Sound field interpolation with unsupervised calibration for freely spaced circular microphone array in rotation-robust beamforming

Shuming Luan
Graduate School of Information Science
Nagoya University
Nagoya, Japan
shuming.luan@g.sp.m.is.nagoya-u.ac.jp

Yukoh Wakabayashi
Department of Computer Science
Toyohashi University of Technology
Toyohashi, Japan
wakayuko@cs.tut.ac.jp

Tomoki Toda
Information Technology Center
Nagoya University
Nagoya, Japan
tomoki@icts.nagoya-u.ac.jp

***Abstract*—This paper introduces a new approach to perform unequally spaced sound field interpolation (SFI) for beamforming using a freely spaced circular microphone array (CMA) that is robust to rotation. Unlike previous methods that required a known microphone distribution, the proposed approach uses unsupervised calibration to estimate the error angle of each microphone. This is accomplished through an innovative optimization problem, making it practical for CMAs with freely spaced microphones. By using only the number of channels and rotation angle, the proposed method enables accurate unequally spaced SFI. Experimental results show that the proposed method outperforms previous methods on freely spaced CMAs.**

***Index Terms*—Sound field interpolation, rotation-robust beamforming, unsupervised calibration, unequally spaced circular microphone array**

I. Introduction

Both humans and humanoid robots rely on the essential function of hearing, and numerous array signal processing techniques have been developed to enhance this function, including source separation and source enhancement. Advanced source separation methods such as beamforming [1], [2], independent low-rank matrix analysis [3], [4], and nonnegative matrix factorization [5], [6] are commonly used. However, these algorithms typically require the sound source and microphones remain stationary, so that the acoustic transfer system (ATS) is time-invariant and performance can be maintained. In cases where there is a time-variant ATS due to moving sources or sensors, time block processing can be used to mitigate the reduced performance. However, if the block length exceeds the time frame of the short-time Fourier transform (STFT), a delay corresponding to the block length is introduced, making it difficult to apply this method in real-time processing. Therefore, there is still considerable room for improvement.

This paper considers an auditory system composed of a circular microphone array (CMA) on the head of a human or humanoid robot. The CMA can rotate with the head to capture sound from a desired source in a noisy environment. However, this rotation results in a time-variant ATS and necessitates the re-estimation of the spatial filter, which is time-consuming, making real-time processing difficult.

Wakabayashi *et al.* proposed a sound field interpolation (SFI) technique [7] to make array signal processing robust to the rotation of an equally spaced CMA. After the CMA has rotated to a new position, based on the sound signal newly recorded at this new position, this method allows us to estimate what the sound signal would be like if it were observed at the original position before rotation. Consequently, the rotated CMA can be treated as a fixed unrotated one, thereby avoiding the need to re-estimate the spatial filter and addressing the bottleneck in online processing.

In practical applications, when a CMA is freely placed on a human or humanoid robot's head, it is not always possible to strictly maintain a uniform distribution of microphones due to hardware constraints and spacing limitations, making an unequally spaced CMA (unes-CMA) more common than an equally spaced one. Building on our research, an enhanced approach called unequally spaced SFI (unes-SFI) has been introduced that is robust to the rotation of an unes-CMA [8]. The method utilizes a modified version of [7] to compensate for errors in microphone positions in an unes-CMA, i.e., angular deviations from corresponding equally spaced positions on the same circle. Thus, an unes-CMA can be regarded as if it were evenly spaced. In this method, it was assumed the errors in each microphone's position were already known. However, these errors are not readily available in most practical circumstances, e.g., a wearable CMA where the microphones can be freely attached and detached by users, or a CMA consisting of two microphones worn as hearing aids on the ears along with several auxiliary microphones placed around the head.

In this study, we develop an innovative method for rotation-robust beamforming that can be applied to a freely distributed CMA without any prior knowledge of the microphone placement. Firstly, we introduce an unsupervised calibration method for an unes-CMA with an unknown distribution. This approach only needs to be executed once then each microphone's position on the CMA can be confirmed without the need for any prior location information. Secondly, we combine the unsupervised calibration and unes-SFI together

as pre-processing before beamforming. Since the distribution information is necessary for implementing unes-SFI, the unsupervised calibration can provide unes-SFI with the calibrated position of each microphone. Hence, when performing SFI on a freely spaced CMA, we only need to know the number of microphones and the rotation angle. And the rotation angle can be readily obtained through various means, such as using an acceleration sensor or employing other estimation methods [9]. Based on simulated experiments, we demonstrate that our proposed method performs very well on a freely spaced CMA.

II. Related Work

It is significant to note all the related works mentioned below share the common objective of avoiding updating the spatial filter after CMA rotates. Thus, the spatial filter before rotation is known in advance. Using the signal after rotation, it is possible to estimate the signal before rotation to directly use the previous spatial filter without re-estimating it.

A. Equally spaced sound field interpolation [7]

Assume that $x(\theta)$ is a continuous periodic function in the time-frequency domain on a circle's circumference, with 2π as the period and $\theta \in [0, 2\pi)$ as the spatial angle. Let x_m, $m \in \{0, ..., M-1\}$, be the STFT complex spectrum of the observed signal captured by an M-channel equally spaced CMA with interval $2\pi/M$. Here, $x(\theta)$ and x_m are both obtained after the CMA rotates Δ rad. The relationship between x_m and $x(\theta)$ can be represented as

$$x_m = x\left(2\pi\frac{m}{M}\right),\ m = 0, ..., M-1. \tag{1}$$

$x(\theta)$ can be reconstructed from x_m if the sampling theorem is satisfied [10]. Thus, SFI can be achieved using the non-integer sample theorem in the Fourier domain. The sound field before rotation, a $(-\Delta)$-rad-rotated sound field $x(2\pi m/M - \Delta)$, corresponds to a δ-sample-shifted discretized sound signal $x_{m+\delta}$, where $\delta = M(-\Delta)/2\pi$. According to the non-integer sample shift theorem in the DFT, $x_{m+\delta}$ is represented as

$$x_{m+\delta} = \sum_{n=0}^{M-1} x_n U_{m,n,\delta}. \tag{2}$$

$U_{m,n,\delta}$ can be numerically calculated [7]. In matrix representation, (2) can also be defined as

$$\begin{bmatrix} x_{0+\delta} \\ \vdots \\ x_{M-1+\delta} \end{bmatrix} = \begin{bmatrix} U_{0,0,\delta} & \cdots & U_{0,M-1,\delta} \\ \vdots & \ddots & \vdots \\ U_{M-1,0,\delta} & \cdots & U_{M-1,M-1,\delta} \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{M-1} \end{bmatrix} = \boldsymbol{U}_M(-\Delta)\boldsymbol{x}(\boldsymbol{0}), \tag{3}$$

where $\boldsymbol{U}_M(-\Delta)$ is the frequency-independent rotation transform matrix.

B. Unequally spaced sound field interpolation [8]

The conceptual diagram of this approach is shown in Fig. 1. We can define $\boldsymbol{\epsilon} = \begin{bmatrix}\epsilon_1 & \cdots & \epsilon_M\end{bmatrix}^\mathsf{T}$ as the vector of error angle, where ϵ_m indicates the angular deviation between the actual position of the mth microphone on the unes-CMA and

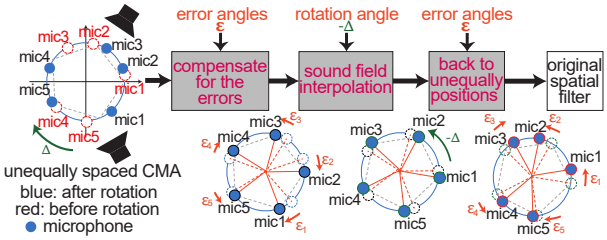

Fig. 1. Conceptual diagram of unequally spaced SFI.

its corresponding position in a uniformly spaced distribution. The sound field function observed by an unes-CMA after rotation can be expressed as

$$\boldsymbol{x}(\boldsymbol{\epsilon}) = \begin{bmatrix} x\left(\frac{2\pi\cdot 0}{M} + \epsilon_1\right) & \cdots & x\left(\frac{2\pi(M-1)}{M} + \epsilon_M\right)\end{bmatrix}^\mathsf{T}. \tag{4}$$

From (3), $\boldsymbol{\epsilon}$ is compensated for in the first step to generate a pseudo-signal recorded by a virtual equally spaced CMA,

$$\hat{\boldsymbol{x}}(\boldsymbol{0}) = \begin{bmatrix} x(0) & \cdots & x\left(\frac{2\pi(M-1)}{M}\right)\end{bmatrix}^\mathsf{T}, \tag{5}$$

which is calculated as

$$\hat{\boldsymbol{x}}(\boldsymbol{0}) = \boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon}), \tag{6}$$

where $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ is defined as

$$\boldsymbol{U}_M(\boldsymbol{\epsilon}) \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{u}_1(\epsilon_1) \\ \vdots \\ \boldsymbol{u}_M(\epsilon_M) \end{bmatrix}. \tag{7}$$

Here, $\boldsymbol{u}_m(\epsilon_n) \in \mathbb{C}^{1\times M}$ is the mth row of $\boldsymbol{U}_M(\epsilon_n) \in \mathbb{C}^{M\times M}$.

Following this, in the second step, we apply SFI to obtain the $(-\Delta)$-rad-rotated result of $\hat{\boldsymbol{x}}(\boldsymbol{0})$, which is captured by this virtual equally spaced CMA before rotation:

$$\boldsymbol{x}(-\Delta) = \begin{bmatrix} x(-\Delta) \cdots x\left(\frac{2\pi(M-1)}{M} - \Delta\right)\end{bmatrix}^\mathsf{T} = \boldsymbol{U}_M(-\Delta)\hat{\boldsymbol{x}}(\boldsymbol{0}). \tag{8}$$

Next, we must convert the virtual equally spaced CMA back to the real unes-CMA, whose signal is expressed as

$$\boldsymbol{x}(\boldsymbol{\epsilon} - \boldsymbol{\Delta}_M) = \begin{bmatrix} x(\epsilon_1 - \Delta) \cdots x\left(\epsilon_M - \Delta + \frac{2\pi(M-1)}{M}\right)\end{bmatrix}^\mathsf{T}, \tag{9}$$

where $\boldsymbol{\Delta}_M$ indicates an M-sized all-Δ vector. This signal can be calculated as

$$\boldsymbol{x}(\boldsymbol{\epsilon} - \boldsymbol{\Delta}_M) = \boldsymbol{U}_M(\boldsymbol{\epsilon})\boldsymbol{x}(-\Delta). \tag{10}$$

By combining (6), (8), and (10), it is possible to calculate the signal before rotation from the signal after rotation:

$$\boldsymbol{x}(\boldsymbol{\epsilon} - \boldsymbol{\Delta}_M) = \boldsymbol{U}_M(\boldsymbol{\epsilon})\boldsymbol{U}_M(-\Delta)\boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon}). \tag{11}$$

Finally, the original spatial filter before rotation can be directly applied to $\boldsymbol{x}(\boldsymbol{\epsilon} - \boldsymbol{\Delta}_M)$ without re-estimation.

III. Sound Field Interpolation with Unsupervised Calibration

A. Overview

A freely spaced CMA is considered in this proposed method. Considering the physical limitations of practical appli-

cations, we assume that microphones cannot be simultaneously placed in the same position. This approach differs from unes-SFI in that it does not need to know each microphone's position. Instead, only the number of microphones and the rotation angle, which could be easily determined, are necessary.

The main idea of our proposed method involves optimizing the error vector $\boldsymbol{\epsilon}$ by iterative updates. Fig. 2 illustrates the conceptual diagram of this main idea, called unsupervised calibration. With this method, we can calibrate each microphone using only the multichannel microphone signals without any prior knowledge of the distribution. The method consists of three steps. First, we assume a set of error values and choose one channel signal as a reference and the remaining $M-1$ channel signals as pseudo-observations. Secondly, we compensate for the angular deviations between each microphone, whose signal is chosen as pseudo observation, and its corresponding microphone on a virtual equally spaced $(M-1)$-channel CMA. The third step involves estimating the reference by applying SFI to the virtual equally spaced $(M-1)$-channel CMA. The estimated reference is then compared to the actual reference, and the error vector is obtained by minimizing the cost function between the two signals. This error vector can be used to perform unes-SFI as before.

B. Formulation

The observation obtained from the unes-CMA is represented by the same equation as (4). As the actual error vector $\boldsymbol{\epsilon}$ is unknown, we can assume a known error vector, denoted as $\boldsymbol{e} = \begin{bmatrix} e_1 & \cdots & e_M \end{bmatrix}^{\mathsf{T}}$, with an initial value of zero vector. Consequently, the initial position of each microphone is assumed to be $\boldsymbol{P}(\boldsymbol{e}) = \begin{bmatrix} 0+e_1 & \cdots & 2\pi(M-1)/M + e_M \end{bmatrix}^{\mathsf{T}}$.

We choose one channel, the first channel for instance, as the reference, $x_{\mathrm{ref}} = x(0+\epsilon_1)$, and select the remaining $M-1$ channels as pseudo-observations, represented as

$$\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{\epsilon}) = \begin{bmatrix} x\left(\frac{2\pi}{M}+\epsilon_2\right) & \cdots & x\left(\frac{2\pi(M-1)}{M}+\epsilon_M\right) \end{bmatrix}^{\mathsf{T}}. \quad (12)$$

By considering this pseudo-observation as a signal recorded by an unequally spaced $(M-1)$-channel CMA, we can apply the first step of unes-SFI to $\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{\epsilon})$ using (6) and position information $\boldsymbol{P}(\boldsymbol{e})$ to estimate the signal of a virtual equally spaced $(M-1)$-channel CMA, $\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{0}) = \begin{bmatrix} x(0) & \cdots & x\left(\frac{2\pi(M-2)}{M-1}\right) \end{bmatrix}^{\mathsf{T}}$. This can be computed as

$$\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{0}) = \boldsymbol{U}_{M-1}(\boldsymbol{e})^{-1}\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{\epsilon}), \quad (13)$$

$$\boldsymbol{U}_{M-1}(\boldsymbol{e}) = \begin{bmatrix} \boldsymbol{v}_1\left(\frac{2\pi\cdot 1}{M}+e_2-0\right) \\ \vdots \\ \boldsymbol{v}_{M-1}\left(\frac{2\pi(M-1)}{M}+e_M-\frac{2\pi(M-2)}{M-1}\right) \end{bmatrix}. \quad (14)$$

Here, $\boldsymbol{v}_m(\epsilon_n) \in \mathbb{C}^{1\times(M-1)}$ is the mth row of rotation matrix $\boldsymbol{U}_{M-1}(\epsilon_n) \in \mathbb{C}^{(M-1)\times(M-1)}$.

After obtaining the equally spaced signal $\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{0})$, we can use conventional SFI to obtain the estimated result of the reference signal, which can be expressed as

$$\hat{x}_{\mathrm{ref}} = \boldsymbol{v}_1(e_1)\boldsymbol{x}_{\mathrm{psd}}(\boldsymbol{0}). \quad (15)$$

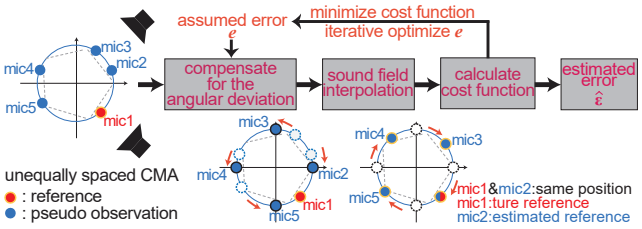

Fig. 2. Conceptual diagram of proposed method's main idea.

It is important to note that the calculation of $\boldsymbol{U}_{M-1}(\boldsymbol{e})$ and $\boldsymbol{v}_1(e_1)$ in this method employs the assumed value $\boldsymbol{e}$ instead of the actual error vector $\boldsymbol{\epsilon}$, as only the former is available and the latter remains unknown.

The cost function can be defined as the difference between $\hat{x}_{\mathrm{ref}}$ and x_{ref}, expressed as follows:

$$\mathcal{L}_1(\boldsymbol{e}) = 10\log_{10}\Big(\sum_{m,t,k}|\hat{x}_{\mathrm{ref},m,t,k} - x_{\mathrm{ref},m,t,k}|^2\Big), \quad (16)$$

where m, t, and k represent the channel, time frame, and frequency bin indexes, respectively.

Next, the second channel's signal $x(2\pi/M+\epsilon_2)$ is selected as the reference, and the same method as mentioned above is used to calculate a new loss function $\mathcal{L}_2(\boldsymbol{e})$. This can be repeated for all M channels, resulting in M loss functions denoted as $\mathcal{L}_1(\boldsymbol{e}), \mathcal{L}_2(\boldsymbol{e}), \ldots, \mathcal{L}_M(\boldsymbol{e})$, where each cost function corresponds to a different reference signal. By combining these cost functions, the final loss function $\mathcal{L}(\boldsymbol{e})$ is obtained:

$$\mathcal{L}(\boldsymbol{e}) = \sum_{i=1}^{M}\mathcal{L}_i(\boldsymbol{e}). \quad (17)$$

Calculating a closed-form solution for $\mathcal{L}(\boldsymbol{e})$ is difficult; however, it is differentiable with respect to $\boldsymbol{e}$, and back propagation can be used to estimate the error vector $\boldsymbol{\epsilon}$. Therefore, we minimize the cost function using a gradient descent method based on the Adam optimizer [11] to find the solution through iterative optimization,

$$\hat{\boldsymbol{\epsilon}} = \arg\min_{\boldsymbol{e}}\mathcal{L}(\boldsymbol{e}). \quad (18)$$

Despite (17) not being strictly convex, it possesses a global minimum. In the realm of gradient descent, attaining the global minimum is typically challenging. More often than not, we find ourselves converging towards a local minimum instead. Nevertheless, even though it may not be the absolute lowest point, this local minimum still showcases commendable effectiveness in performance. Please note that our calibration method does not use a deep neural network, and the optimization only searches for the optimum of $\boldsymbol{e}$ based on frame-wise observation.

After obtaining the estimated error vector $\hat{\boldsymbol{\epsilon}}$, we can apply the previous unes-SFI to generate the $(-\Delta)$-rad-rotated signal:

$$\hat{\boldsymbol{x}}(\boldsymbol{\epsilon}-\boldsymbol{\Delta}_M) = \boldsymbol{U}_M(\hat{\boldsymbol{\epsilon}})\boldsymbol{U}_M(-\Delta)\boldsymbol{U}_M(\hat{\boldsymbol{\epsilon}})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon}). \quad (19)$$

Finally, we can estimate the sound field as it would have been observed in its original position before rotation without requiring the exact value of $\boldsymbol{\epsilon}$.

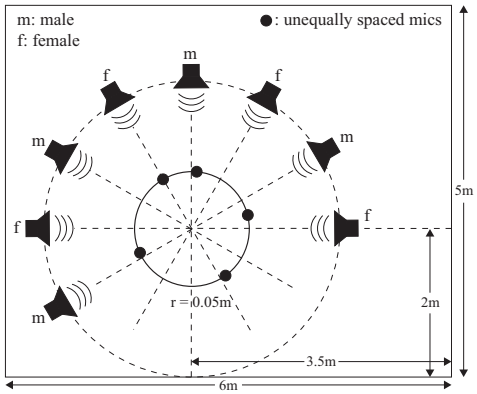

Fig. 3. Simulated environment in the experiments.

IV. Experimental evaluation

A. Setup

We conducted simulated experiments on the SiSEC database [12], each utterance of which was sampled at 16 kHz, to evaluate the efficacy and robustness of the proposed method. We selected eight speech signals, including four female and four male voices, as sound sources from various directions, as shown in Fig 3. To simulate a reverberant environment, we used an RIR generator [13] based on the image method [14] to create room impulse responses (RIRs) with a reverberation time of approximately 100 ms. We then convolved these RIRs with the sound source signals to obtain microphone signals. To perform in the time-frequency domain, we applied the STFT using a 1/8-shifted Blackman window with a length of 64 ms.

An M-channel CMA with 0.05 m radius was utilized to capture sound signals in a noise-free room. To generate an unes-CMA, an error was added to the position of each microphone. This error, denoted as ω_i (deg), $i \in \{0, ..., M-1\}$, followed a Gaussian distribution with zero mean and a standard deviation of 100 deg. Ten different unes-CMAs were produced. We additionally considered an unknown mismatch for each microphone's position, ϵ_i (deg), to make the actual distribution of the unes-CMA unknown. During unes-SFI, we assumed that the angular error of each microphone was ω_i because only ω_i was available, whereas the actual angular error should be $\omega_i+\epsilon_i$. We anticipated that this would degrade the performance of unes-SFI and that the proposed method could overcome the influence of unknown mismatches. ϵ_i was also subject to a Gaussian distribution with zero mean and variances ranging from 0 to 500 degrees-squared in increments of 10. For each variance, there were a total of 100 samples.

For the first experiment, we evaluated the performance in terms of the signal-to-error ratio (SER), which is defined as

$$\mathrm{SER}_{m,k} = 10\log_{10}\left(\frac{\sum_t |x_{m,t,k}|^2}{\sum_t |\hat{x}_{m,t,k} - x_{m,t,k}|^2}\right), \qquad (20)$$

where the signal $x_{m,t,k}$ and its estimate $\hat{x}_{m,t,k}$ are represented in the time–frequency domain. The values of M and the rotational angle were set to 5–6 and 10, 20, 30 deg, respectively. For the second experiment, we compared the performance of source enhancement using the minimum variance distortionless response (MVDR) beamformer [15]–[17], and evaluated the results using the source-to-distortion ratio (SDR) and source-to-interference ratio (SIR) [18] at a sampling frequency of 16 kHz. The beamformer's filter was estimated as:

$$\mathbf{w}_{\mathrm{MVDR}} = \frac{\mathbf{\Phi}_{nn}^{-1}\tilde{\mathbf{h}}_0}{\tilde{\mathbf{h}}_0^H \mathbf{\Phi}_{nn}^{-1}\tilde{\mathbf{h}}_0}, \qquad (21)$$

where $\mathbf{\Phi}_{nn}$ and $\tilde{\mathbf{h}}_0$ respectively denoted the covariance matrix of the interference signal and relative transfer function (RTF) [19]. Two random sources were mixed into the observation, with an angle between them of $30, 60, \ldots, 180$ deg. In this experiment, we set M and ϕ to be 5 and 20, 30 deg, respectively.

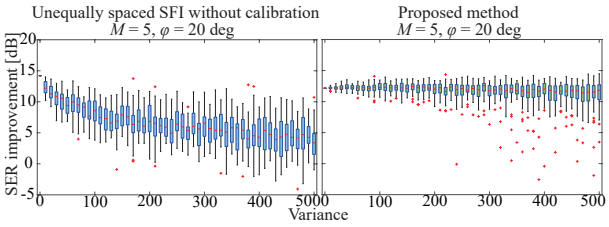

Fig. 4. Boxplots of the relationship between the variance of the mismatch and the SER improvement.

B. Evaluation results of sound field interpolation

Fig. 4 depicts the relationship between the variance of mismatch and SER improvement, with M set to 5 and ϕ set to 20 deg. The SER improvement measures the increase in the SER score achieved through processing. The baseline is the case without any processing, where the SER is computed by comparing the uninterpolated signals with the true signals. Since SFI struggles to estimate the higher-frequency components, we calculated the mean SER improvement over 0–1 kHz and five channels for each sample. As the variance of mismatch increases, the performance of unes-SFI without calibration deteriorates significantly. In contrast, our proposed method delivers a more consistent and superior result. Specifically, our method could accurately estimate the microphones' positions and maintain excellent interpolation performance despite substantial unknown mismatches.

Fig. 5 shows SER improvement for different M and ϕ, with a variance of mismatch set to 200. We calculated mean SER improvement over 0–1 kHz and M channels. The proposed method outperformed unes-SFI without calibration significantly. We also evaluated the performance of unes-SFI with known mismatch. While the proposed method had slightly lower improvement, which is expected due to the lack of prior information, the degeneracy was minimal and acceptable.

C. Evaluation results of source enhancement

First, the MVDR beamformer filter was computed at the original microphone position before any rotation. This filter, denoted as $\boldsymbol{w}$, was applied to the multichannel STFT spectrogram obtained from the unes-CMA without rotation, resulting in a reference performance denoted as **No-Rot**. This is the most favorable scenario as true signals, instead of interpolated

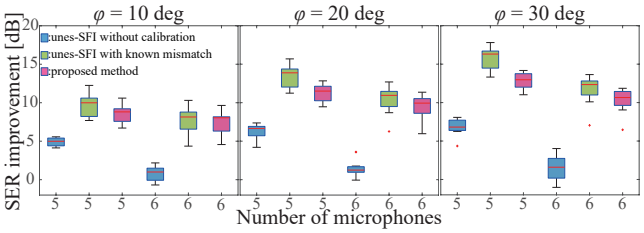

Fig. 5. Boxplots of mean SER improvement.

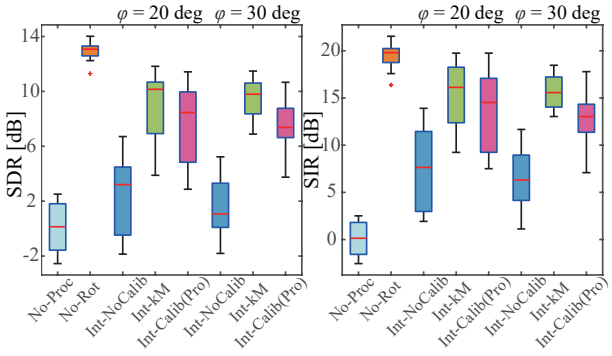

Fig. 6. Boxplots of SDR and SIR obtained by MVDR beamformer with different processing methods.

signals, are used for MVDR. After rotation, the spectrograms obtained from unes-SFI without calibration (**Int-NoCalib**), unes-SFI with known mismatches (**Int-kM**), and the proposed method (**Int-Calib(Pro)**), were processed by $\boldsymbol{w}$ to generate the estimated target signal. The unprocessed cases (**No-Proc**) and **No-Rot** were used as baselines for comparison.

Fig. 6 presents the SDR and SIR results for different methods when the variance of mismatch is 200. **Int-kM** and **Int-Calib(Pro)** performed better than **Int-NoCalib** and showed results closest to the highest performance, regardless of the simulated environment. The difference between **Int-kM** and **Int-Calib(Pro)** was still acceptably small. These results suggest our proposed method with unsupervised calibration can maintain robustness to the rotation of an unes-CMA and enhance the performance of array signal processing, even when the exact microphone distribution is unknown.

V. Conclusion

This paper introduced a novel framework of SFI with unsupervised calibration to improve beamforming robustness to unes-CMA rotation. By estimating error angles, we achieved unes-SFI without prior knowledge of microphone distribution. Experimental results showed high performance even with unknown mismatches. Future work includes addressing challenges such as non-circular microphone arrays, high-frequency component estimation, and real environment applications.

Acknowledgment

This work was partly supported by JST CREST Grant Number JPMJCR19A3, Japan.

References

[1] K. Yamaoka, N. Ono, S. Makino, and T. Yamada, "Time-frequency-bin-wise switching of minimum variance distortionless response beamformer for underdetermined situations," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 7908–7912.

[2] Y. Kubo, T. Nakatani, M. Delcroix, K. Kinoshita, and S. Araki, "Mask-based MVDR beamformer for noisy multisource environments: Introduction of time-varying spatial covariance model," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 6855–6859.

[3] 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 Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.

[4] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari, and N. Ono, "Independent deeply learned matrix analysis for determined audio source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 10, pp. 1601–1615, 2019.

[5] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, 2009.

[6] R. Scheibler and N. Ono, "Fast and stable blind source separation with rank-1 updates," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 236–240.

[7] Y. Wakabayashi, K. Yamaoka, and N. Ono, "Rotation-Robust Beamforming Based on Sound Field Interpolation with Regularly Circular Microphone Array," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 771–775.

[8] S. Luan, Y. Wakabayashi, and T. Toda, "Modified Sound Field Interpolation Method for Rotation-robust Beamforming with Unequally Spaced Circular Microphone Array," in *EUSIPCO 2022-2022 European Signal Processing Conference (EUSIPCO)*, 2022, pp. 344–348.

[9] G. Lian, Y. Wakabayashi, T. Nakashima, and N. Ono, "Self-rotation angle estimation of circular microphone array based on sound field interpolation," in *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2021, pp. 1016–1020.

[10] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] S. Araki, F. Nesta, E. Vincent, Z. Koldovskỳ, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 Signal Separation Evaluation Campaign (SiSEC2011):-audio source separation," in *International Conference on Latent Variable Analysis and Signal Separation*, 2012, pp. 414–422.

[13] E. A. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, p. 1, 2006.

[14] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[15] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[16] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[17] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2004.

[18] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[19] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, 2015.