consists spatial information from some or all sources. We solve equation (6) for each column of the mixing matrix and then form the single vector $\mathbf{a} = \sum_n \mathbf{a}_n$. If the number of sources, P, is known, we select the directions that correspond to the P highest values in vector $\mathbf{a}$ as the putative source directions. If the number of sources is unknown, a threshold value (typically 0.85) is set such that the putative source directions are chosen as the directions corresponding to elements in the vector, $\mathbf{a}$, that are above the threshold.

3.1. Simulation

To evaluate the ICA-SR DOA estimation method described above, we consider two sources in a reverberant room and vary the cross-correlation between these two sources. By varying the cross-correlation, we modify the degree of statistical independence between the two sources. We expect that when the cross-correlation is high, that DOA estimation using ICA alone should fail, but that the ICA-SR method may still work. We compute the cross-correlation, ρ, of two sources as:

$$\rho = E\{s_1 s_2\},$$

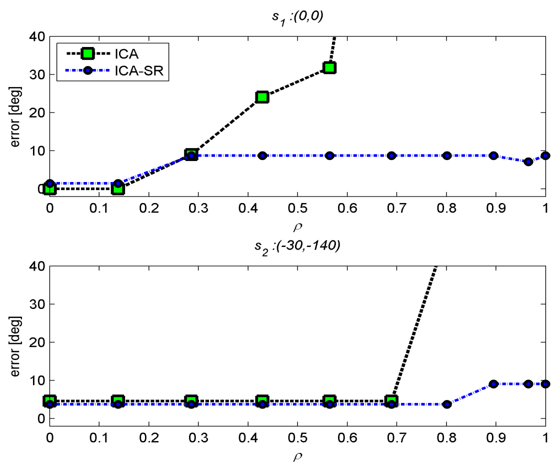

Fig. 2. The error in the estimated angular position as a function of the cross-correlation is shown for both the ICA-only and ICA-SR algorithms. The top and bottom figures corresponds to the source positions $(0, 0)$ and $(-30, -140)$, respectively.

where E denotes the statistical expectation operator. Two zero mean sources s_1 and s_2 are said to be uncorrelated, if their cross-correlation value is zero, $\rho = 0$. For the simulations, we vary ρ evenly in ten steps between 0 and 1.

Other details for the simulations are as follows. The SMA used consists of two concentric spherical arrays of 12 omnidirectional microphones. There are 12 microphones located on the surface of a rigid sphere with a radius of 3 cm; the other 12 microphones are located on the surface of an open sphere with a radius of 15 cm. The sound sources are place in a room with size 14 m$\times$10 m$\times$3 m. The SMA is located at (7 m, 4 m, 1.3 m) relative to the corner of the room. The two sources are located at a distance of two meters from the microphone array at the angular positions of s_1: $(0, 0)$ and s_2: $(-30, -140)$, respectively. The configuration of the sources relative to the SMA are shown in Fig. 1.

The reverberation time of the room, RT60, is approximately 450 ms and the signal-to-reverberation ratio (SRR) is about -3.8 dB. The reverberant impulse responses between the sources and the microphone array sensors have been calculated using MCROOMSIM, a multichannel room acoustics simulator [13]. The source signals are speech signals of approximately 4 s in duration with a sampling frequency of 16 kHz and the plane-wave dictionary comprises 2562 directions which were evenly distributed over the sphere, resulting in an angular resolution of approximately 4°. We band-pass filter the order-2 HOA signals prior to applying ICA and SR so that they contain only the frequencies where the HOA signals can be considered as instantaneous mixtures (for more details see [4]). The ICA algorithm is applied to signals using FastICA [14], an Independent Component Analysis package for the Matlab environment.

The DOA estimation results are shown in Fig. 2. The difference between the true and estimated angular position is plotted for both the ICA-only and ICA-SR methods as a function of the cross-correlation, ρ, between the two sources. As shown in Fig. 2, the DOA estimation for the ICA-SR algorithm is more robust than for the ICA-only algorithm as ρ increases. In particular, note that when $\rho > 0.5$, the ICA-only method fails to find two sources, whereas the ICA-SR method can still localize two sources when $\rho = 1.0$ – because it includes spatial information via the sparse recovery.

4. ROBUST DOA ESTIMATION: SPARSE RECOVERY FOLLOWED BY ICA

In this section, we describe a complementary approach to that presented in Section 3. Previously, we applied the ICA algorithm and followed it with sparse recovery. We now want to examine what happens when we flip the order and first apply sparse recovery followed by ICA. In [11], we presented a DOA estimation algorithm where the source signals are localized by applying an SR algorithm to solve equation (2). In that work, we first rewrite equation (2) to make the use of time windows (length N) explicit:

$$\mathbf{B} = \mathbf{Y}_{\text{dict}}\mathbf{S}\,, \tag{7}$$

where

$$\mathbf{S} = [\mathbf{s}(t), \mathbf{s}(t+1), \cdots, \mathbf{s}(t+N-1)]\ ,$$
$$\mathbf{B} = [\mathbf{b}(t), \mathbf{b}(t+1), \cdots, \mathbf{b}(t+N-1)]\ .$$

We then assume that the sound field can be explained by a minimum number of plane wave sources. In other words, we assume in equation (7) that most of the rows of $\mathbf{S}$ can be taken as zero because there are no signals in those directions. We then formulate the DOA estimation as the following SR problem:

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{S}\|_{12} \\ &\text{subject to} \quad \mathbf{B} = \mathbf{Y}_{\text{dict}}\mathbf{S}\,, \end{aligned} \tag{8}$$

where $\|\mathbf{S}\|_{12}$ is the l_{12} norm of $\mathbf{S}$ which is defined as:

$$\|\mathbf{S}\|_{12} = \sum_{v=1}^{V} \sqrt{\sum_{n=1}^{N} s_v(t+n-1)^2}\ .$$

We solve the SR problem using the IRLS algorithm and apply regularization to achieve better results. The source directions are estimated as those directions corresponding to the most energetic rows in $\mathbf{S}$.

Although the SR DOA method can achieve surprising super-resolution acoustic imaging, there are still limits in resolution that arise when the sources are near to each other. In this work, we examine whether we can improve results by applying ICA to the outputs of the sparse recovery approach. We refer to this approach as the SR-ICA method. To explain the SR-ICA method, we begin by supposing we have applied the SR DOA method and found a k-sparse solution $\mathbf{S}_{\text{sr}}$ to the SR problem (8). The number k is typically chosen greater than the number of sources. Within the SR framework, this means we solve for the matrix $\mathbf{W}$ such that: $\mathbf{WB} = \mathbf{S}_{\text{sr}}$, where $\mathbf{S}_{\text{sr}}$ has only k rows. The idea is to now apply ICA to the k signals in $\mathbf{S}_{\text{sr}}$ and see whether we can improve our solution. Applying ICA to $\mathbf{S}_{\text{sr}}$ gives:

$$\mathbf{WB} = \mathbf{S}_{\text{sr}} = \mathbf{GS}_{\text{ica}}, \tag{9}$$

where $\mathbf{G}$ is the ICA mixing matrix. Intuitively, we expect the mixing matrix $\mathbf{G}$ to be close to the identity matrix for most situations. However, when two statistically independent sources cannot be resolved by the SR method, ICA may still have the chance to separate the sources. To examine this issue, we can compare the acoustic image maps generated from the derived $\mathbf{Y}$-matrices:

$$\begin{aligned} \mathbf{Y}_{\text{sr}} &= \text{pinv}(\mathbf{W}) \\ \mathbf{Y}_{\text{ica}} &= \text{pinv}(\mathbf{W})\mathbf{G} \end{aligned} \tag{10}$$

The acoustic image maps are generated from the derived $\mathbf{Y}$-matrices by correlating them with the plane-wave dictionary $\mathbf{Y}_{\text{dict}}$ as described in the second paragraph of Section 3.

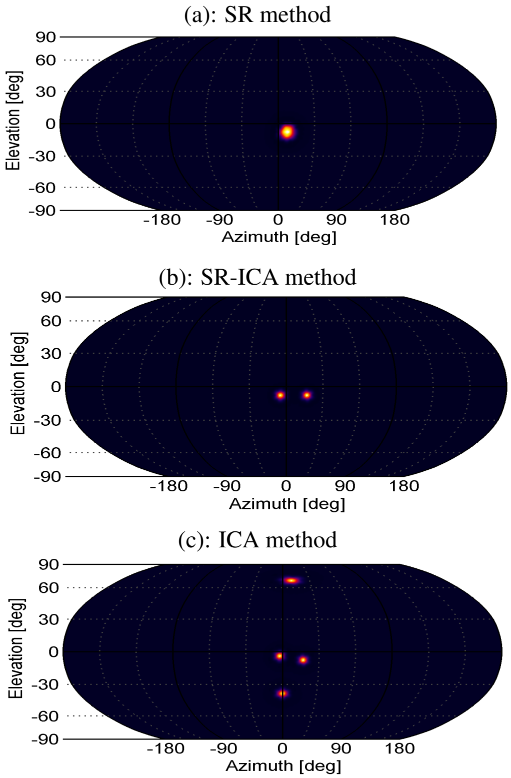

Fig. 3. An acoustic image map is shown for (a) the SR-only method, (b) the SR-ICA method, and (c) the ICA-only method.

Fig. 3 shows some simulation results comparing the SR method with the SR-ICA method for two sources s_1 and s_3 positioned close to each other at $(0, 0)$ and $(0, 16)$, respectively (see Fig. 1). We use a time window of length $N = 64000$ which corresponds to 4 s and we set $k = 6$. Other details for the simulations are as described in Section 3.1. Fig. 3 shows the acoustic image maps (i.e., the acoustic energy as a function of the direction in space) for the SR-only and SR-ICA algorithms. As shown in Fig. (3-a), there is only one peak – meaning the SR-based algorithm fails to estimate position of two sources. However, the SR-ICA method finds the two sources as shown in Fig. (3-b). It is interesting to consider the results for the ICA-only algorithm which are shown in Fig. (3-c). We see that the ICA-only method identifies too many source positions and finds four peaks. Thus, the SR-ICA method achieves a more robust DOA estimation.

5. CONCLUSION

In this paper, we describe methods for combining the ICA and sparse recovery algorithms to achieve more robust source localization using a spherical microphone array. The ICA-only and SR-only algorithms estimate the source positions based on statistical and spatial information, respectively. By combining these two algorithms, we are able to incorporate both statistical and spatial information for source localization. Results show that the ICA-SR and SR-ICA methods can outperform the ICA-only and SR-only algorithms and in some situations can even estimate the sources positions when the other algorithms fail completely. In future work, we will try to unify the two criteria of independence and sparsity in spatial domain to obtain a single algorithm for both source separation and localization.

References

[1] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[2] J. Meyer and G.W. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," *In Proceedings of the ICASSP*, pp. 1781–1784, 2002.

[3] A. Parthy, C. Jin, and A. van Schaik, "Optimisation of co-centred rigid and open spherical microphone arrays," *In Proceedings of the AES 120th Convention, Paris, France*, 2006.

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

[5] H. Teutsch and W. Kellermann, "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," *In Proceedings of Joint Workshop on Hands-Free Communication and Microphone Arrays*, p. 1314, 2005.

[6] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," *In Proceedings of the ICASSP*, pp. 113–116, 2011.

[7] S. Yan, H. Sun, U.P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 19, pp. 361–371, 2011.

[8] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical microphone array beamforming," *In Speech Processing in Modern Communication: Challenges and Perspectives, Chap. 11*, pp. 281–305, 2010.

[9] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," *In Proceedings of the ICASSP*, pp. 5276–5279, 2008.

[10] 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*, pp. 117–120, 2011.

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

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

[13] A.Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," *In Proceedings of the International Symposium on Room Acoustics*, 2010, `http://www.ee.usyd.edu.au/carlab/mcroomsim.htm`.

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