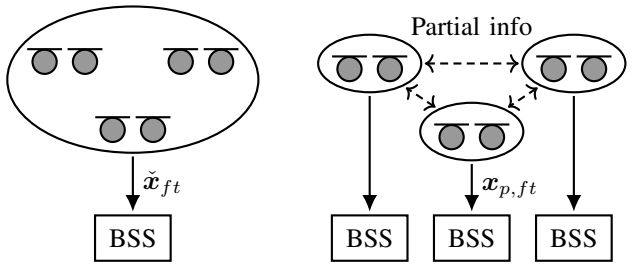
Auxiliary-Function-Based Decentralized Independent Vector Analysis for Distributed Microphone Arrays

Kouei Yamaoka*, Katsuhiro Morita*, Norihiro Takamune*, and Hiroshi Saruwatari*
* The University of Tokyo, Japan
E-mail: kouei_yamaoka@ipc.i.u-tokyo.ac.jp

***Abstract*—This paper proposes a novel blind source separation (BSS) method for distributed microphone arrays, namely, auxiliary-function-based decentralized independent vector analysis (IVA). Previously, a decentralized IVA algorithm based on the natural gradient method was proposed for image signals. Unlike conventional BSS approaches, which aggregate all observations at a central processing unit, the decentralized IVA utilizes signals only within small groups of nearby microphones, referred to as subarrays, and shares not the signals themselves, but their power information among subarrays. In this paper, we first formulate the optimization problem of decentralized IVA for audio signals captured by distributed microphone arrays. We then derive a decentralized IVA algorithm as a modified version of auxiliary-function-based IVA to improve the convergence speed and stability of the conventional decentralized IVA. Experimental results demonstrate that the proposed algorithm converges faster and more stably than the natural gradient-based approach and achieves better performance than centralized IVA.**



(a) Centralized BSS (b) Decentralized BSS

Fig. 1: Centralized vs. decentralized BSS. Centralized BSS aggregates all observations at a central processor, whereas decentralized BSS allows each subarray to perform BSS locally, sharing only limited information such as signal power, thereby reducing communication load and preserving privacy.

I. Introduction

Blind source separation (BSS) [1]–[3] is a technique for estimating individual source signals from observed mixtures without prior information about the mixing system, such as microphone positions. It plays a crucial role in the preprocessing stage of audio applications, including automatic speech recognition and hearing aids. Independent vector analysis (IVA) [4]–[6] and independent low-rank matrix analysis (ILRMA) [7] are widely used techniques for BSS. The optimization algorithms used in IVA have evolved from gradient-based methods [4], [5] to those based on the auxiliary function method, which offer faster and more stable convergence [6].

Recently, the framework of distributed microphone arrays or wireless acoustic sensor networks has gained significant attention, where multiple microphones or microphone arrays are spatially distributed over a wide area [8]–[10]. This framework enables BSS in large-scale environments that cannot be covered by a single conventional microphone array. In such a setup, small recording devices, such as smartphones and voice recorders, can function as individual microphone arrays, offering high convenience and scalability.

There are two main approaches to performing BSS on distributed microphone arrays. The first is to collect the signals observed by each microphone array at a central processor and treat them as if they were recorded by a single large array [11], as shown in Fig. 1(a). This approach is hereafter referred to as centralized BSS. It is a simple and effective method, as conventional techniques developed for single microphone arrays can be directly applied. Moreover, by utilizing all observed signals, centralized BSS fully leverages the benefits of distributed microphone arrays; namely, the large number of microphones provides rich spatial information that enhances the performance of array signal processing. However, since conventional array signal processing methods require strict time synchronization, the problem known as sampling rate offset (SRO) must be addressed in advance [12]–[14]. Additionally, when using a large number of microphones, the computational complexity of BSS algorithms becomes significant; e.g., standard IVA requires $\mathcal{O}(FM^3)$ computations, where F and M denote the number of frequency bins and microphones, respectively [15]. Finally, there is concern that when individuals use their personal devices as components of a distributed microphone array, aggregating information may lead to privacy issues, such as voiceprint leakage.

The second approach is to use only a limited subset of the available microphones. A naïve solution is to perform BSS individually within each microphone array using only local observations (hereinafter referred to as local BSS). While this approach completely avoids the challenges of centralized BSS, it also loses the benefits of spatially distributed sound information. Another option is to select an optimal subset of microphones [16]–[18]; however, this also involves discarding information captured by the remaining arrays.

In this paper, we propose a new decentralized IVA algorithm as an intermediate between centralized and local IVA, as shown in Fig. 1(b). Specifically, we first apply decentralized IVA, originally proposed for image signals [19], to acoustic

979-8-3315-7206-8/25/$31.00

(speech) signals. Second, since the original method is based on optimization using the natural gradient, we develop a faster and more stable algorithm by adopting the auxiliary function method [6]. Experimental results show that the proposed method converges faster and more stably than the conventional decentralized IVA. Moreover, the proposed decentralized algorithm outperforms centralized IVA in separation performance.

II. Conventional BSS on Distributed Microphone Arrays

A. Formulation

In this paper, we consider the problem of BSS using multiple microphone arrays, each referred to as a subarray. Let N, M_p and P denote the number of sources, microphones in the pth subarray, and subarrays, respectively, where $p \in \{1 \ \cdots \ P\}$ is the subarray index. The source and observed signals at the pth subarray in the short-time Fourier transform (STFT) domain are modeled as

$$\boldsymbol{s}_{ft} = [s_{ft1} \ \cdots \ s_{ftN}]^{\mathsf{T}} \in \mathbb{C}^N, \tag{1}$$

$$\boldsymbol{x}_{p,ft} = \left[x_{p,ft1} \ \cdots \ x_{p,ftM_p}\right]^{\mathsf{T}} \in \mathbb{C}^{M_p}, \tag{2}$$

where $f \in \{1 \ \cdots \ F\}$ and $t \in \{1 \ \cdots \ T\}$ are the indices of frequency bins and time frames, respectively. The superscript T denotes the transpose. We also denote the source index as $n \in \{1 \ \cdots \ N\}$.

In BSS, the observations at subarrays are modeled as

$$\boldsymbol{x}_{p,ft} = \boldsymbol{A}_{p,f}\boldsymbol{s}_{ft}, \tag{3}$$

where $\boldsymbol{A}_{p,f} \in \mathbb{C}^{M_p \times N}$ denotes the frequency-dependent and time-invariant mixing matrix. Under the determined condition, i.e., $M_p = N$, and assuming $\boldsymbol{A}_{p,f}$ is nonsingular, BSS can be achieved by the following demixing process:

$$\boldsymbol{y}_{p,ft} = \boldsymbol{W}_{p,f}\boldsymbol{x}_{p,ft}, \tag{4}$$

where $\boldsymbol{W}_{p,f} = \boldsymbol{A}_{p,f}^{-1}$, and $\boldsymbol{y}_{p,ft} = [y_{p,ft1} \ \cdots \ y_{p,ftN}]^{\mathsf{T}} \in \mathbb{C}^N$ denotes the separated signals at subarray p, which are estimates of the sources $\boldsymbol{s}_{ft}$. The goal of BSS is to obtain $\boldsymbol{y}_{p,ft}$ by estimating $\boldsymbol{W}_{p,f}$ without any prior information, such as the mixing matrix $\boldsymbol{A}_{p,f}$ and microphone positions.

In this paper, we assume the following conditions:

- The window length used in the STFT analysis is sufficiently longer than the impulse responses from sources to microphones, allowing the time-domain convolution to be approximated as multiplication in the STFT domain;
- The number of sources N is equal to the number of microphones in each subarray, i.e., $M_p = N$ for all p, which can be satisfied, e.g., by applying dimension reduction via principal component analysis (PCA).

B. Local BSS

In local BSS, source separation can be straightforwardly performed by estimating $\boldsymbol{W}_{p,f}$ and applying (4) at each subarray independently. Well-established BSS methods, such as IVA [4]–[6] and ILRMA [7], can be employed without modification for this process.

This approach relies solely on observations within each subarray; therefore, it inherently avoids privacy concerns, even when the distributed array is constructed from personal recording devices, as no information is shared beyond the local scope. However, this localized processing comes at the cost of one of the key advantages of distributed microphone arrays, namely, the ability to exploit spatial information collected across a wide area.

C. Centralized BSS

In contrast to local BSS, centralized BSS uses all observations simultaneously. First, the SROs between the microphone arrays are estimated and compensated to achieve synchronization [12]–[14]. Then, a conventional BSS algorithm is applied as if the signals were captured by a single microphone array. This procedure is illustrated in Fig. 1(a).

In the distributed microphone arrays, the total number of microphones, $M = \sum_{p=1}^{P} M_p$, is often greater than the number of sound sources, N. Under such overdetermined conditions, i.e., $M > N$, PCA is commonly applied beforehand to reduce the dimensionality and match the number of observations to that of sources [20]–[22]. With the projection matrix $\boldsymbol{B}_f \in \mathbb{C}^{N \times M}$ obtained through the PCA procedure, the separation can be performed as

$$\boldsymbol{y}_{ft} = \check{\boldsymbol{W}}_f \boldsymbol{B}_f \check{\boldsymbol{x}}_{ft} \in \mathbb{C}^N, \tag{5}$$

where

$$\check{\boldsymbol{x}}_{ft} = [\underbrace{x_{1,ft1} \ \cdots \ x_{1,ftM}}_{\boldsymbol{x}_{1,ft}^{\mathsf{T}}} \ \cdots \ \underbrace{x_{P,ft1} \ \cdots \ x_{P,ftM}}_{\boldsymbol{x}_{P,ft}^{\mathsf{T}}}]^{\mathsf{T}} \in \mathbb{C}^M. \tag{6}$$

Applying back-projection [23] onto the desired channel yields separated signals on any subarray. Note that BSS algorithms for overdetermined situations have also been proposed; see, e.g., [24], [25].

The advantage of centralized BSS is that conventional BSS algorithms can be directly applied, as signals from all microphones are aggregated. Moreover, since full observed information is available, high separation performance can be expected. However, in typical use cases assumed by conventional BSS studies, microphones are placed a few centimeters apart to avoid spatial aliasing. In large-scale arrays with meter-level spacing, the reliability of phase information becomes uncertain. In addition, centralized processing poses challenges in terms of computational complexity and user privacy.

D. Decentralized BSS

The idea of decentralized BSS, originally proposed for image signal processing, provides an intermediate approach between local and centralized BSS, as illustrated in Fig. 1(b). An algorithm based on this idea, decentralized IVA, was introduced in [19]. In that framework, most of the separation is carried out within each subarray, thereby reducing computational complexity. Moreover, only information for which an accurate reconstruction of the original observed waveform is inherently difficult is exchanged, which lowers privacy risks by avoiding the transmission of raw audio (e.g., from a user's smartphone) to external devices. Such audio may

contain sensitive information, including private conversations or background sounds, potentially raising privacy concerns.

III. Proposed Method

A. Motivation

As described in the previous section, decentralized BSS offers advantages over centralized BSS in terms of computational complexity and privacy protection. Moreover, since phase information of the observed signals is used only within each subarray, the method is expected to be more robust to the spatial separation between subarrays. However, existing decentralized IVA has been proposed for image signals, i.e., real-valued data, and its performance on acoustic signals, which are typically complex-valued in the STFT domain, remains unclear. In addition, the original method relies on optimization using the natural gradient method [19], whereas recent advances in BSS are predominantly based on the auxiliary function approach due to its improved convergence properties [6], [26].

To address these issues, this paper introduces an auxiliary-function-based optimization for decentralized IVA. This extension also serves as a foundation for developing decentralized variants of more advanced BSS algorithms introduced after IVA, such as ILRMA [7].

B. Formulation of decentralized IVA in STFT Domain

This section formulates the decentralized IVA in the STFT domain. We consider the following observation model:

$$\tilde{\boldsymbol{x}}_{f't} = \tilde{\boldsymbol{A}}_{f'} \tilde{\boldsymbol{s}}_{f't}, \tag{7}$$

where $f' = f + (p-1)F$ denotes the combined index of frequency bin and subarray, ranging over $f' \in \{1 \ \cdots \ F'\}$ with $F' = PF$. The observation $\tilde{\boldsymbol{x}}_{f't} \in \mathbb{C}^{M_p}$ and mixing matrix $\tilde{\boldsymbol{A}}_{f'} \in \mathbb{C}^{M_p \times N}$ are constructed by concatenating $\boldsymbol{x}_{p,ft}$ and $\boldsymbol{A}_{p,f}$ from each subarray along the frequency axis. The source signal $\tilde{\boldsymbol{s}}_{f't} \in \mathbb{C}^N$ is defined by replicating $\boldsymbol{s}_{ft}$ across all P subarrays and stacking them along the frequency axis, for notational convenience. Again, we assume $M_p = N$ for all p. If $\tilde{\boldsymbol{A}}_{f'}$ is nonsingular, the separated signal can be obtained similarly to the conventional BSS as:

$$\tilde{\boldsymbol{y}}_{f't} = \tilde{\boldsymbol{W}}_{f'} \tilde{\boldsymbol{x}}_{f't}. \tag{8}$$

Let $\tilde{\boldsymbol{w}}_{f'n} \in \mathbb{C}^{M_p}$ and $\boldsymbol{w}_{p,fn} \in \mathbb{C}^{M_p}$ denote the demixing vectors corresponding to source n, where

$$\tilde{\boldsymbol{W}}_{f'} = \left[\tilde{\boldsymbol{w}}_{f'1}^{\mathsf{H}} \ \cdots \ \tilde{\boldsymbol{w}}_{f'N}^{\mathsf{H}}\right]^{\mathsf{T}}, \tag{9a}$$

$$\boldsymbol{W}_{p,f} = \left[\boldsymbol{w}_{p,f1}^{\mathsf{H}} \ \cdots \ \boldsymbol{w}_{p,fN}^{\mathsf{H}}\right]^{\mathsf{T}}. \tag{9b}$$

Here, $\tilde{\boldsymbol{W}}_{f'}$ and $\tilde{\boldsymbol{w}}_{f'n}$ are formed by concatenating $\boldsymbol{W}_{p,f}$ and $\boldsymbol{w}_{p,fn}$ across all subarrays in the frequency direction, respectively. The superscript H denotes the Hermitian transpose.

On the basis of the above mixing model, we define a cost function $\mathcal{J}(\mathcal{W})$ of the same form as the conventional IVA [6]:

$$\mathcal{J}(\mathcal{W}) = \sum_{t=1}^{T} \sum_{n=1}^{N} G(\tilde{\boldsymbol{y}}_{tn}) - 2T \sum_{f'=1}^{F'} \log\left|\det \tilde{\boldsymbol{W}}_{f'}\right|, \tag{10}$$

$$\tilde{\boldsymbol{y}}_{tn} = [\tilde{y}_{1,tn} \ \cdots \ \tilde{y}_{F'tn}]^{\mathsf{T}} \in \mathbb{C}^{F'}. \tag{11}$$

Here, $\mathcal{W} = \{\tilde{\boldsymbol{W}}_{f'}\}_{f'=1}^{F'}$ denotes the set of demixing matrices for all frequency bins and subarrays, and $G(\tilde{\boldsymbol{y}}_{f't}) = -\log p(\tilde{\boldsymbol{y}}_{tn})$ is the contrast function. This paper employs the spherical Laplace distribution as in previous studies [4]–[6]:

$$p(\tilde{\boldsymbol{y}}_{tn}) \propto \exp(-\|\tilde{\boldsymbol{y}}_{tn}\|_2), \tag{12}$$

where $\|\cdot\|_2$ denotes the Euclidean norm. As in conventional methods, decentralized IVA is formulated as a minimization problem of (10) with respect to $\mathcal{W}$.

C. Auxiliary-Function-Based Decentralized IVA

The cost function (10) is closely related to auxiliary-function-based IVA (AuxIVA) and can be optimized in a similar manner [6]. Assuming the contrast function $G(\tilde{\boldsymbol{y}}_{tn})$ follows the spherical Laplace distribution, which is a type of spherically symmetric super-Gaussian distributions, the following $\mathcal{Q}(\mathcal{W}, \mathcal{R})$ serves as an auxiliary function for $\mathcal{J}(\mathcal{W})$; that is, $\mathcal{J}(\mathcal{W}) \leq \mathcal{Q}(\mathcal{W}, \mathcal{R})$ holds for any $\mathcal{W}$ and $\mathcal{R}$:

$$\mathcal{Q}(\mathcal{W}, \mathcal{R}) = \sum_{f'=1}^{F'} \mathcal{Q}_{f'}(\mathcal{W}, \mathcal{R}), \tag{13}$$

$$\mathcal{Q}_{f'}(\mathcal{W}, \mathcal{R}) = \frac{1}{2} \sum_{n=1}^{N} \tilde{\boldsymbol{w}}_{f'n}^{\mathsf{H}} \tilde{\boldsymbol{V}}_{f'n} \tilde{\boldsymbol{w}}_{f'n} - 2\log\left|\det \tilde{\boldsymbol{W}}_{f'}\right| + \mathrm{C}, \tag{14}$$

$$\tilde{\boldsymbol{V}}_{f'n} = \frac{1}{T} \sum_{t=1}^{T} \frac{\tilde{\boldsymbol{x}}_{f't} \tilde{\boldsymbol{x}}_{f't}^{\mathsf{H}}}{r_{tn}}, \tag{15}$$

where r_{tn} denotes the auxiliary variable, $\mathcal{R} = \{r_{tn}\}_{t=1,n=1}^{TN}$ denotes the set of auxiliary variables, and C is a constant term independent of $\tilde{\boldsymbol{W}}_{f'}$. Equality $\mathcal{J}(\mathcal{W}) = \mathcal{Q}(\mathcal{W}, \mathcal{R})$ holds if and only if

$$r_{tn} = \|\tilde{\boldsymbol{y}}_{tn}\|_2. \tag{16}$$

As shown above, the auxiliary function in decentralized AuxIVA is essentially the same as in the original AuxIVA formulation, except that the frequency index f is replaced with f', which also includes the subarray index.

1) Demixing Matrix Updates: First, we minimize $\mathcal{Q}(\mathcal{W}, \mathcal{R})$ with respect to $\mathcal{W}$. For subarray $p \in \{1 \ \cdots \ P\}$, by setting $\partial \mathcal{Q}(\mathcal{W}, \mathcal{R}) / \partial \boldsymbol{w}_{p,fn}^{*} = 0$, we obtain the following hybrid exact-approximate joint diagonalization (HEAD) problem:

$$\boldsymbol{w}_{p,fl}^{\mathsf{H}} \boldsymbol{V}_{p,fn} \boldsymbol{w}_{p,fn} = \delta_{ln}, \tag{17}$$

where δ_{ln} is the Kronecker delta, $l \in \{1 \ \cdots \ N\}$, and the weighted covariance matrix $\boldsymbol{V}_{p,fn} = \tilde{\boldsymbol{V}}_{f'n}$ is the local covariance within subarray. The superscript $*$ denotes the complex conjugate.

Now, since the HEAD problem (17) is equivalent to that in the conventional AuxIVA applied to the pth subarray, the following update rules are obtained:

$$\boldsymbol{w}_{p,fn} \leftarrow (\boldsymbol{W}_{p,f} \boldsymbol{V}_{p,fn})^{-1} \boldsymbol{e}_n, \tag{18}$$

$$\boldsymbol{w}_{p,fn} \leftarrow \frac{\boldsymbol{w}_{p,fn}}{\sqrt{\boldsymbol{w}_{p,fn}^{\mathsf{H}} \boldsymbol{V}_{p,fn} \boldsymbol{w}_{p,fn}}}, \tag{19}$$

TABLE I: Comparison of centralized, local, and decentralized IVA algorithms.

	Utilized information	Shared signal	Size of demixing matrix at f
Centralized	Inter-channel correlation	$\boldsymbol{x}_{p,ft}$	$M \times M$
Local	-	-	$M_p \times M_p$ $(\times P)$
Decentralized	Higher-order correlation	$\sum_f \lvert y_{p,ftn}\rvert^2$	$M_p \times M_p$ $(\times P)$

where $\boldsymbol{e}_n$ is a unit vector in which only the nth component is unity and zeros elsewhere.

2) Auxiliary variable updates: Next, we minimize $\mathcal{Q}(\mathcal{W}, \mathcal{R})$ with respect to $\mathcal{R}$. From the equality condition, the update rules are given by:

$$r_{tn} = \sqrt{\sum_{p=1}^{P} \sum_{f=1}^{F} \left| \boldsymbol{w}_{p,fn}^{\mathsf{H}} \boldsymbol{x}_{p,ft} \right|^2}, \tag{20}$$

$$\boldsymbol{V}_{p,fn} = \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{x}_{p,ft} \boldsymbol{x}_{p,ft}^{\mathsf{H}}}{r_{tn}}. \tag{21}$$

From the above, we obtain the update rules (18)–(21) for the proposed decentralized AuxIVA.

D. Discussion

Table I summarizes a comparison of the centralized, local, and proposed decentralized algorithms for AuxIVA.

1) Signal Model: In the model described by (7), the observed signals from each subarray are concatenated along the frequency axis. This structure allows the model to exploit not only signal correlations within subarrays but also higher-order correlations across them. This is analogous to conventional IVA, which also exploits higher-order correlations arising from frequency dependencies. Even if the microphone arrays are spatially separated, it is assumed that the frequency-domain activation patterns of each sound source are consistent across subarrays. This co-occurrence of source activity is a common assumption in conventional IVA. Thus, this formulation can be considered reasonable. This shared activation assumption may also help mitigate permutation ambiguity between subarrays.

2) Algorithm: From (18) and (19), it can be seen that the demixing matrix updates are performed independently within each subarray p. For updating the auxiliary variables, (20) requires aggregated information from all subarrays (i.e., the summation over p), where only information exchanged beyond subarrays is the power of the separated signals. Since it cannot directly reconstruct the original waveforms, the proposed algorithm is more privacy-preserving.

3) Computational Complexity: The size of the demixing matrix in decentralized IVA is $M_p \times M_p$ per frequency bin, which is the same as in local IVA. This is always smaller than that of centralized IVA, which is $M \times M$, since $M = \sum_p M_p$, $M_p \geq 2$, and $P \geq 1$ are required to perform IVA. Hence, decentralized IVA is computationally efficient while still leveraging the spatial diversity provided by the distributed microphone array.

TABLE II: Positions of sources and microphone arrays.

Label	(x, y) coordinates
Sources 1 and 2	(3 m, 4 m), (5 m, 1 m)
Microphone array 1	(2 m, 3 m), (2.04 m, 3 m)
Microphone array 2	(7 m, 3 m), (7.04 m, 3 m)

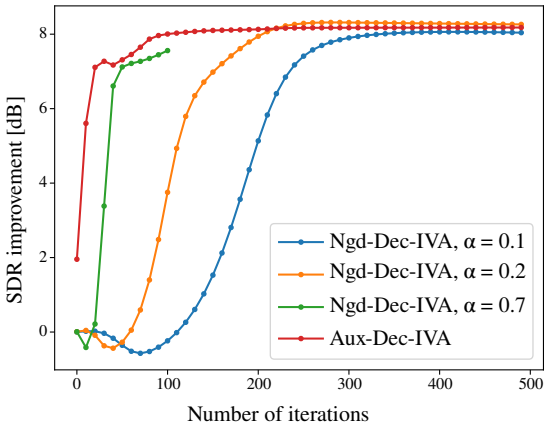


Fig. 2: Comparison of convergence speed of decentralized IVA algorithms, where α denotes the step size.

IV. Numerical Experiments

In this section, we conduct two numerical experiments using the Python package "Pyroomacoustics" [27] to evaluate the effectiveness of the proposed method. The first experiment compares the decentralized IVA based on the natural gradient method [19] with the proposed approach. The second experiment evaluates the separation performance of centralized, local, and decentralized AuxIVA.

A. Convergence

1) Experimental Condition: In this experiment, we assume a distributed microphone array setup in a relatively large room environment of $9\,\mathrm{m} \times 7.5\,\mathrm{m} \times 3\,\mathrm{m}$. We simulate a scenario with two sound sources and two microphone arrays, each equipped with two microphones. Dry speech signals from the development dataset of community-based Signal Separation Evaluation Campaign (SiSEC) [28] were used as source signals. The dataset comprises eight speech signals, two each from Japanese male, Japanese female, English male, and English female speakers, each lasting $10\,\mathrm{s}$. The sound sources were adjusted to achieve an signal-to-noise ratio (SNR) of $0\,\mathrm{dB}$, and then mixed through room impulse response (RIR) generated using the Pyroomacoustics package.

Microphone and source positions are summarized in Table II, with all of them placed at a height of $1.5\,\mathrm{m}$. The reverberation time was approximately $200\,\mathrm{ms}$, which may slightly deviate from the desired value due to the approximate modeling in pyroomacoustics. We used a Hann window of $128\,\mathrm{ms}$ with half-overlap for the STFT. The initial demixing matrices for both decentralized IVA algorithms were set to identity matrices. We evaluated performance using the signal-to-distortion ratio (SDR) [29], [30].

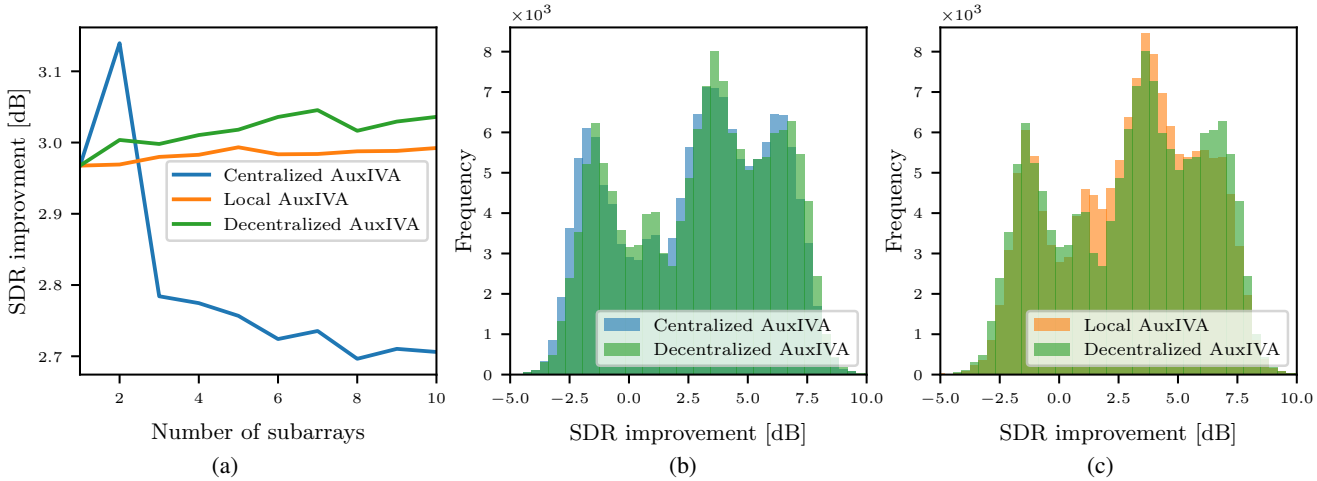


Fig. 3: Experimental results. (a) SDR improvement as a function of the number of subarrays. (b) Histogram of SDR improvement: centralized vs. decentralized. (c) Histogram of SDR improvement: local vs. decentralized.

2) Result and Discussion: The average SDR improvement over iterations is shown in Fig. 2. In the natural-gradient-based decentralized IVA (Ngd-Dec-IVA), convergence becomes faster as the step size α increases; however, it remains slower than the proposed auxiliary-function-based decentralized IVA (Aux-Dec-IVA). Although Ngd-Dec-IVA slightly outperforms the proposed method when $\alpha = 0.2$, it diverges at higher step sizes, making step size tuning challenging. In contrast, the proposed method exhibits faster and more stable convergence, achieving separation performance comparable to that of Ngd-Dec-IVA. This result is consistent with previous findings for AuxIVA [6].

B. Performance Evaluation

1) Experimental Condition: Next, two sound sources and one to ten two-channel microphone arrays were randomly placed in the same room as described in Section IV-A. All sound sources and arrays were positioned at least $50\,\mathrm{cm}$ away from the walls, without any constraints on their relative positions. The heights of the sound sources and arrays were set to $1.5\,\mathrm{m}$ and $1\,\mathrm{m}$, respectively. Two sound sources were randomly selected from the eight types introduced in Section IV-A. The average reverberation time was approximately $312\,\mathrm{ms}$. A Hann window of $512\,\mathrm{ms}$ with half-overlap was used for the STFT. Note that we assumed perfect synchronization across all microphones, i.e., no SRO was present.

We compared three methods: centralized, local, and decentralized AuxIVA. For the centralized method, PCA was applied in advance (except in the case of a single microphone array) to handle the overdetermined condition. Back-projection [23] was then applied to the first channel of each subarray to produce the same number of outputs as the other two methods. No whitening was applied for the local and decentralized methods. The number of iterations of all methods was 100.

2) Result and Discussion: Fig. 3 shows the result of 5000 trials. Note that all methods yield theoretically identical results when only a single subarray is used. As shown in Fig. 3(a) (average SDR improvement), the proposed decentralized IVA outperforms the centralized one when the number of subarrays exceeds two. Fig. 3(b) presents a histogram of the SDR improvements for each trial and for each number of subarrays. It can be observed that the proposed method consistently outperforms the conventional method across all conditions. In this experiment, the microphone arrays were spaced relatively far apart, which may have limited the effectiveness of dimensionality reduction using PCA. These results confirm the effectiveness of the proposed decentralized approach for audio signals, compared to the centralized one.

Despite the above, the performance of the proposed and local AuxIVA is nearly identical. From Fig. 3(c), we observe that the proposed method tends to yield more distinct separation results in terms of SDR improvement compared to local IVA. In particular, there is a noticeably higher concentration of cases around an SDR improvement of $7.5\,\mathrm{dB}$ with the proposed method, demonstrating its effectiveness. On the other hand, there is also an increase in cases with negative SDR improvement. This may be attributed to two situations: i) when all sources are located far from a microphone array, separation becomes inherently difficult; ii) when the sources are sufficiently distant from each other, little improvement can be expected due to sufficiently high input SDR. These results suggest that appropriate channel selection is critical for BSS on distributed microphone arrays.

V. Conclusions

In this paper, we proposed a decentralized IVA based on the auxiliary function method for distributed microphone arrays. This novel BSS framework utilizes the signals within subarrays and those powers shared across subarrays, reducing the risk of privacy invasion while maintaining separation performance. Experimental results demonstrated that the proposed method achieves faster and more stable convergence than the conventional natural-gradient-based decentralized IVA, similar to the advantages of AuxIVA. Moreover, it showed promising performance for BSS with distributed microphone arrays.

The proposed method assumes that each subarray has more microphones than sound sources, which may limit its applicability in practical scenarios. Future work also includes conducting experiments in cases of three or more sound sources and developing decentralized versions of more advanced BSS algorithms, such as ILRMA.

ACKNOWLEDGMENT

This work was supported by the Kajima Foundation's Support Program for International Joint Research Activities (2024-kyodoshin-05) and JSPS KAKENHI Grant Numbers 24K23854 and 25K21220.

REFERENCES

[1] S. Makino, T.-W. Lee, and H. Sawada, *Blind Speech Separation* (Signals and Communication Technology). Dordrecht: Springer, 2007.

[2] S. Makino, *Audio Source Separation* (Signals and Communication Technology). Cham: Springer, 2018.

[3] E. Vincent, T. Virtanen, and S. Gannot, Eds., *Audio Source Separation and Speech Enhancement*, 1st ed. John Wiley & Sons, 2018.

[4] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. ICA*, pp. 601–608, 2006.

[5] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.

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

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

[8] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *Proc. SCVT*, pp. 1–6, 2011.

[9] A. Bertrand, S. Doclo, S. Gannot, N. Ono, and T. van Waterschoot, "Special issue on wireless acoustic sensor networks and ad hoc microphone arrays," *Signal Process.*, vol. 107, pp. 1–3, 2015.

[10] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Commun. Mobile Comput.*, 2017.

[11] L. Wang and S. Doclo, "Correlation maximization-based sampling rate offset estimation for distributed microphone arrays," *IEEE/ACM Trans. ASLP*, vol. 24, no. 3, pp. 571–582, 2016.

[12] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Signal Process.*, vol. 107, pp. 185–196, 2015.

[13] A. Chinaev, P. Thüne, and G. Enzner, "Double-cross-correlation processing for blind sampling-rate and time-offset estimation," *IEEE/ACM Trans. ASLP*, vol. 29, pp. 1881–1896, 2021.

[14] Y. Masuyama, K. Yamaoka, T. Kawamura, and N. Ono, "Efficient joint optimization of sampling rate offsets using entire multichannel signal," *IEEE/ACM Trans. ASLP*, vol. 32, pp. 1816–1828, 2024.

[15] T. Nakashima, R. Ikeshita, N. Ono, S. Araki, and T. Nakatani, "Fast online source steering algorithm for tracking single moving source using online independent vector analysis," in *Proc. ICASSP*, pp. 1–5, 2023.

[16] M. Gunther, H. Afifi, A. Brendel, H. Karl, and W. Kellermann, "Network-aware optimal microphone channel selection in wireless acoustic sensor networks," in *Proc. ICASSP*, pp. 820–824, 2021.

[17] X.-L. Zhang, "Deep ad-hoc beamforming," *Comput. Speech Lang.*, vol. 68, pp. 1–18, 2021.

[18] R. Ikeshita, T. Nakatani, T. Ochiai, and S. Araki, "Maximizing predicted signal-to-distortion ratio: A new microphone selection criterion for beamforming in acoustic sensor networks," *IEEE/ACM Trans. ASLP*, vol. 33, pp. 2259–2274, 2025.

[19] N. P. Wojtalewicz, R. F. Silva, V. D. Calhoun, A. D. Sarwate, and S. M. Plis, "Decentralized independent vector analysis," in *Proc. ICASSP*, pp. 826–830, 2017.

[20] M. Joho, H. Mathis, and R. H. Lamber, "Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture," in *Proc. ICA*, pp. 81–86, 2000.

[21] S. Winter, H. Sawada, and S. Makino, "Geometrical interpretation of the PCA subspace approach for overdetermined blind source separation," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–11, 2006.

[22] C. Osterwise and S. L. Grant, "On over-determined frequency domain BSS," *IEEE/ACM Trans. ASLP*, vol. 22, no. 5, pp. 956–966, 2014.

[23] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomput.*, vol. 41, pp. 1–24, 2001.

[24] T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind separation for convolutive mixtures of speech based on multistage ICA using subarray processing," in *Proc. ICASSP*, vol. 1, pp. I-225–228, 2004.

[25] R. Scheibler and N. Ono, "Independent vector analysis with more microphones than sources," in *Proc. WASPAA*, pp. 185–189, 2019.

[26] S. Araki, N. Ito, R. Haeb-Umbach, G. Wichern, Z.-Q. Wang, and Y. Mitsufuji, "30+ years of source separation research: Achievements and future challenges," in *Proc. ICASSP*, pp. 1–5, 2025.

[27] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. ICASSP*, pp. 351–355, 2018.

[28] E. Vincent, S. Araki, and P. Bofill, "The 2008 Signal Separation Evaluation Campaign: A community-based approach to large-scale evaluation," in *Proc. ICA*, 2009.

[29] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.

[30] R. Scheibler, "SDR — Medium rare with fast computations," in *Proc. ICASSP*, pp. 701–705, 2022.