ROBUST SOURCE SEPARATION WITH DIFFERENTIAL MICROPHONE ARRAYS AND INDEPENDENT LOW-RANK MATRIX ANALYSIS

Dexin Li[1], *Gongping Huang*[2], *Yanqiang Lei*[3], *Jingdong Chen*[1], *and Jacob Benesty*[4]

[1]CIAIC, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
[2]Technion–Israel Institute of Technology, Technion City, Haifa 3200003, Israel
[3]CVTE Research, 6 Yunpu 4th Road, Huangpu District, Guangzhou 510000, China
[4]INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada

ABSTRACT

Acoustic source separation has been an active and important area of research in the field of acoustic signal processing. This paper deals with this problem using small and compact differential microphone arrays (DMAs) so that the resulting technology can be used in a broad range of small devices in voice communication and human-machine interfaces. A straightforward way to achieve source separation with DMAs is through differential beamforming. Although it has frequency-invariant beampatterns and high directivity in comparison with other existing beamforming methods with the same number of sensors, differential beamforming with small DMAs has limited spatial gain, generally leading to insufficient separation performance. To circumvent this limitation, we propose in this work a method to combine differential beamforming with an independent vector analysis (IVA) based algorithm. Specifically, differential beamformers are designed and applied to separate sound sources from different directions. Then, differential beamformers' outputs are used as inputs for the independent low-rank matrix analysis (ILRMA) algorithm, a widely used IVA method for blind source separation. The advantage of this proposed method consists of at least three aspects: 1) improving the source separation performance, 2) helping deal with the permutation problem, and 3) helping improve the convergence of ILRMA.

Index Terms— Differential microphone arrays, beamforming, source separation, independent low-rank matrix analysis.

1. INTRODUCTION

In many speech communication and human-machine interaction systems, it is important to separate different sound sources. One way to achieve separation is through microphone array beamforming, which extracts the signal of interest with a spatial filter [1–3]. Since most smart devices can only use small-aperture microphone arrays, it is nature to adopt DMAs and differential beamforming [4–7], which can achieve higher directivity as compared to other beamforming methods with the same number of microphones. However, even with differential beamforming, the separation performance with small-aperture microphone arrays is insufficient for most applications.

Another way to achieve separation is through the so-called blind source separation (BSS) technique [8–12]. Most widely used algorithms for BSS are based on independent component analysis (ICA) [10, 13–15], which assumes that the source signals are statistically independent. The core issue of ICA based source separation is to estimate a demixing system [10, 13], which is very challenging in reverberant environments due to the inherent problems of permutation [16–22] and scale ambiguity with ICA. The independent vector analysis (IVA) was subsequently proposed to mitigate the permutation problem [20–22]. One of the widely used IVA algorithms is the independent low-rank matrix analysis (ILRMA) [23–25], which was shown better than the ICA based methods in dealing with permutation. However, permutation still exists with ILRMA. Another prominent issue with ILRMA is that it suffers from slow convergence. Due to the existence of permutation and scale ambiguity, source separation performance with ILRMA in reverberant environments is also limited.

To improve the source separation performance, efforts were made in the literature to combine ICA and adaptive beamforming [26, 27]. But the existing work mainly focused on using geometric and direction-of-arrival information [26–29]. In this paper, we propose a method that combines differential beamforming and the ILRMA method, which works with small-aperture microphone arrays. The differential beamformer is designed with the Jacobi-Anger expansion method and is applied to the microphone observation signals. The beamformers' outputs are used as inputs to ILRMA. Simulations demonstrated the good properties and superior performance of this proposed method.

2. SIGNAL MODEL AND PROBLEM FORMULATION

We consider the signal model in room acoustic environments in which a microphone array with M sensors captures signals from N sources in some noise field. The observation signals are expressed as

$$y_m(t) = \sum_{n=1}^{N} a_{mn}(t) * s_n(t) + v_m(t), \tag{1}$$

for $m = 1, 2, \ldots, M$, where $a_{mn}(t)$ denotes the acoustic impulse response from the unknown speech source, $s_n(t)$, to the mth microphone, $*$ stands for linear convolution, and $v_m(t)$ is the additive noise at microphone m. We assume that all the source signals and noise are independent, zero mean, real, and reasonably broadband.

Transforming all the signals into the short-time Fourier transform (STFT) domain, we can rewrite the signal model in (1) as [5]

$$Y_m(\omega, l) = \sum_{n=1}^{N} A_{mn}(\omega) S_n(\omega, l) + V_m(\omega, l), \tag{2}$$

where ω is the angular frequency, l is the time-frame index, and $Y_m(\omega, l)$, $A_{mn}(\omega)$, $S_n(\omega, l)$, and $V_m(\omega, l)$ are the STFTs of $y_m(t)$, $a_{mn}(t)$, $s_n(t)$, and $v_m(t)$, respectively. In a vector form, (2) can be arranged as

$$\begin{aligned}\mathbf{y}(\omega, l) &= \left[Y_1(\omega, l) \quad Y_2(\omega, l) \quad \cdots \quad Y_M(\omega, l)\right]^T \\ &= \mathbf{A}(\omega)\,\mathbf{s}(\omega, l) + \mathbf{v}(\omega, l),\end{aligned} \tag{3}$$

where the superscript ^{T} denotes the transpose of a vector or a matrix, $\mathbf{A}(\omega)$ is the $M \times N$ mixing matrix whose (m, n)th element is

This work was supported in part by the Key Program of National Science Foundation of China (NSFC) under Grant No. 61831019, the NSFC and Israel Science Foundation (ISF) joint research program under Grant No. 61761146001 and 2514/17, and the NSFC "Distinguished Young Scientists Fund" under Grant No. 61425005.

$A_{mn}(\omega)$,

$$\mathbf{s}(\omega,l) = [S_1(\omega,l) \quad S_2(\omega,l) \quad \cdots \quad S_N(\omega,l)]^T \tag{4}$$

is the source vector of length N, and $\mathbf{v}(\omega,l)$ is the noise vector of length M defined in a similar way to $\mathbf{y}(\omega,l)$.

The objective of source separation is to recover all the source signal components in $\mathbf{s}(\omega,l)$ from the mixture observation vector $\mathbf{y}(\omega,l)$.

3. ILRMA ALGORITHM

The ILRMA algorithm aims at estimating a demixing matrix [24]:

$$\mathbf{W}(\omega) = [\mathbf{w}_1(\omega) \quad \mathbf{w}_2(\omega) \quad \cdots \quad \mathbf{w}_N(\omega)]^H, \tag{5}$$

so that an estimate of the source vector is obtained as

$$\hat{\mathbf{s}}(\omega,l) = \mathbf{W}(\omega)\,\mathbf{y}(\omega,l), \tag{6}$$

where the superscript ^{H} is the conjugate-transpose operator. The paramount issue in ILRMA is how to find the demixing matrix $\mathbf{W}(\omega)$, which can be achieved through non-negative matrix factorization (NMF) [22]. Let us assume an isotropic complex Gaussian distribution of the nth separated source signal, i.e.,

$$p\left(\hat{\mathbf{S}}_n\right) = \prod_{\omega,l} \frac{1}{\pi\lambda_n(\omega,l)} \exp\left(-\frac{|\hat{S}_n(\omega,l)|^2}{\lambda_n(\omega,l)}\right), \tag{7}$$

where $\hat{\mathbf{S}}_n$ is a matrix consisting of $\hat{S}_n(\omega,l)$ from all the frequencies and time frames, i.e., $\left\{\hat{\mathbf{S}}_n\right\}_{\omega,l} = \hat{S}_n(\omega,l)$, and $\lambda_n(\omega,l)$ is the time-frequency-varying source variance. Then, the negative log-likelihood can be defined and used as the cost function to estimate the demixing matrix, which is written as [24]

$$\begin{aligned}\mathcal{J}\left(\{\mathbf{W}(\omega)\}_1^{\Omega}, \boldsymbol{\Lambda}\right) = &-\sum_n \log p\left(\hat{\mathbf{S}}_n\right) \\ &- 2L\sum_{\omega} \log|\det \mathbf{W}(\omega)|,\end{aligned} \tag{8}$$

where $\boldsymbol{\Lambda}$ is a parameter set consisting of $\lambda_n(\omega,l)$ for all the n's, ω's, and l's.

The demixing matrix $\mathbf{W}(\omega)$ is then estimated with an iterative projection method as follows [22, 24]:

$$\mathbf{U}_n(\omega) = \frac{1}{L}\sum_l \frac{1}{\lambda_n(\omega,l)}\mathbf{y}(\omega,l)\,\mathbf{y}^H(\omega,l), \tag{9}$$

$$\mathbf{w}_n(\omega) \leftarrow [\mathbf{W}(\omega)\,\mathbf{U}_n(\omega,l)]^{-1}\mathbf{e}_n, \tag{10}$$

$$\mathbf{w}_n(\omega) \leftarrow \mathbf{w}_n(\omega)\left[\mathbf{w}_n^H(\omega)\,\mathbf{U}_n(\omega)\,\mathbf{w}_n(\omega)\right]^{-\frac{1}{2}}, \tag{11}$$

where $\mathbf{e}_n$ denotes a vector with its nth element being 1 and all the others being 0. The separated source signals are then obtained as

$$\hat{S}_n(\omega,l) = \mathbf{w}_n^H(\omega)\,\mathbf{y}(\omega,l). \tag{12}$$

In ILRMA, the time-frequency-varying variance $\lambda_n(\omega,l)$ is modeled by the NMF as $\lambda_n(\omega,l) = \sum_k t_n(\omega,k)\,v_n(k,l)$, with $t_n(\omega,k) \geq 0$ and $v_n(k,l) \geq 0$ being the NMF basis and activation variables, respectively, and $k = 1, 2, \ldots, K$ denotes the basis index. The NMF variables $t_n(\omega,k)$ and $v_n(k,l)$ are updated based on a convergence-guaranteed majorization-minimization (MM) algorithm [24, 30]:

$$t_n(\omega,k) \leftarrow t_n(\omega,k)\left[\frac{\sum_l \frac{|\hat{S}_n(\omega,l)|^2}{\lambda_n^2(\omega,l)} v_n(k,l)}{\sum_l \frac{1}{\lambda_n(\omega,l)} v_n(k,l)}\right]^{\frac{1}{2}}, \tag{13}$$

$$v_n(k,l) \leftarrow v_n(k,l)\left[\frac{\sum_\omega \frac{|\hat{S}_n(\omega,l)|^2}{\lambda_n^2(\omega,l)} t_n(\omega,k)}{\sum_\omega \frac{1}{\lambda_n(\omega,l)} t_n(\omega,k)}\right]^{\frac{1}{2}}, \tag{14}$$

$$\lambda_n(\omega,l) \leftarrow \sum_k t_n(\omega,k)\,v_n(k,l). \tag{15}$$

To avoid divergence during the optimization, some normalization should be applied at each iteration [24]. The scalar of the separated sources can be determined based on the power of the observation signals with the help of the back-projection (BP) method [17].

As most ICA based source separation methods, ILRMA suffers from the permutation and convergence problems. Moreover, as will be shown in Section 6, the separation performance of ILRMA with small size microphone arrays is limited and is insufficient for most applications. It is therefore indispensable to study how to further improve the source separation performance.

4. SOURCE SEPARATION VIA DMAS

Assume that the incidence angle of the nth source signal to the array is parameterized by the azimuthal angle $\theta_{\mathrm{s},n}$, the objective of beamforming is to recover it with a spatial filter $\mathbf{h}_n(\omega)$ of length M [5]:

$$Z_n(\omega,l) = \mathbf{h}_n^H(\omega)\,\mathbf{y}(\omega,l). \tag{16}$$

Then, the beampattern is

$$\mathcal{B}[\mathbf{h}_n(\omega), \theta] = \mathbf{h}_n^H(\omega)\,\mathbf{d}(\omega,\theta), \tag{17}$$

where $\mathbf{d}(\omega,\theta)$ is the steering vector of length M. In our study, we consider a microphone array of generic planar geometry. The steering vector can be written as [31]

$$\mathbf{d}(\omega,\theta) = \left[e^{\jmath\frac{\omega r_1}{c}\cos(\theta-\psi_1)} \quad \cdots \quad e^{\jmath\frac{\omega r_M}{c}\cos(\theta-\psi_M)}\right]^T, \tag{18}$$

where $\jmath$ is the imaginary unit with $\jmath^2 = -1$, r_m $(m = 1, 2, \ldots, M)$ is the distance from the mth microphone to the origin point of the coordinate system, c is the speed of sound in the air, and ψ_m is the angular position of the mth array element. Generally, it is required that the beampattern has a gain of 1 in the look direction and of less than 1 in other directions so that the signal of interest from the look direction passes through the beamformer without distortion while signals from other directions are suppressed.

It is well known that the differential beamformer's directivity pattern can be written as a frequency-independent cosine function of the direction θ [4]. Specifically, the frequency-independent directivity pattern of an Qth-order differential beamformer with its main beam pointing in the direction $\theta_{\mathrm{s},n}$ can be written as [32]

$$\mathcal{B}_Q(\theta - \theta_{\mathrm{s},n}) = \sum_{q=-Q}^{Q} b_{Q,q}e^{\jmath q\left(\theta-\theta_{\mathrm{s},n}\right)}, \tag{19}$$

where $b_{Q,q}$, $q = 0, \pm1, \ldots, \pm Q$ are real coefficients determining the shape of the directivity pattern [33]. The form of the target directivity pattern in (19) consists of the steering information; therefore, we can develop steerable differential beamformers with microphone arrays of planar geometry [34].

Then, the problem of differential beamforming becomes one of designing an optimal beamforming filter, $\mathbf{h}_n(\omega)$, such that the resulting beampattern is equal or is as close as possible to the desired frequency-independent directivity pattern [35], i.e.,

$$\mathcal{B}\left[\mathbf{h}_n(\omega), \theta\right] \rightarrow \mathcal{B}_Q\left(\theta - \theta_{\mathrm{s},n}\right), \ \forall\omega. \tag{20}$$

This can be achieved by the so-called Jacobi-Anger expansion of the exponential function that appears in the beamformer's beampattern. Then, the differential beamforming filter is solved by maximizing the WNG [32], whose solution is

$$\mathbf{h}_n(\omega) = \mathbf{\Upsilon}^*\left(\theta_{\mathrm{s},n}\right)\mathbf{\Psi}^H(\omega)\left[\mathbf{\Psi}(\omega)\mathbf{\Psi}^H(\omega)\right]^{-1}\mathbf{b}_Q, \tag{21}$$

where the superscript * is the complex-conjugate operator,

$$\mathbf{\Psi}(\omega) = \begin{bmatrix} (-\jmath)^{-Q}\boldsymbol{\psi}_{-Q}^H(\omega) \\ \vdots \\ \boldsymbol{\psi}_0^H(\omega) \\ \vdots \\ (-\jmath)^{Q}\boldsymbol{\psi}_{Q}^H(\omega) \end{bmatrix} \tag{22}$$

is a $(2Q+1) \times M$ matrix, with

$$\boldsymbol{\psi}_q(\omega) = \left[J_q\left(\frac{\omega r_1}{c}\right)e^{-\jmath q\psi_1} \ \cdots \ J_q\left(\frac{\omega r_M}{c}\right)e^{-\jmath q\psi_M}\right]^T \tag{23}$$

being a vector of length M, $J_q(\cdot)$ being the qth-order Bessel function of the first kind,

$$\mathbf{\Upsilon}\left(\theta_{\mathrm{s},n}\right) = \mathrm{diag}\left[e^{\jmath Q\theta_{\mathrm{s},n}}, \ \ldots, \ 1, \ \ldots, \ e^{-\jmath Q\theta_{\mathrm{s},n}}\right]$$

is a diagonal matrix, and $\mathbf{b}_Q = \left[b_{Q,-Q} \ \cdots \ b_{Q,0} \ \cdots \ b_{Q,Q}\right]^T$ is a vector consisting of the $2Q+1$ real coefficient determining the target directivity pattern [32].

A differential beamformer can achieve source separation through spatial filtering. But as ILRMA, the separation performance is generally not sufficient, particularly in reverberant environments.

5. SOURCE SEPARATION VIA DMAS/ILRMA

In this section, we investigate the potential of combining differential beamforming and ILRMA to improve the source separation performance. Briefly, we propose to use the differential beamformers' outputs as the inputs of ILRMA. Then, the NMF variables $t_n(\omega, k)$ and $v_n(k, l)$ in (13) and (14) are updated as follows:

$$t_n(\omega, k) \leftarrow t_n(\omega, k)\left[\frac{\sum_l \frac{|Z_n(\omega,l)|^2}{\lambda_n^2(\omega,l)} v_n(k,l)}{\sum_l \frac{1}{\lambda_n(\omega,l)} v_n(k,l)}\right]^{\frac{1}{2}}, \tag{24}$$

$$v_n(k, l) \leftarrow v_n(k, l)\left[\frac{\sum_\omega \frac{|Z_n(\omega,l)|^2}{\lambda_n^2(\omega,l)} t_n(\omega,k)}{\sum_\omega \frac{1}{\lambda_n(\omega,l)} t_n(\omega,k)}\right]^{\frac{1}{2}}, \tag{25}$$

$$\lambda_n(\omega, l) = \sum_k t_n(\omega, k)\, v_n(k, l). \tag{26}$$

All the other steps of the algorithms are the same as the conventional ILRMA. This combination provides the following benefits. First, differential beamforming provides a robust and more accurate update of the NMF variables $t_n(\omega, k)$ and $v_n(k, l)$, which helps improve both the source separation performance as well as the convergence of ILRMA. Second, since the differential beamformers' outputs contain the *a priori* information about the source signals, the permutation problem is naturally solved.

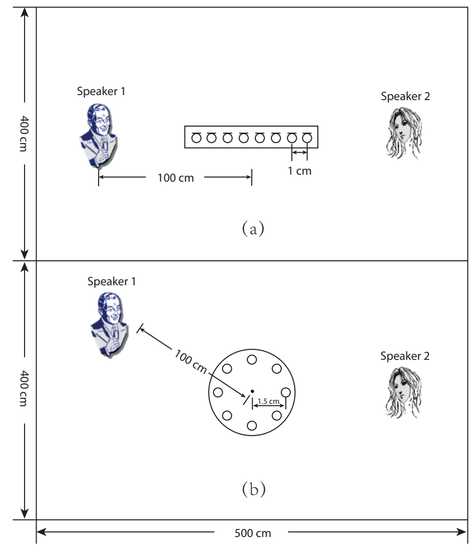

Fig. 1. Layout of the simulation setup: (a) with a ULA and (b) with a UCA.

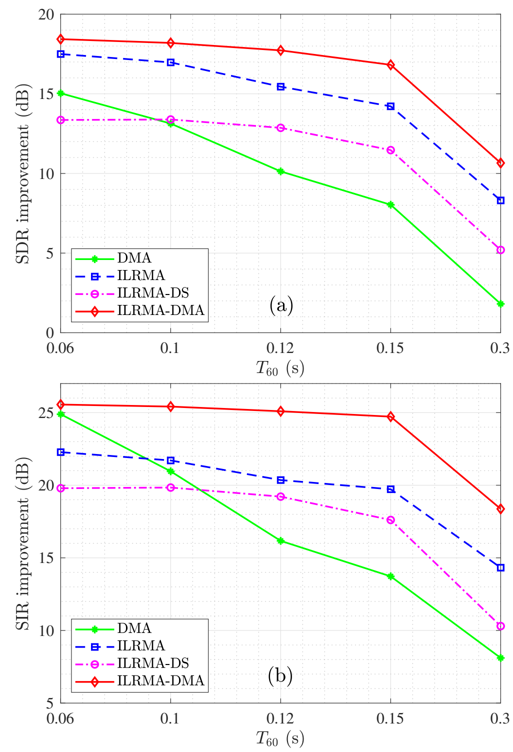

Fig. 2. SDR and SIR improvements for ILRMA, DMA, ILRMA-DMA, and ILRMA-DS with the ULA in different reverberation conditions.The two sources are located at 0° and 180°, respectively.

6. EXPERIMENTS

6.1. Experimental Setup

In this section, we study the performance of the proposed source separation method (called ILRMA-DMA for simplicity) through simulations. We consider two different microphone arrays: the first one is a uniform linear array (ULA) consisting of eight omnidirectional microphones with an interelement spacing of 1.0 cm, and the second one is a uniform circular array (UCA) consisting of eight omnidirectional microphones with a radius of 1.5 cm. The layout of the simulation

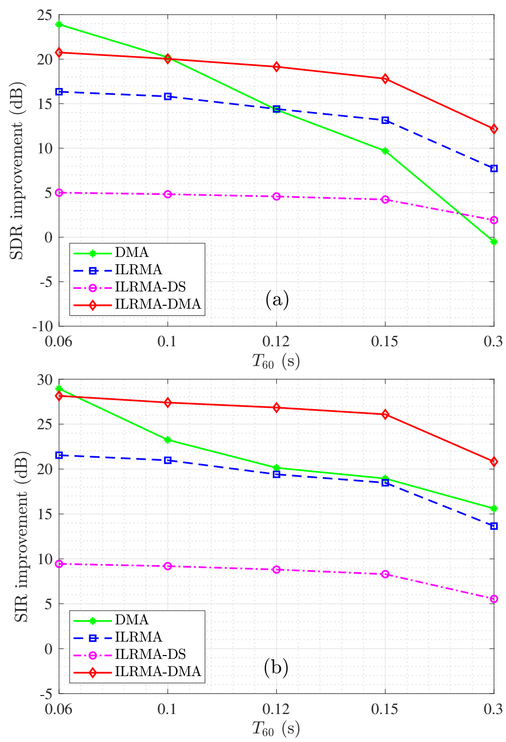

Fig. 3. SDR and SIR improvements for ILRMA, DMA, ILRMA-DMA, and ILRMA-DS, with the UCA in different reverberation conditions.The two sources are located 0° and 90°, respectively.

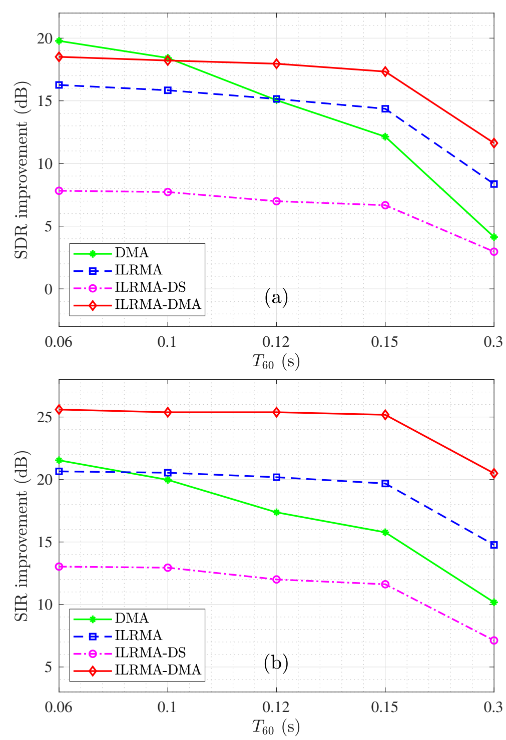

Fig. 4. SDR and SIR improvements for ILRMA, DMA, ILRMA-DMA, and ILRMA-DS with the UCA in different reverberation conditions.The two sources are located at 0° and 135°, respectively.

setup is illustrated in Fig. 1. The size of the room is $5\ \mathrm{m} \times 4\ \mathrm{m} \times 3\ \mathrm{m}$. The microphone array is placed in the room so that its center is at $[2.5, 2, 1.5]$ m. Two loudspeakers are placed 1 meter away from the center of the microphone array, which play back some pre-recorded speech signals to simulate two sound sources. For ULA, the two loudspeakers are placed at, respectively, 0° and 180°. In the UCA situation, we consider two cases: 1) the two loudspeakers are placed at, respectively, 0° and 90°; and 2) the two loudspeakers are placed at, respectively, 0° and 135°. The speech signals played back from the two loudspeakers are taken from the SiSEC2011 database [36]. To make simulations repeatable, the acoustic channel impulse responses from the loudspeakers to the eight microphones are generated with the image model method [37]. Then, the microphone signals are generated by convolving the source signal with the corresponding impulse responses. In our simulations, the reflection coefficients of the six walls are assumed to be the same. They are set to $0.1, 0.3, 0.5, 0.6, 0.8$ and the corresponding reverberation times, T_{60}, are approximately 0.06 s, 0.1 s, 0.12 s, 0.15 s, 0.3 s, respectively.

Differential beamformers may suffer from white noise amplification. To avoid this, we used the minimum-norm method to design the 1st-order cardioid, 2nd- and 3rd-order supercardioid for frequency bands below 500 Hz, $500 - 2000$ Hz, and $2000 - 8000$ Hz, respectively, when ULA is used, and for frequency bands below 2500 Hz, $2500 - 4000$ Hz and $4000 - 8000$ Hz, respectively, when UCA is used. The source separation performance are evaluated with the source-to-distortion ratio (SDR) and source-to-interference ratio (SIR). The SDR and SIR are global performance measures, detailed definition of which can be found in [38]. We implemented the performance measures using the *BSS_EVAL* MATLAB toolbox [39].

6.2. Results and Analysis

We first study the case with a ULA. For each reverberation condition, the performance measures are averaged over 4 groups arbitrarily chosen from the 10 trials in the database. For the purpose of comparison, the performance of the conventional ILRMA, combination of ILRMA and the delay-and-sum (DS) beamformer, and DMA is also presented. Note that, in all the methods, the principal component analysis is exploited to reduce mixtures' dimensionality [24]. The SDR and SIR improvement are plotted in Fig. 2. While almost all the studied methods improved both the SDR and SIR, the proposed ILRMA-DMA yielded best performance. The improvement in SDR and SIR reduces as the reverberation time increases. This is understandable since as the reverberation time becomes larger, it is more difficult to separate different sound signals. Nevertheless, as seen, the proposed ILRMA-DMA yielded significant SDR and SIR improvement in all the studied cases.

For UCA, we first study the case with one source propagating from 0° while the other propagating from 90°; the results are plotted in Fig. 3. We then study the case with one source propagating from 0° while the other from 135°; the results are plotted in Fig. 4. In both cases, the advantage of the ILRMA-DMA over the other studied methods is clearly seen. Note also from our simulations that the ILRMA-DMA has better convergence where the ILRMA-DMA converges within 50 iterations while the ILRMA needs more than 80 iterations to converge.

7. CONCLUSIONS

This paper dealt with the problem of blind acoustic source separation in reverberant environments with small size microphone arrays. We proposed a separation method that combines differential beamforming with the ILRMA algorithm, in which the differential beamformers' outputs are used as the inputs of the ILRMA algorithm. Simulations demonstrated that the proposed ILRMA-DMA method yielded much better performance than either ILRMA or DMA. Moveover, ILRMA-DMA converges faster than the ILRMA algorithm. An additional advantage of the proposed method is that the permutation problem is naturally solved as the differential beamformers' outputs contain the *a priori* direction information about the source signals.

8. REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[2] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Singapore: Wiley-IEEE Press., 2018.

[3] O. Schwartz, S. Gannot, and E. A. Habets, "Multispeaker LCMV beamformer and postfilter for source separation and noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 940–951, 2017.

[4] G. W. Elko, "Differential microphone arrays," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pp. 11–65, Springer, 2004.

[5] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2012.

[6] J. Chen, J. Benesty, and C. Pan, "On the design and implementation of linear differential microphone arrays," *J. Acoust. Soc. Am.*, vol. 136, pp. 3097–3113, Dec. 2014.

[7] G. Huang, J. Benesty, I. Cohen, and J. Chen, "A simple theory and new method of differential beamforming with uniform linear microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, 2020.

[8] C. Jutten and J. Herault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.

[9] S. Haykin: Ed., *Unsupervised Adaptive Filtering (Volume I: Blind Source Separation)*. John Wiley & Sons, The United States of America, 2000.

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, The United States of America, 2001.

[11] S. Makino, T.-W. Lee, and H. Sawada, *Blind Speech Separation*. Springer, The Netherlands, 2007.

[12] S. Makino, *Audio Source Separation*. Springer International, 2018.

[13] P. Comon, "Independent component analysis, a new concept?," *IEEE Trans. Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[14] B. Anthony and S. Terrence, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[15] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, pp. 1483–1492, 1997.

[16] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. IEEE ICASSP*, vol. 5, pp. 3140–3143, 2000.

[17] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, 2001.

[18] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 12, no. 5, pp. 530–538, 2004.

[19] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of binwise separated signals for permutation alignment in frequency-domain bss," *Proc. IEEE Int. Sym. Circuits and Systems*, pp. 3247–3250, 2007.

[20] A. Hiroe, "Solution of permutation problem in frequency domain ica using multivarite probability density functions," *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation*, pp. 601–608, 2006.

[21] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: an extension of ica to multivariate components," *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation*, pp. 165–172, 2006.

[22] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. IEEE WASPAA*, pp. 189–192, 2011.

[23] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Efficient multichannel nonegative matrix factorization exploiting rank-1 spatial model," *Proc. IEEE ICASSP*, pp. 276–280, 2015.

[24] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.

[25] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo, "Generalized independent lowrank matrix analysis using heavy-tailed distributions for blind source separation," *EURASIP J. Adva. Signal Process.*, vol. 2018, no. 28, pp. 1–25, 2018.

[26] L. C. Parra and C. V. Alvino, "Geometric source separation: merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech, Audio Process.*, vol. 10, no. 6, pp. 352–562, 2002.

[27] G. Reuven, S. Gannot, and I. Cohen, "Dual-source transfer-function generalized sidelobe canceller," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 711–727, 2008.

[28] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation algorithms for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," *Proc. IEEE IWAENC*, pp. 275–278, 2003.

[29] S. M. Golan, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signal," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, 2009.

[30] C. Févotte, N. Bertin, and J. L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence. with application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, 2009.

[31] R. A. Monzingo and T. W. Miller, *Introduction to Adaptive Arrays*. SciTech Publishing, Inc, Raleigh, NC, 2004.

[32] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1140–1153, 2017.

[33] G. W. Elko and J. Meyer, "Microphone arrays," in *Springer Handbook of Speech Processing* (J. Benesty, M. M. Sondhi, and Y. Huang, eds.), ch. 48, pp. 1021–1041, Berlin, Germany: Springer-Verlag, 2008.

[34] G. Huang, J. Chen, and J. Benesty, "Insights into frequency-invariant beamforming with concentric circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 12, pp. 2305–2318, 2018.

[35] G. Huang, J. Chen, and J. Benesty, "On the design of differential beamformers with arbitrary planar microphone array," *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. EL66–EL70, 2018.

[36] S. Araki, "The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation," *Proc. Latent Variable Anal. Signal Separation*, pp. 414–422, 2012.

[37] J. Allen and D. Berkley, "Image method for efficiently simulating smallroom acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[38] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

[39] C. Févotte, R. Gribonval, and E. Vincent, "BSS_EVAL toolbox user guide - Revision 2.0," *Tech. Rep. 1706, IRISA*, 2005. Available: http://www.irisa.fr/metiss/bss_eval/