# Unequally Spaced Sound Field Interpolation for Rotation-Robust Beamforming

Shuming Luan , Yukoh Wakabayashi , *Member, IEEE*, and Tomoki Toda , *Senior Member, IEEE*

*Abstract*—In this paper, we present an enhanced method designed to facilitate sound field interpolation (SFI) for rotation-robust beamforming using unequally spaced circular microphone arrays (unes-CMAs). Unlike the previous approach that necessitated an equally spaced circular microphone array (es-CMA), our method addresses the challenge of handling non-uniformly spaced microphones, making it suitable for real-world applications where unes-CMAs are more prevalent. Our proposed method enables the estimation of a virtual signal of an unes-CMA before rotation, derived from the observed signal after rotation. A modified SFI technique is utilized to compensate for the positional errors of microphones on an unes-CMA and to estimate a virtual signal at equally spaced positions after rotation. As an intermediate step, the previous SFI method is utilized to obtain equally spaced signals before rotation. Subsequently, the target signal of the unes-CMA before rotation is reconstructed, effectively achieving rotation-robust beamforming on the unes-CMA. Moreover, we provide an in-depth analysis of our proposed method's properties. We conducted simulated experiments, including online beamforming applications, to evaluate its performance. The experimental results demonstrated that our method effectively mitigates the adverse effects of unequal microphone placement, yielding significant improvements in estimating the signal before rotation under various conditions. Moreover, our proposed method consistently outperformed the previous approach, significantly enhancing the performance of beamforming.

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

## I. INTRODUCTION

AUDITORY information plays a pivotal role in daily communication for both human beings and humanoid robots. To enhance auditory data acquisition, a multitude of signal processing techniques have been developed, including source separation and source enhancement. Notably, the performance of advanced methods for source separation has been remarkably improved by altering spatial models, for example, beamforming [1], [2], [3], using source models such as independent vector

analysis [4], [5], [6], [7], nonnegative matrix factorization [8], [9], and using a variational autoencoder [10], [11], or their combinations such as independent low-rank matrix analysis [12], [13] and using a multichannel variational autoencoder [14], [15], [16]. However, it is essential to emphasize that most contemporary state-of-the-art approaches necessitate a time-invariant acoustic transfer system (ATS) to maintain their efficacy. The ATS refers to a system consisting of the sound source, the microphone sensors, and the acoustic transfer functions from the source to the sensors. This requirement implies that not only the sound sources but also the positions of the microphones must remain static during the execution of source separation and enhancement processes. Any modification in any aspect of these three components will lead to a time-variant ATS, necessitating the re-estimation of the spatial filter and making real-time processing challenging.

In this paper, we investigate an auditory system that comprises a circular microphone array (CMA) situated on the head of a human or humanoid robot. Specifically, we investigate an unequally spaced CMA (unes-CMA), wherein the angular intervals between adjacent microphones need not be uniform. In real-world scenarios, achieving a strictly uniform distribution of microphones on a CMA affixed to a human or humanoid robot's head is often impractical owing to hardware constraints and spatial limitations. Examples of unes-CMAs include wearable CMAs to which users can attach and detach microphones freely, or CMAs equipped with two microphones functioning as hearing aids on the ears, complemented by several auxiliary microphones positioned around the head. Such unes-CMAs can flexibly rotate with the head, facilitating sound capture from desired sources in noisy environments. In our analysis, we duly account for this typical variation in the ATS, that is, the rotation of the CMA. The CMA rotation introduces ATS variability, necessitating spatial filter re-estimation. This estimation typically relies on statistical information such as the covariance matrix, a computationally intensive process in most array signal processing methods. Consequently, the rotation of the CMA in our assumed scenario poses challenges for real-time online processing in practical environments.

The alteration in the ATS can be categorized into two distinct cases: situations involving moving sources and those involving moving sensors. In scenarios where a time-variant ATS arises owing to moving sources, reduced performance can be mitigated by time block processing [17], [18]. However, a challenge emerges when the block length surpasses the time frame of short-time Fourier transform (STFT), introducing a delay

corresponding to the block length, thereby impeding real-time processing applications. As a solution, Taseska and Habets [19] addressed online source separation by sequentially estimating the covariance matrix with direction-of-arrival (DOA) information estimation. In general, numerous methodologies have been proposed to tackle the challenges posed by moving sources.

Our method primarily focuses on cases involving moving sensors, which remains an area not well sufficiently researched. One possible approach to tackling this issue is the "stop-perceive-act" principle [20], which means halting movement to acquire data for estimation. The purpose of the "stop-perceive-act" principle is to maintain a time-invariant ATS when implementing array signal processing methods. However, this approach is impractical and unnatural in real-world applications, as the movement of a human or humanoid robot is typically continuous. Such an approach may hinder the microphone array from receiving new signals during motion, imposing behavioral constraints that limit natural interaction with the environment. If the movement is not excessively fast, the time-variant ATS can be approximated as a time-invariant ATS. Consequently, the "stop-perceive-act" principle can be satisfied, allowing the spatial filter to be applied without re-estimation. While this method can provide a certain level of effectiveness, the performance is susceptible to attenuation, similar to time block processing. The continuous movement, even if slow, will perpetuate the existence of a time-variant ATS. Hence, real-time processing necessitates the spatial filter to be updated continuously to achieve optimal performance. Therefore, continuous slow movement still impacts real-time processing. Tourbabin and Rafaely [21] proposed an alternative approach to DOA estimation using microphone arrays mounted on moving humanoid robots. They presented a motion compensation matrix based on the spherical harmonic domain to account for the robot's motion, utilizing the Wigner-D matrix to generate the rotation matrix. Additionally, Casebeer et al. [22] introduced a learning-based method employing a recurrent neural network for estimating time-varying spatial covariance matrices to address rapid pose changes in wearable devices.

An innovative beamforming framework [23], [24] robust to the rotation of a CMA has been proposed to address the challenges posed by moving sensors in online processing. This technique employs sound field interpolation (SFI) based on a non-integer sample shift theorem. SFI uses the sound field's periodicity along the circumference of a circle and the relationship between CMA sensing and sound field discretization. Upon the rotation of the CMA to a new position, before beamforming, this method utilizes the sound signal newly recorded at this position to perform SFI. This enables the estimation of how the sound signal would be if observed at its original position before rotation. Consequently, the rotated CMA can be treated as a fixed, unrotated one, eliminating the need to update the beamformer's filter. The previous filter can be directly applied to arbitrary sound fields when the CMA rotates. Extensive evaluations have demonstrated that this method enables a robust estimation of a low-frequency-band spectrum and achieves remarkable performance when applied to an existing beamformer, even under CMA rotation. Furthermore, the experimental results

of online beamforming highlighted the superiority of estimating the sound signal before the rotation of CMA compared to estimating the spatial information after rotation. It is critical to note that this framework requires an equally spaced CMA (es-CMA) for sound field sensing, as the discretized sound field needs to exhibit periodic behavior. Any angular deviation from equally spaced positions on the CMA's circumference, causing an unes-CMA, significantly diminishes the method's efficacy. In practical applications, unes-CMAs are more prevalent than es-CMAs, posing challenges to real-world implementation.

In [25], Ma et al. discussed two methods, namely cross-spectral matrix based on modal decomposition (CSM-MD) and cross-spectral matrix based on linear interpolation (CSM-LI). CSM-MD employs Fourier interpolation on the sound pressures at all real microphones to determine the sound pressures at the virtual rotating array (VRA) microphones. This CSM-MD method bears similarity to the SFI method in [23], [24], as both use linear interpolation on all microphones in the Fourier domain. However, Wakabayashi et al. [23], [24] offered clearer matrix-form analytical expressions for the interpolation. Due to the utilization of the periodicity of the sound pressure on the circumference of a circle, CSM-MD, like SFI, also requires the use of an es-CMA. Therefore, CSM-MD cannot be directly applied to the scenario in our study, where an unes-CMA is considered. Conversely, CSM-LI can be applied to an unes-CMA, as it employs linear interpolation only on the two neighboring microphones in the time domain. However, it has been established in [25] that CSM-LI exhibits a lower spectrum reconstruction capability, leading to inferior source enhancement performance. On the other hand, regarding computational efficiency, results from [25] indicate that CSM-MD and SFI consume approximately ten times more computation time than CSM-LI when the number of microphones is 64, primarily because CSM-MD and SFI involve pressures from all real microphones, whereas CSM-LI only utilizes pressures from two neighboring microphones. In our study, given that the number of microphones is much smaller than 64, which is more realistic for a wearable CMA in daily applications, CSM-LI's advantage in running speed will be completely lost.

This study focuses on a specific scenario previously described, namely, the rotation of an unes-CMA. In this context, two main challenges need to be addressed: ATS variation and the errors associated with the microphones' positions. With a focus on these two issues, we propose an improved approach known as unequally spaced SFI (unes-SFI), which is robust to a rotating unes-CMA. The initial step of this proposed framework involves compensating for the non-uniform microphone distribution before tackling ATS variation. Using the signals recorded by the unes-CMA, our method efficiently estimates the hypothetical sound signals that would be observed if each microphone were correctly positioned at its corresponding equally spaced location. Consequently, the unes-CMA is effectively treated as a virtual es-CMA. The complex issue of unes-CMA rotation is thus simplified to es-CMA rotation, reducing the challenge to the ATS variation–a matter that has been addressed in a prior work. Using the sound signals from the virtual es-CMA before rotation, our proposed method can successfully generate the sound signals of the actual unes-CMA before rotation. Thus, through the

implementation of our novel technique, we can virtually regard the time-variant ATS on the unes-CMA as a time-invariant ATS on an es-CMA. Note that our proposed framework is based on the concept presented in previous works [23], [24], which we utilize as an intermediate step. Nevertheless, our framework remains based on SFI and the non-integer sample shift theorem. We have modified some previous formulations to compensate for the errors associated with the microphones' positions on an unes-CMA.

This paper is an extension of a previously presented conference paper [26], wherein we introduced the concept of unes-SFI and its application to beamforming. On the basis of the conference paper's findings, this study expands on the analysis of unes-SFI by investigating its formulation and properties. Moreover, we conduct extensive experiments under various conditions to thoroughly investigate unes-SFI 's performance and extend its application to online beamforming processing. The subsequent sections of this paper are organized as follows. Section II begins with a comprehensive explanation of prior work on SFI, which serves as the foundation for the proposed approach. In Section III, we introduce and analyze the proposed unes-SFI method, highlighting its advancements and refinements. In Section IV, we outline the experimental setup and present the validation results obtained through rigorous evaluation. Finally, in Section V, we summarize the key findings and discuss potential avenues for future research.

## II. BACKGROUND

In this section, we present previous research studies [23], [24] on a time-variant ATS utilizing an es-CMA. Initially, we introduce the derivation process of the previous research studies. Subsequently, we conduct a new straightforward analysis of the singularity exhibited by the rotation transform matrix. This matrix represents the rotation of the es-CMA and plays a crucial role in generating the sound signal before rotation.

### A. Formulation

It is important to emphasize that the primary aim of the earlier research studies is to circumvent the need for updating the spatial filter after the CMA rotates. Consequently, the spatial filter before rotation is predetermined and remains known. By utilizing the signal observed after rotation, we can estimate the signal before rotation, enabling the direct usage of the pre-existing spatial filter without re-estimation. Thus, we shall consider a continuous sound field function denoted by $x(\theta)$ in the time–frequency domain, which is observed after the CMA rotates by $-\Delta$ rads, as shown in Fig. 1. $x(\theta)$ can be generally expressed as a sum of sinusoidal and cosinusoidal components in a Fourier series representation. Evidently, $x(\theta)$ exhibits periodicity with a period of $2\pi$, wherein $\theta \in [0, 2\pi)$ represents the spatial angle. When capturing the sound field using an es-CMA, it corresponds to discretizing the continuous sound field function $x(\theta)$, effectively sampling $x(\theta)$ at the locations where the microphones are positioned. To ensure that the discretized sound field function remains periodic as well, it is imperative to position the microphones on the CMA at equidistant intervals. Consequently, $x(\theta)$ can be discretized using an $M$-channel es-CMA with an
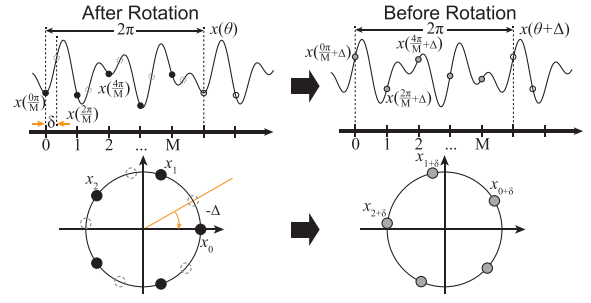


Fig. 1. Continuous sound field on a circle's circumference and the discretized sound field function with a $\delta$ sample shift.

interval of $2\pi/M$, leading to the representation of the observed signal in the $m$th channel as given by

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

where $x_m$ is the observed signal at the $m$th microphone after rotating the CMA. Note that there is no assumption on the signal model here. We do not need to worry about the specifics of the signal model. In other words, $x_m$ can be any type of signal model in any acoustic environment.

Assuming that the sampling theorem [27] holds, we can confidently reconstruct the continuous sound field function $x(\theta)$ from the discretized sound signal $x_m$. Hence, SFI becomes feasible, utilizing the non-integer sample shift theorem in the Fourier domain. Specifically, the sound field before rotation, denoted by a discretized $\Delta$-rad-rotated sound field function $x(2\pi m/M + \Delta)$, aligns consistently with the $\delta$-sample-shifted sound signal $x_{m+\delta}$ observed using an es-CMA,

$$x_{m+\delta} = x\left(\frac{2\pi m}{M} + \Delta\right), \qquad (2)$$

where $\delta = M\Delta/2\pi$. Using the non-integer sample shift theorem in the discrete Fourier transform (DFT), we can represent $x_{m+\delta}$ with $x_0, x_1, \ldots, x_{M-1}$ as

$$x_{m+\delta} = \frac{1}{M} \sum_{k=-M/2+1}^{M/2} \left(\mathscr{F}_{\mathrm{D}}[x_m] e^{j\Delta k}\right) e^{j\frac{2\pi mk}{M}}$$

$$= \sum_{n=0}^{M-1} x_n U_{m,n,\delta}, \qquad (3)$$

where $\mathscr{F}_{\mathrm{D}}[x_m]$ is the DFT of $x_m$. $U_{m,n,\delta}$ is the coefficient of SFI, which is computed by applying the $sinc$ function, as demonstrated below:

$$U_{m,n,\delta}$$

$$= \begin{cases} \frac{1-e^{jL\pi}}{M} + \frac{\mathrm{sinc}\left(\frac{L}{2}\right)\cos\left(\frac{M+2}{2M}L\pi\right)}{\mathrm{sinc}\left(\frac{L}{M}\right)}, & M \text{ is even}^1 \\ \frac{1}{M} + \frac{M-1}{M}\frac{\mathrm{sinc}\left(\frac{L(M-1)}{2M}\right)\cos\left(\frac{M+1}{2M}L\pi\right)}{\mathrm{sinc}\left(\frac{L}{M}\right)}, & M \text{ is odd,} \end{cases} \qquad (4)$$

where $L = n - m - \delta$ and $j = \sqrt{-1}$. According to Euler's formula, (4) in the case of an even $M$ can also be

---

[1]Compared with [23], we slightly modified the formulation of $U_{m,n,\delta}$ when $M$ is even. In [24], the formulation has been corrected.

written as

$$\text{Re}\left(U_{m,n,\delta}\right) = \frac{1 - \cos L\pi}{M} + \frac{\text{sinc}\left(\frac{L}{2}\right)\cos\left(\frac{M+2}{2M}L\pi\right)}{\text{sinc}\left(\frac{L}{M}\right)}, \quad (5)$$

$$\text{Im}\left(U_{m,n,\delta}\right) = -\frac{j\sin L\pi}{M}. \quad (6)$$

In matrix representation, the SFI formulation (3) can also be defined as

$$\boldsymbol{x}(\Delta) = \begin{bmatrix} x_{0+\delta} & \cdots & x_{M-1+\delta} \end{bmatrix}^{\mathsf{T}}$$

$$= \begin{bmatrix} U_{0,0,\delta} & \cdots & U_{0,M-1,\delta} \\ \vdots & \ddots & \vdots \\ U_{M-1,0,\delta} & \cdots & U_{M-1,M-1,\delta} \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{M-1} \end{bmatrix}$$

$$= \boldsymbol{U}_M(\Delta)\boldsymbol{x}, \quad (7)$$

where $\boldsymbol{U}_M(\Delta)$ is the rotation transform matrix, and $\boldsymbol{x} = \begin{bmatrix} x_0 & \cdots & x_{M-1} \end{bmatrix}^{\mathsf{T}}$ is the multichannel signal of the CMA after rotating $-\Delta$ rads in the time–frequency domain and is equal to $\boldsymbol{x}(0)$. Significantly, while all formulations are delineated within the confines of the time–frequency domain, it is noteworthy that this interpolation method in [23], [24] is not domain-restricted. In other words, even if the signal pertains to the time-domain, this interpolation technique remains applicable. It should also be emphasized that $\boldsymbol{U}_M(\Delta)$ takes the form of a cyclic matrix and remains independent of the observation's frequency.

### B. Analysis of the Rotation Transform Matrix

Given the application of the inverse matrix of the modified rotation transform matrix in our proposed method unes-SFI, it is pertinent to analyze certain properties of $\boldsymbol{U}_M(\Delta)^{-1}$. In the subsequent discussion, we designate the position of the es-CMA before rotation as the "reference position".

Evidently, if an es-CMA undergoes an initial rotation of $\Delta$ rads, followed by a subsequent rotation of $-\Delta$ rads, it will return to its reference position. This implies that the inverse matrix of $\boldsymbol{U}_M(\Delta)$ is equivalent to $\boldsymbol{U}_M(-\Delta)$. Furthermore, from the assertion in [24], $\boldsymbol{U}_M(\Delta)$ is also a unitary matrix.

In (4), when the number of microphones, $M$, is even, we encounter the challenge of dealing with the Nyquist frequency (Nyqf) component. As illustrated in (4), the numerator of the first term in the case of an even $M$, $e^{jL\pi} = (-1)^{n-m}e^{-j\delta\pi}$, represents the Nyqf component. In a previous paper [24], three approaches to handling this Nyqf component were proposed:

- Complex Nyqf (CoN): The entire Nyqf component is directly employed in its complex value form.
- Real Nyqf (ReN): Only the real part is considered.
- Zero Phase Nyqf (ZPN): A zero value is substituted for $\delta$ in only the Nyqf component.

In the subsequent part of this subsection, we focus on the scenario where $M$ is even. This limitation is applied to facilitate our investigation of the effect of the Nyqf component. By examining (5), (6), and the commutative property of matrix multiplication, we ascertain that without the imaginary part (6), $\boldsymbol{U}_M(\Delta) \cdot \boldsymbol{U}_M(-\Delta)$ is no longer equal to the identity matrix in

both ReN and ZPN. Consequently, the es-CMA cannot return to its original position after an initial rotation of $\Delta$ rads followed by a rotation of $-\Delta$ rads. Furthermore, the property of a unitary matrix is also no longer maintained.

Essentially, the inverse matrix of $\boldsymbol{U}_M(\Delta)$ is always existent, except for an exceptional scenario in ReN. Returning to the initial stages of the derivation for SFI, according to [24], (3) can be simplified as

$$x_{m+\delta} = \mathscr{F}_{\text{D}}^{-1}\left(\mathscr{F}_{\text{D}}\left[x_m\right]e^{j\Delta k}\right). \quad (8)$$

By substituting the Fourier transform $\mathscr{F}_{\text{D}}$ with the DFT matrix $\boldsymbol{F}$, we can translate (8) into the matrix representation

$$\boldsymbol{x}(\Delta) = \boldsymbol{F}^{-1}\boldsymbol{E}\left(\Delta\right)\boldsymbol{F}\boldsymbol{x}(0), \quad (9)$$

where

$$\boldsymbol{F} = \frac{1}{\sqrt{M}} \begin{bmatrix} e^{-j\frac{2\pi}{M}\cdot 0\cdot 0} & \cdots & e^{-j\frac{2\pi}{M}\cdot 0\cdot(M-1)} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{M}\cdot(M-1)\cdot 0} & \cdots & e^{-j\frac{2\pi}{M}\cdot(M-1)\cdot(M-1)} \end{bmatrix}, \quad (10)$$

and

$$\boldsymbol{E}\left(\Delta\right) = \text{diag}\left(e^{j\Delta\lceil 1-M/2\rceil}, \ldots, e^{j\Delta\lceil M/2\rceil}\right) \quad (11)$$

is a diagonal matrix of phase rotation for the $\delta$-sample shift, where $\lceil \bullet \rceil$ indicates the ceiling function.

The specific case where $\Delta$ is equal to $\pi/M$ results in the last element in $\boldsymbol{E}(\Delta)$ being $e^{j\pi/2}$. In ReN, where the imaginary part is neglected, the phase rotation matrix $\boldsymbol{E}_{\text{ReN}}(\pi/M)$ can be calculated as

$$\boldsymbol{E}_{\text{ReN}}\left(\frac{\pi}{M}\right) = \text{diag}\left(\cos\left(\frac{(1-M)\pi}{2M}\right), \ldots, 0\right), \quad (12)$$

which is a singular matrix. Therefore, in ReN, the rotation transform matrix $\boldsymbol{U}_M(\pi/M) = \boldsymbol{F}^{-1}\boldsymbol{E}_{\text{ReN}}(\pi/M)\boldsymbol{F}$ also lacks the corresponding inverse matrix. That is, in ReN, when $\Delta$ is equal to $\pi/M$, the inverse matrix of $\boldsymbol{U}_M(\Delta)$ does not exist. This abnormal situation warrants further discussion, which will be addressed in the next section.

## III. PROPOSED SOUND FIELD INTERPOLATION METHOD WITH UNEQUALLY SPACED CMA

In this section, we introduce a novel SFI method named unes-SFI specifically designed to address two key challenges: the time-variant ATS and the unes-CMA. Additionally, a comprehensive discussion on unes-SFI will be presented. Here, it is significant to note that the proposed method shares a common objective with previous research studies [23], [24], that is, to circumvent the need for spatial filter updates after the CMA rotates.

### A. Overview

In our proposed method, we assume that the error angle vector $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 & \cdots & \epsilon_M \end{bmatrix}^{\mathsf{T}}$ is already known beforehand, where $\epsilon_m \in (-2\pi/M, 2\pi/M)$ indicates the angular deviation between the actual position of the $m$th microphone on the unes-CMA and its corresponding position in a uniformly spaced distribution. As
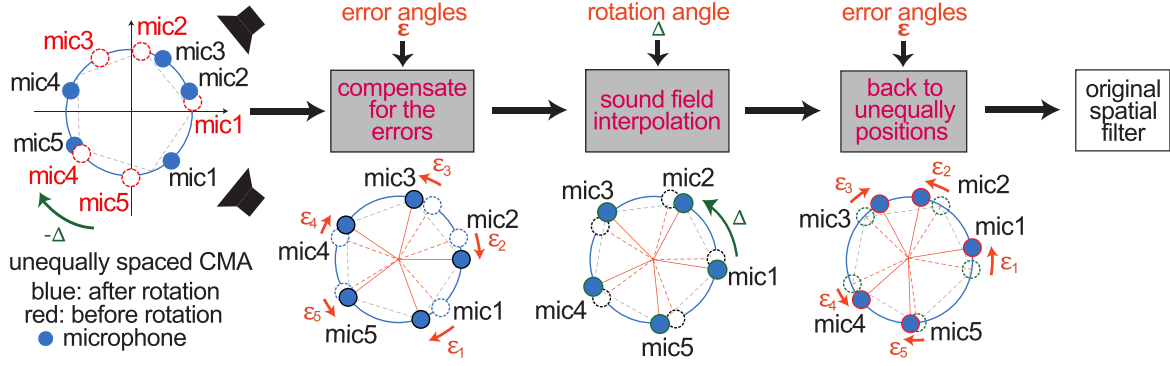
Fig. 2. Conceptual diagram of unes-SFI.

shown in Fig. 2, the proposed method encompasses three distinct steps. In the initial step, we utilize $\epsilon$ and the observation recorded by the unes-CMA after rotation to estimate the pseudo-signal observed by a virtual es-CMA. Subsequently, by incorporating the rotation angle $\Delta$, which can be readily obtained by various methods, such as using an acceleration sensor or employing other estimation methods [28], [29], [30], we generate the sound signal captured by this virtual es-CMA before rotation. Finally, we recover the non-uniform distribution, obtaining the signal captured by the unes-CMA before rotation. This allows us to directly apply the estimated results to other array signal processing methods.

### B. Formulation

In this subsection, we introduce the formulation of the proposed method. The error vector is $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 & \cdots & \epsilon_M \end{bmatrix}^\mathsf{T}$, and accordingly, the sound field function observed using an unes-CMA after rotating $-\Delta$ rads is expressed as

$$\boldsymbol{x}(\boldsymbol{\epsilon}) = \left[ x\left(\frac{2\pi \cdot 0}{M} + \epsilon_1\right) \quad \cdots \quad x\left(\frac{2\pi(M-1)}{M} + \epsilon_M\right) \right]^\mathsf{T}. \quad (13)$$

In (13), $\epsilon_m$ can be interpreted as the rotation angle from the position of the $m$th microphone in the es-CMA to the $m$th channel's position in the unes-CMA. The resultant rotated sound signal $\boldsymbol{x}(\epsilon_m)$ can be computed using the rotation transform matrix $\boldsymbol{U}_M(\epsilon_m)$. Furthermore, from (7), we observe that when targeting the $m$th channel signal of $\boldsymbol{x}(\Delta)$, only the $m$th row of matrix $\boldsymbol{U}_M(\Delta)$ is required. Likewise, considering the $m$th channel signal of the unes-CMA, $x(2\pi(m-1)/M + \epsilon_m)$, which simultaneously corresponds to the $m$th channel signal of $\boldsymbol{x}(\epsilon_m)$, we can define the relationship with the pseudo observation recorded using a virtual es-CMA,

$$\hat{\boldsymbol{x}}(0) = \left[ x(0) \quad \cdots \quad x\left(\frac{2\pi(M-1)}{M}\right) \right]^\mathsf{T}, \quad (14)$$

as follows:

$$x\left(\frac{2\pi(m-1)}{M} + \epsilon_m\right) = \boldsymbol{u}_m(\epsilon_m)\hat{\boldsymbol{x}}(0), \quad (15)$$

where $\hat{\bullet}$ denotes a pseudo observation from a virtual CMA and $\boldsymbol{u}_m(\epsilon_l) \in \mathbb{C}^{1 \times M}$ is the $m$th row of the rotation transform matrix $\boldsymbol{U}_M(\epsilon_l)$. Referring to (15), we can represent the entirety of the

observation recorded using an unes-CMA after rotation as

$$\boldsymbol{x}(\boldsymbol{\epsilon}) = \boldsymbol{U}_M(\boldsymbol{\epsilon})\hat{\boldsymbol{x}}(0), \quad (16)$$

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

$$\boldsymbol{U}_M(\boldsymbol{\epsilon}) \overset{\text{def}}{=} \left[ \boldsymbol{u}_1^\mathsf{T}(\epsilon_1) \quad \cdots \quad \boldsymbol{u}_M^\mathsf{T}(\epsilon_M) \right]^\mathsf{T}. \quad (17)$$

From (16), $\hat{\boldsymbol{x}}(0)$ can be calculated as

$$\hat{\boldsymbol{x}}(0) = \boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon}). \quad (18)$$

Consequently, in the first step, $\boldsymbol{\epsilon}$ is compensated for, yielding a virtual sound signal with a uniform distribution derived from the observation of an unes-CMA. In essence, the unes-CMA is virtually transformed into the es-CMA using the inverse matrix of $\boldsymbol{U}_M(\boldsymbol{\epsilon})$.

In the second step, SFI is employed to obtain the $\Delta$-rad-rotated result of $\hat{\boldsymbol{x}}(0)$. This result corresponds to the sound signal captured by the virtual es-CMA before rotation:

$$\hat{\boldsymbol{x}}(\Delta) = \left[ x(\Delta) \cdots x\left(\frac{2\pi(M-1)}{M} + \Delta\right) \right]^\mathsf{T} = \boldsymbol{U}_M(\Delta)\hat{\boldsymbol{x}}(0). \quad (19)$$

Subsequently, we need to convert the virtual equally spaced signal before rotation, $\hat{\boldsymbol{x}}(\Delta)$, back to the real unequally spaced signal, represented as

$$\boldsymbol{x}(\Delta + \boldsymbol{\epsilon}) = \left[ x(\Delta + \epsilon_1) \cdots x\left(\frac{2\pi(M-1)}{M} + \Delta + \epsilon_M\right) \right]^\mathsf{T}. \quad (20)$$

This signal can be computed using a formula similar to (16):

$$\boldsymbol{x}(\Delta + \boldsymbol{\epsilon}) = \boldsymbol{U}_M(\boldsymbol{\epsilon})\hat{\boldsymbol{x}}(\Delta). \quad (21)$$

By combining (18), (19), and (21), we can calculate the signal before rotation from the signal after rotation on an unes-CMA:

$$\boldsymbol{x}(\Delta + \boldsymbol{\epsilon}) = \boldsymbol{U}_M(\boldsymbol{\epsilon})\boldsymbol{U}_M(\Delta)\boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon}). \quad (22)$$

Upon completing these three sequential steps, the original spatial filter before rotation can be directly applied to $\boldsymbol{x}(\Delta + \boldsymbol{\epsilon})$ without re-estimation, enabling an efficient online processing.

### C. Analysis of the Compensation Matrix

As elucidated in the preceding section, three approaches are available for handling the Nyqf component: CoN, ReN, and

ZPN. In the subsequent analysis, we will investigate the property of the compensation matrix $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ under the impact of different Nyqf components in the proposed unes-SFI when the number of microphones, $M$, is even.

*1) Periodicity:* Evidently, in the previous SFI for an es-CMA [23], [24], the interpolation accuracy exhibits periodicity with respect to the rotation angle. This is due to the fact that rotating the es-CMA by the angle between adjacent microphones simply results in a shift in microphone index. This highlights the fact that the rotation transform matrix $\boldsymbol{U}_M(\Delta)$ transforms into an $M$-cyclic permutation matrix under such circumstances.

In our proposed unes-SFI, this periodicity is expected to persist, even when the angles between adjacent microphones differ from each other. For instance, assuming that $\boldsymbol{x}(\boldsymbol{\epsilon})$ is observed after the unes-CMA undergoes a rotation of $-\Delta$ rads, which corresponds to the angular deviation between two channels, specifically the $a$th and $b$th channels ($a \neq b$),

$$-\Delta = \left( \frac{2\pi(b-1)}{M} + \epsilon_b \right) - \left( \frac{2\pi(a-1)}{M} + \epsilon_a \right), \quad (23)$$

then the following equation holds:

$$x\left( \Delta + \frac{2\pi(b-1)}{M} + \epsilon_b \right) = x\left( \frac{2\pi(a-1)}{M} + \epsilon_a \right). \quad (24)$$

The left-hand side of (24) corresponds to the signal of the $b$th channel before rotating $-\Delta$ rads, whereas the right-hand side represents the signal of the $a$th channel after rotation. Thus, from (24), we can deduce that the position of the $b$th microphone before rotation aligns with the position of the $a$th microphone after rotation when the rotation angle is equal to the angular deviation between these two channels. Consequently, even in an unes-CMA, the interpolation accuracy of a particular channel remains cyclic with respect to the rotation angle, reaching its maximum whenever the rotation angle is equivalent to the angular deviation from other channels.

Equation (24) can be analytically proven. By applying the unes-SFI formulation (22), we can calculate the estimated signal of the $b$th channel before rotation as follows:

$$x\left( \Delta + \frac{2\pi(b-1)}{M} + \epsilon_b \right) = \boldsymbol{u}_b(\epsilon_b)\boldsymbol{U}_M(\Delta)\boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon}), \quad (25)$$

where $\boldsymbol{u}_b(\epsilon_b)\boldsymbol{U}_M(\Delta)$ can be expressed as

$$\boldsymbol{u}_b(\epsilon_b)\boldsymbol{U}_M(\Delta) = \boldsymbol{u}_b(\epsilon_b + \Delta) = \boldsymbol{u}_b\left( \epsilon_a + \frac{2\pi(a-b)}{M} \right). \quad (26)$$

By substituting $m = b$ and $\delta = M(\epsilon_a + 2\pi(a-b)/M)/2\pi$ into $L$ in (4), we obtain

$$L = n - b - \frac{M\left( \epsilon_a + \frac{2\pi(a-b)}{M} \right)}{2\pi} = n - a - \frac{M\epsilon_a}{2\pi}. \quad (27)$$

According to (4) and (7), it is easy to know that

$$\boldsymbol{u}_b(\epsilon_b)\boldsymbol{U}_M(\Delta) = \boldsymbol{u}_a(\epsilon_a). \quad (28)$$

Using (28), we can obtain (24) by simplifying the right-hand side of (25):

$$x\left( \Delta + \frac{2\pi(b-1)}{M} + \epsilon_b \right) = \boldsymbol{u}_a(\epsilon_a)\boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}\boldsymbol{x}(\boldsymbol{\epsilon})$$

$$= x\left( \frac{2\pi(a-1)}{M} + \epsilon_a \right). \quad (29)$$

However, (24) only holds in CoN and ReN. In ZPN, because $\delta$ is ignored, different from (27), $L$ becomes

$$L = n - b \neq n - a = n - b + (b - a). \quad (30)$$

Hence, the Nyqf components of the $b$th and $a$th channels are

$$e_b^{jL\pi} = (-1)^{n-b},$$

$$e_a^{jL\pi} = (-1)^{n-b+(b-a)} = (-1)^{n-b} \cdot (-1)^{b-a}. \quad (31)$$

The satisfaction of $e_b^{jL\pi} = e_a^{jL\pi}$ relies on the condition that $b - a$ is even. For the remaining terms in (4), where $\delta$ is not replaced by zero, the equivalence relation in (27) continues to hold. Consequently, (24) is not universally applicable in ZPN unless $b - a$ is even. Thus, the maximum interpolation accuracy of a specific microphone can be cyclically attained only when the microphone rotates to the position of other microphones and simultaneously skips an odd number of microphones.

*2) Singularity:* As can be seen from in (22), the inverse matrix of $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ holds a crucial significance in unes-SFI . Consequently, unes-SFI performance is significantly impacted by the singularity of $\boldsymbol{U}_M(\boldsymbol{\epsilon})$. Under normal circumstances, $\boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}$ exists, except in cases where two microphones are positioned at the same location on the unes-CMA, which is physically impossible in practical applications. However, in the preceding section, it was highlighted that an abnormal situation arises in ReN where $\boldsymbol{U}_M(\Delta)$ becomes singular. In light of this, we will further investigate whether a similar result arises for unes-SFI in ReN and endeavor to establish a more universally applicable conclusion for $\boldsymbol{U}_M(\boldsymbol{\epsilon})$.

Firstly, when $M$ is even, according to (5), the cos function in the latter term of $\text{Re}(U_{m,n,\delta})$ can be rewritten as

$$\cos\left( \frac{M+2}{2M}L\pi \right) = \cos\left( \frac{1}{2}L\pi + \frac{1}{M}L\pi \right)$$

$$= \cos\left( \frac{1}{2}L\pi \right)\cos\left( \frac{1}{M}L\pi \right) - \sin\left( \frac{1}{2}L\pi \right)\sin\left( \frac{1}{M}L\pi \right), \quad (32)$$

whereas the other part in the latter term is reformulated as

$$\frac{\text{sinc}\left( \frac{L}{2} \right)}{\text{sinc}\left( \frac{L}{M} \right)} = \frac{2}{M} \cdot \frac{\sin\left( \frac{L\pi}{2} \right)}{\sin\left( \frac{L\pi}{M} \right)}. \quad (33)$$

To ensure the validity of (33) for $\forall n, m \in [1, M]$, it is imperative to assume that $\delta \neq 0$, thereby preventing a zero denominator when $n = m$.

Thus, (5) is simplified to

$$\text{Re}\left( U_{m,n,\delta} \right) = \frac{(1 - \cos L\pi)}{M} + \frac{2}{M} \cdot \frac{\sin\left( \frac{L\pi}{2} \right)}{\sin\left( \frac{L\pi}{M} \right)}.$$

$$\left[\cos\left(\frac{L\pi}{2}\right)\cos\left(\frac{L\pi}{M}\right) - \sin\left(\frac{L\pi}{2}\right)\sin\left(\frac{L\pi}{M}\right)\right]$$

$$= \frac{1}{M}\left[1 - \cos\left(L\pi\right) + \frac{\sin\left(L\pi\right)\cos\left(\frac{L\pi}{M}\right)}{\sin\left(\frac{L\pi}{M}\right)} - 2\sin^2\left(\frac{L\pi}{2}\right)\right]$$

$$= \frac{1}{M} \cdot \frac{\sin\left(L\pi\right)\cos\left(\frac{L\pi}{M}\right)}{\sin\left(\frac{L\pi}{M}\right)} = -\frac{\sin\left(\delta\pi\right)}{M} \cdot V_{m,n,\delta}, \quad (34)$$

where $V_{m,n,\delta}$ is defined as

$$V_{m,n,\delta} = \frac{(-1)^{n-m}\cos\left(\frac{L\pi}{M}\right)}{\sin\left(\frac{L\pi}{M}\right)} = (-1)^{n-m}\cot\left(\frac{L\pi}{M}\right). \quad (35)$$

In ReN, as the imaginary part (6) is neglected, $U_{m,n,\delta}$ is simplified to $\mathrm{Re}(U_{m,n,\delta})$. Consequently, $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ in (17) can be redefined as

$$\boldsymbol{U}_M(\boldsymbol{\epsilon}) = \begin{bmatrix} -\frac{\sin(\delta_1\pi)}{M} & & \\ & \ddots & \\ & & -\frac{\sin(\delta_M\pi)}{M} \end{bmatrix}$$

$$\cdot \begin{bmatrix} V_{0,0,\delta_1} & \cdots & V_{0,M-1,\delta_1} \\ \vdots & \ddots & \vdots \\ V_{M-1,0,\delta_M} & \cdots & V_{M-1,M-1,\delta_M} \end{bmatrix}$$

$$= \boldsymbol{D}_M(\boldsymbol{\epsilon}) \cdot \boldsymbol{V}_M(\boldsymbol{\epsilon}), \quad (36)$$

where $\delta_i = M\epsilon_i/2\pi \in (-1,1)$, $i \in [1,M]$. Then, the determinant of $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ can be calculated as

$$\det\left(\boldsymbol{U}_M(\boldsymbol{\epsilon})\right) = \det\left(\boldsymbol{D}_M(\boldsymbol{\epsilon})\right) \cdot \det\left(\boldsymbol{V}_M(\boldsymbol{\epsilon})\right). \quad (37)$$

As previously indicated, it is assumed that $\delta_i$ is not equal to zero for $\forall i \in [1,M]$, consequently rendering $\det(\boldsymbol{D}_M(\boldsymbol{\epsilon}))$ as a non-zero constant. As a result, the matrix $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ is singular only when $\det(\boldsymbol{V}_M(\boldsymbol{\epsilon})) = 0$.

Here, for simplicity, we give a concrete example with $M = 2$ and $\boldsymbol{\epsilon} = [\epsilon_1 \quad \epsilon_2]^\mathsf{T}$. Then, $\boldsymbol{V}_M(\boldsymbol{\epsilon})$ is calculated as

$$\boldsymbol{V}_M(\boldsymbol{\epsilon}) = \begin{bmatrix} V_{0,0,\delta_1} & V_{0,1,\delta_1} \\ V_{1,0,\delta_2} & V_{1,1,\delta_2} \end{bmatrix}$$

$$= \begin{bmatrix} -\cot\left(\frac{\delta_1}{2}\pi\right) & \cot\left(\frac{(\delta_1-1)}{2}\pi\right) \\ \cot\left(\frac{(\delta_2+1)}{2}\pi\right) & -\cot\left(\frac{\delta_2}{2}\pi\right) \end{bmatrix}. \quad (38)$$

The determinant of $\boldsymbol{V}_M(\boldsymbol{\epsilon})$ can be obtained as

$$\det\left(\boldsymbol{V}_M(\boldsymbol{\epsilon})\right) = -\frac{4\cos\left(\frac{(\delta_1+\delta_2)}{2}\pi\right) \cdot \cos\left(\frac{(\delta_1-\delta_2)}{2}\pi\right)}{\cos\left(\frac{(\delta_1+\delta_2)}{2}\pi\right)^2 - \cos\left(\frac{(\delta_1-\delta_2)}{2}\pi\right)^2}. \quad (39)$$

Obviously, when $\delta_1 + \delta_2 = 1$ or $\delta_1 - \delta_2 = 1$, the determinant of $\boldsymbol{V}_M(\boldsymbol{\epsilon})$ will be zero. However, $\delta_1 - \delta_2 = 1$, which corresponds to $\epsilon_1 - \epsilon_2 = 2\pi/M$, indicates that two microphones are overlapped and placed in the same position. These conditions were

previously noted as physically implausible. Thus, $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ becomes singular when $\delta_1 + \delta_2 = 1$, which can also be expressed as $\epsilon_1 + \epsilon_2 = \pi$.

When there is a zero-value $\delta$, $e.g.$, $\delta_i = 0$ for the $i$th channel, the $i$th rows of $\boldsymbol{D}_M(\boldsymbol{\epsilon})$ and $\boldsymbol{V}_M(\boldsymbol{\epsilon})$, denoted by $\boldsymbol{d}_i(\epsilon_i)$ and $\boldsymbol{v}_i(\epsilon_i)$, respectively, will be adjusted as

$$\boldsymbol{d}_i(\epsilon_i) = \boldsymbol{v}_i(\epsilon_i) = \boldsymbol{I}_i, \quad (40)$$

where $\boldsymbol{I}_i$ is the $i$th row of an $M \times M$ identity matrix.

Taking $M = 4$ and $\boldsymbol{\epsilon} = [\epsilon_1 \quad \cdots \quad \epsilon_4]^\mathsf{T}$ with $\epsilon_4 = 0°$ as an example, the determinant of $\boldsymbol{V}_M(\boldsymbol{\epsilon})$ can be calculated as

$$\det\left(\boldsymbol{V}_M(\boldsymbol{\epsilon})\right) = \frac{4\cos\left(\frac{\delta_2-\delta_3+1}{4}\pi\right)}{\cos\left(\frac{\delta_1+1}{4}\pi\right)\sin\left(\frac{\delta_1}{2}\pi\right)} \cdot \frac{\sin\left(\frac{\delta_1-\delta_2-1}{4}\pi\right)}{\sin\left(\frac{\delta_2}{4}\pi\right)\cos\left(\frac{\delta_2}{2}\pi\right)}$$

$$\cdot \frac{\cos\left(\frac{\delta_1-\delta_3}{4}\pi\right)}{\sin\left(\frac{\delta_3+1}{4}\pi\right)\sin\left(\frac{\delta_3}{2}\pi\right)} \cdot \cos\left(\frac{\delta_1+\delta_2+\delta_3}{4}\pi\right). \quad (41)$$

Similarly, $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ is singular only when $\delta_1 + \delta_2 + \delta_3 = 2$, which can be reformulated as $\epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 = \pi$.

Revisiting the earlier conclusion in Section II-B, we can evidently see that in ReN, $\boldsymbol{U}_M(\Delta)$ becomes singular when $\Delta$ is equal to $\pi/M$, which in turn implies that the sum of $\Delta$ values from all $M$ channels is also equal to $\pi$. As a result, a more general conclusion can be drawn, stating that $\boldsymbol{U}_M(\boldsymbol{\epsilon})^{-1}$ will not exist when

$$\sum_{i=1}^{M} \epsilon_i = \pi. \quad (42)$$

We will further experimentally investigate whether (42) holds when $M$ is set to larger than 4 in Section IV-B2.

Note that despite the fact that the abnormal situation mentioned earlier is not taken into account, $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ may still occasionally be a nearly singular matrix. In such instances, we apply singular value decomposition (SVD) [31] on $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ and disregard the eigenvalues and eigenvectors associated with abnormally small condition numbers. Subsequently, the factorized coefficients of this truncated SVD [32] are employed to compute the inverse matrix of $\boldsymbol{U}_M(\boldsymbol{\epsilon})$.

## IV. EXPERIMENTAL EVALUATION

### A. Setup

To evaluate the performance and robustness of our proposed method to the rotation of an unes-CMA, simulation experiments were conducted. The SiSEC database [33] was used, with each utterance sampled at 16 kHz. From this database, eight speech signals were selected, consisting of four female and four male voices, as sound sources emanating from various directions, as illustrated in Fig. 3. To emulate a reverberant environment, sound signals were convolved with room impulse responses (RIRs) simulated by an RIR generator [34] on the basis of the image method [35]. This process yielded microphone signals with an approximate reverberation time of 100 ms. For conducting in the time–frequency domain, we employed the STFT utilizing a 1/8-shifted Blackman window with a length of 64 ms.
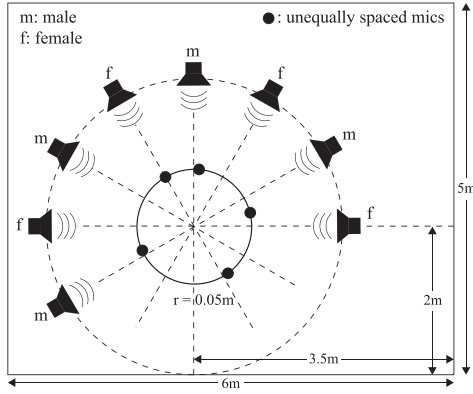
Fig. 3.    Simulated environment in the experiments.

The sound signals were recorded using an $M$-channel CMA with a radius of $0.05\,\text{m}$ in a noise-free room. To create an unes-CMA, an angle error, denoted by $\epsilon_i(°)$, $i \in \{1, \ldots, M\}$, was introduced to the position of each microphone. The angle error for each microphone followed a Gaussian distribution with zero mean and variances ranging from $(0°)^2$ to $(\sqrt{500°})^2$ in increments of $(\sqrt{10°})^2$. All the errors were independently and identically distributed. For each Gaussian distribution with a specific variance, 100 samples were generated. The simulation process proceeded as follows: initially, the sound field was simulated after the unes-CMA rotated $\Delta$ rads. Subsequently, this sound signal was employed to estimate the observation signals before rotation at the reference position, with the rotational angle $\phi = \Delta\pi/180°$ being a known value.

In the initial experiment, we assessed the performance in a scenario involving a single source, where sound sources were not mixed. The evaluation was based on the signal-to-error ratio (SER) [23], [24], [36], [37], [38] defined as

$$\text{SER}_{m,k} = 10 \log_{10} \left( \frac{\sum_t |x_{m,t,k}|^2}{\sum_t |\hat{x}_{m,t,k} - x_{m,t,k}|^2} \right), \qquad (43)$$

where $x_{m,t,k}$ is the time–frequency domain signal and $\hat{x}_{m,t,k}$ is its estimate. $m$, $t$, and $k$ denote the channel, time frame, and frequency bin, respectively. We conducted this experiment by varying the number of microphones, $M$, within the range of 3–6, and manipulating the rotation angle $\phi$.

In the second experiment, we employed the minimum variance distortionless response (MVDR) beamformer [39], [40], [41] to compare the source enhancement performance characteristics of different methods. The evaluation was based on the source-to-distortion ratio (SDR) [42]. From [23], [24], we utilized the covariance matrix of the interference signal and the relative transfer function (RTF) [43], [44] to estimate the beamformer's filter. The RTF was calculated using the RIR from the target source to each microphone. Then, two sources were randomly selected and mixed into the observation, with an angular separation between them set at $30°, 60°, \ldots, 180°$. This enabled us to simulate 12 environments, with two patterns at each angle.

### B. Results of Sound Field Interpolation

*1) Interpolation Accuracy:* Initially, we focus solely on the sound source in the direction of $0°$. Fig. 4 presents several examples of SER results obtained using the previous method [23], [24] and the proposed unes-SFI when the rotation angle $\phi$ ranges from $20°$ to $30°$ and $M$ is varied from 4 to 6. The mean SER of all $M$ channels is shown for a specific standard deviation ($10°$) of the error angle $\epsilon_i$. The results in Fig. 4 highlight that, in general, the proposed unes-SFI demonstrates superior capability to estimate the spectrum compared with the previous method. However, it should be noted that higher-frequency components are relatively challenging for both methods. To simplify the analysis, we limited the frequency range to 0–3 kHz for SER evaluation and averaged the SERs in decibels in all subsequent experiments.

*2) Effect of the Nyqf Component:* To present the results clearly and concisely, we consider two specific distributions of the unes-CMA with $M = 5$ and 6, with a standard deviation of $10°$: $[-8°, 85°, 150°, 221°, 295°]$ and $[2°, 46°, 117°, 171°, 253°, 295°]$. Here, we focus solely on the sound source in the direction of $0°$. Figs. 5 and 6 show the variation of the SER concerning the rotation angle, with the vertical axis representing the average SER over the frequency range of 0–3 kHz for the first channel, and the horizontal axis illustrating the rotation angle of the unes-CMA. The baseline curve represents the SER without any interpolation. When $M = 6$, the results of the previous method [23], [24] with ReN are obtained, which has been proven to be the most effective method in previous research [24]. Additionally, the SERs of unes-SFI with CoN, ReN, and ZPN are also displayed.

As observed in the figures, owing to erroneously treating the unes-CMA as an es-CMA, the previous method consistently performed worse than unes-SFI when $M = 5$ and unes-SFI with CoN and ReN when $M = 6$. However, it occasionally outperforms unes-SFI with ZPN only at a few rotation angles around $169°$ and $293°$. We also find that the SER of unes-SFI exhibits periodicity, as previously analyzed. When the first channel is rotated to the same position as another channel before rotation, the SER becomes maximum. However, in ZPN with $M = 6$, the SERs significantly differ from those in CoN and ReN when the first channel rotates to the second, fourth, and sixth channels, whereas they remain the same for the first, third, and fifth channels. This reflects the effect of ignoring the Nyqf component and supports the earlier conclusion that in ZPN, the interpolation accuracy of a specific microphone cyclically is maximum when the microphone rotates to the position of other microphones before rotation and skips an odd number of microphones simultaneously. Furthermore, when $M = 6$, ReN exhibits slightly higher performance at any rotation angle than CoN. As explained in [23], [24], the reason is that the presence of the complex-valued Nyqf component in CoN may adversely impact the performance of SFI. And it should be noted that the time consumption of ZPN, ReN, and CoN is generally the same because these three methods share the same formulation for calculation, as depicted in (4), and the difference among these three methods lies solely in how the Nyqf component is
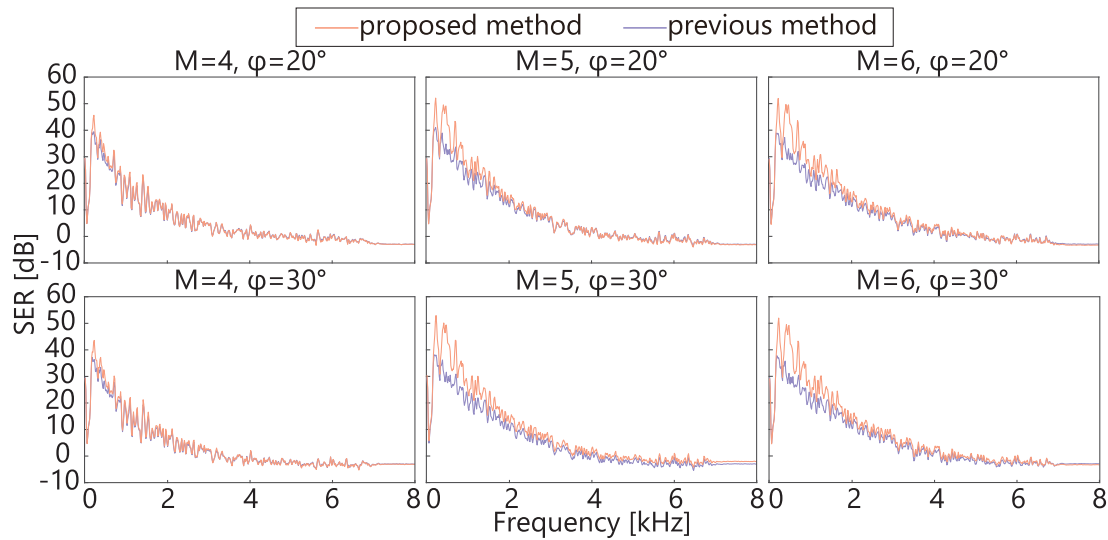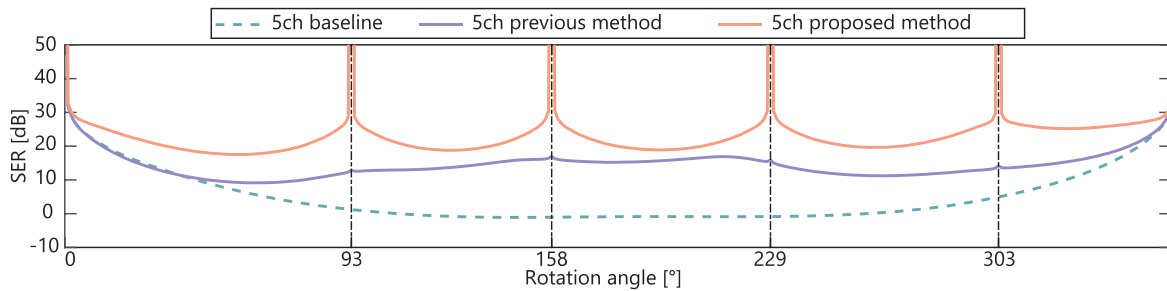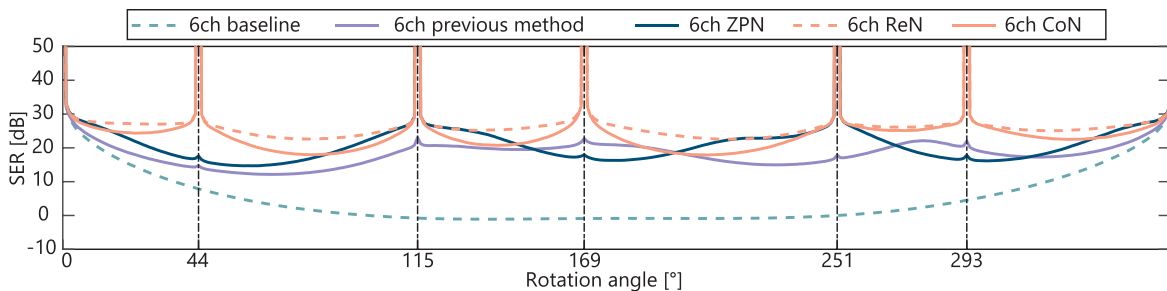
Fig. 4. Examples of SERs as a function of frequency.



Fig. 5. Dependence of SER on the rotation angle with $M = 5$, where the baseline indicates the cases without any interpolation.



Fig. 6. Dependence of SER on the rotation angle with $M = 6$, where the baseline indicates the cases without any interpolation, and ZPN, ReN, and CoN indicate neglecting the Nyqf component, considering only the real part of the Nyqf component and employing the Nyqf component's complex value form.

handled. Different approaches to handling this Nyqf component are unlikely to significantly impact the time consumption.

To assess the impact of the Nyqf component on the singularity of the compensation matrix $\boldsymbol{U}_{M}(\boldsymbol{\epsilon})$, we maintained the same microphone distributions as in the previous experiment, but allowed the position of the first channel's microphone to vary along the unes-CMA. Specifically, we moved it around a circle along the unes-CMA, ranging from $-8°$ to $351°$ when $M = 5$ and from $2°$ to $361°$ when $M = 6$. The condition number of $\boldsymbol{U}_{M}(\boldsymbol{\epsilon})$ was employed to quantify the singularity of $\boldsymbol{U}_{M}(\boldsymbol{\epsilon})$, where a larger condition number indicates a more singular matrix [45].

At each new position of the moving microphone, we rotated the unes-CMA by $20°$ and calculated both the condition number and SER.

Figs. 7 and 8 show the dependences of the condition number of the compensation matrix and the SER on the moving microphone position. The vertical axes on the left and right sides, respectively correspond to the condition number and the average SER of the microphone in the second channel of the initial distribution before relocating the microphone in the first channel. Irrespective of the moving microphone position, we consistently utilize the microphone located at $85°$ ($M = 5$) and
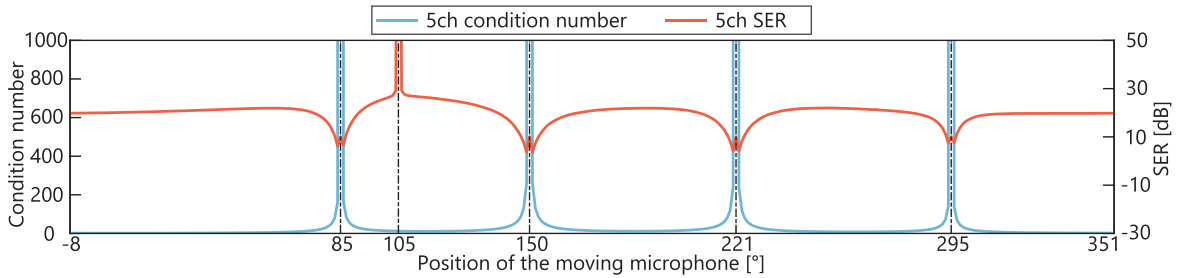
Fig. 7. Dependences of condition number and SER on the position of the moving microphone with $M = 5$.
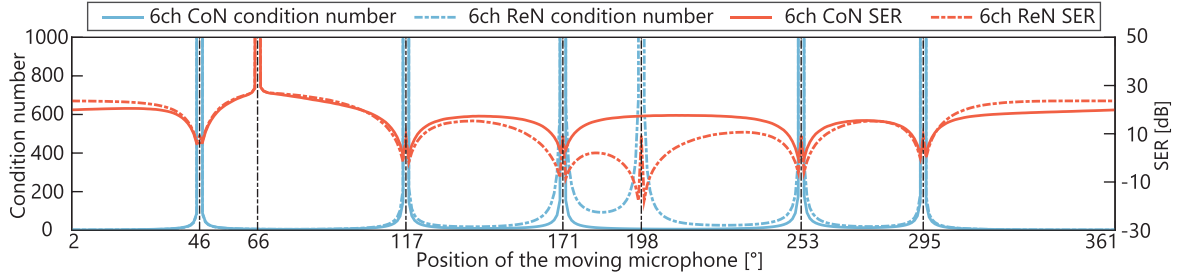


Fig. 8. Dependences of condition number and SER on the position of the moving microphone with $M = 6$, where ReN and CoN indicate considering only the real part of the Nyqf component and employing the Nyqf component's complex value form, respectively.

$46°$ ($M = 6$) before rotation to compute the SER result. Note that this choice is maintained even if the moving microphone alters the position order of channels, given the possibility of the channel position being not second after the microphone moved. The horizontal axis represents the moving microphone's position before rotation. For $M = 6$, only results for CoN and ReN are shown, as ZPN was previously found to be ineffective and unsuitable.

As observed in the figures, when the moving microphone is close to another microphone, the corresponding two rows in $\boldsymbol{U}_M(\boldsymbol{\epsilon})$ become similar to each other. This leads to a larger condition number, indicating a more singular $\boldsymbol{U}_M(\boldsymbol{\epsilon})$, owing to which the SER decreases. At certain positions, such as $105°$ in Fig. 7 and $66°$ in Fig. 8, the SER results are extraordinarily high, because after rotating by $20°$, the microphone used to calculate the SER aligns with the position of the moving microphone before rotation.

For $M = 5$ and $M = 6$ in CoN, there are four and five positions of the moving microphone causing the abnormally large condition number, respectively, where these positions coincide with the other fixed microphones. However, when $M = 6$ in ReN, there are six such positions, although we expected only five. The additional position angle is $198°$. At this position, the 6-channel distribution in unes-CMA is $[46°, 117°, 171°, 198°, 253°, 295°]$. Thus, the error vector $\boldsymbol{\epsilon}$ is $[46°, 57°, 51°, 18°, 13°, -5°]$. This arrangement leads to the sum of all error angles being equal to $180°$, which validates our previous conclusion in (42).

Note that the SERs at these singular positions are slightly higher than those nearby, owing to the application of truncated SVD at these positions, which reduces $\boldsymbol{U}_M(\boldsymbol{\epsilon})$'s singularity.

From Fig. 6, when $M = 6$, the interpolation accuracy of CoN is slightly lower than that of ReN, but the degree of degeneracy
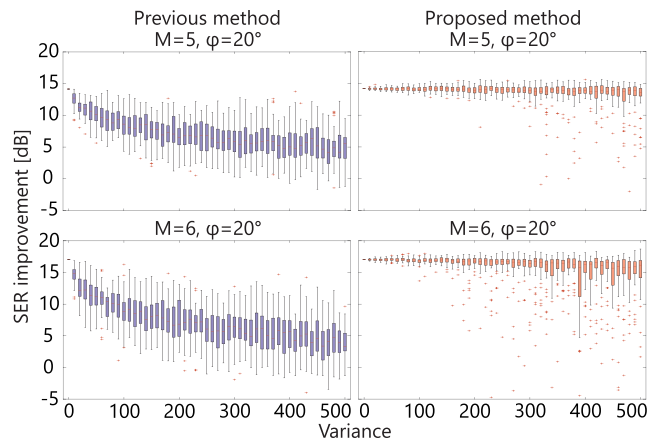


Fig. 9. Boxplots of the relationship between the variance of the error angle and the SER improvement at frequencies up to 3 kHz relative to the cases without interpolation.

is minimal and acceptable. From Fig. 8, it is evident that ReN can lead to an unexpected abnormal situation where the proposed unes-SFI fails owing to a singular compensation matrix. Consequently, CoN is deemed the most reasonable approach to handling the Nyqf component compared with the other two methods. Therefore, in subsequent evaluations, CoN will be employed when $M$ is an even number.

*3) Robustness to the Variance of Angle Error:* Fig. 9 shows the relationship between the variance of the error angles and the SER improvement for cases where $M$s are 5 and 6, and $\phi$ is $20°$, with the sound source located at $0°$. The SER improvement quantifies the increase in SER achieved through signal processing. The baseline used in SER improvement is obtained without any processing, where the SER is computed
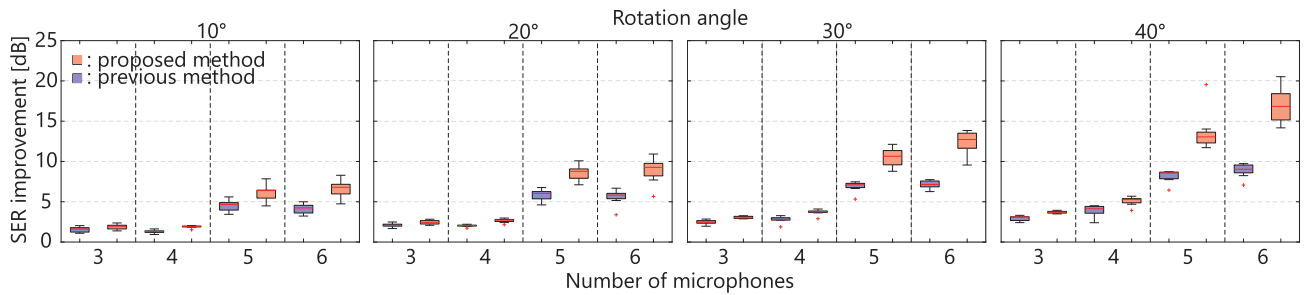
Fig. 10. Boxplots of mean SER improvement at frequencies up to 3 kHz for $M$ channels in eight situations.

by comparing the uninterpolated signal after rotation with the target signal before rotation. Each box in the graph represents the mean SER improvement over $M$ channels for each sample, resulting in 100 data points in each box.

The results clearly demonstrate that as the variance of the errors of angles increases, the SER improvement of the previous method [23], [24] experiences significant degradation. In contrast, our proposed method maintains its excellent interpolation performance, even with substantial errors of angles. This underscores the impracticality of directly applying ordinary SFI to an unes-CMA and highlights the advantages of employing our novel technique.

*4) Channelwise SER Improvements:* Fig. 10 illustrates the channelwise SER improvements obtained for various numbers of microphones $M$ and rotation angles $\phi$ with the standard deviation of error set to $10°$. Here, we use eight sound sources situated in different directions. The mean SER improvement relative to cases without interpolation is calculated over $M$ channels. Each box in Fig. 10 contains eight samples, corresponding to the mean SER improvement of eight sound sources.

The results demonstrate that the proposed method consistently exhibits greater SER improvement than the previous technique [23], [24] across all situations. As anticipated, in the proposed method, an increase in the number of microphones leads to enhanced performance owing to the higher spatial sampling rate. Conversely, in the previous method, using more microphones does not always lead to an improved SER and may even result in a poorer performance. This is observed when changing the number of microphones $M$ from 3 to 4 and from 5 to 6. This phenomenon can be attributed to the introduction of more errors in the rotation transform matrix $\boldsymbol{U}_M(\Delta)$ with more microphones, where the benefits of a higher spatial sampling rate do not outweigh the adverse effects of errors.

Furthermore, the proposed method achieves a greater SER improvement with an increase in the rotation angle. This can be attributed to the inferior performance of the case without any processing at higher rotation angles, where the proposed method's capability to compensate for microphone positions becomes more crucial.

*5) Robustness to the Error in Rotation Angle:* The rotation angle is a critical prior knowledge that must be known in advance for our proposed method. In cases where the rotation angles are inaccurately measured or estimated, the performance of our method could be affected. In this subsection, we explore a scenario where the rotation angle is not accurately measured and
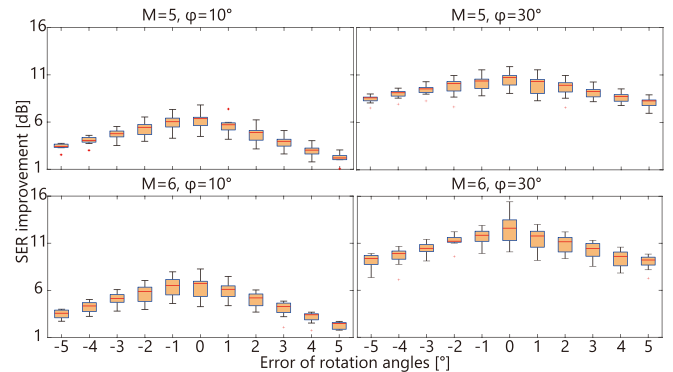


Fig. 11. Boxplots of mean SER improvement under different errors of rotation angle estimation.
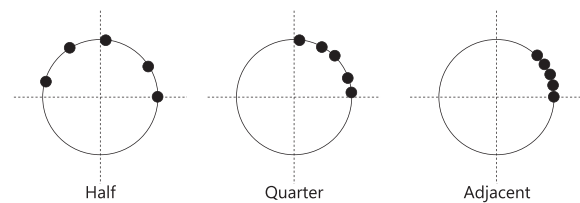


Fig. 12. Examples of extreme distributions.

investigate the robustness of our proposed method to errors in the rotation angle. In various methods for rotation angle estimation, the error in rotation angle estimation can be limited to within $5°$. Consequently, we present the channelwise SER improvement under varying errors in rotation angle estimation, ranging from $-5°$ to $5°$. For simplicity, here we only focus on the situations where the number of microphones $M$ is 5 and 6, rotation angle $\phi$ is $10°$ and $30°$. We continue to employ eight sound sources positioned in various directions. The mean SER improvement, relative to cases without interpolation, is computed over $M$ channels. Each box encompasses eight samples, corresponding to the mean SER improvement of eight sound sources. As depicted in Fig. 11, a degradation in SER improvement is observed with an increase in the absolute value of the error in rotation angle estimation, aligning with our expectations. However, the extent of degradation is not considerable, and our method remains effective to a certain degree in reconstructing the sound signal before rotation.

*6) Performance in Some Extreme Distributions:* The microphones in the unes-CMA are generally distributed unevenly
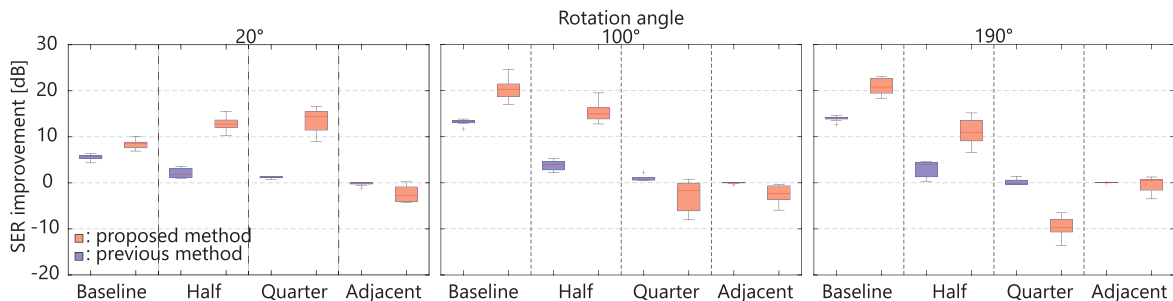
Fig. 13.    Boxplots of mean SER improvement at frequencies up to 3 kHz relative to the cases without interpolation in extreme distributions, where **Baseline**, **Half**, **Quarter**, and **Adjacent** represent distributions throughout the entire circle, distributions spanning only half of the circle, distributions spanning only a quarter of the circle, and distributions with microphones placed next to each other, respectively.

throughout the circular array, as shown in Fig. 3. However, in some atypical cases, the microphones may exhibit central clustering within certain regions of the unes-CMA. In this study, we considered three such extreme distributions when the number of microphones is 5: distributions spanning only half of the circle (**Half**), distributions spanning only a quarter of the circle (**Quarter**), and distributions with microphones placed next to each other at an angular interval of $1°$ (**Adjacent**). Examples of these extreme distributions are shown in Fig. 12. The baseline configuration corresponds to the typical scenario where microphones are unequally spaced throughout the entire circle (**Baseline**).

To evaluate the generalization capacity of the proposed method, we analyzed the SER improvement results at rotation angles of $20°$, $100°$, and $190°$, as illustrated in Fig. 13. The rotation angles of $100°$ and $190°$ were chosen to examine the performance of **Half** and **Quarter** when the microphones' positions after rotation fall outside the prerotation distribution range. The results indicate that the proposed method can accurately estimate the target signal with **Half** and **Quarter** at a rotation angle of $20°$. However, when rotating by $100°$ and $190°$, the proposed method demonstrates some effectiveness only for **Half**, although the SER improvement is notably smaller than that for the **Baseline**. **Adjacent** presents a significant challenge. In the case of **Adjacent**, it is observed that the proposed method cannot provide satisfactory results regardless of the rotation angle. These behaviors are expected as **Quarter** and **Adjacent** exhibit a small sampling range and limited spatial information available for interpolation beyond the original distribution.

Surprisingly, **Adjacent**'s performance is less inferior than **Quarter**'s when the rotation angles are $100°$ and $190°$, contrary to our initial expectations. This unexpected outcome can be attributed to the application of truncated SVD in Section III-C2, which effectively mitigated the issues with the compensation matrix for **Adjacent**, preventing extremely erratic performance. If the truncated SVD were not applied, all rows in $U_M(\epsilon)$ of **Adjacent** would be similar to each other, leading to a nearly singular $U_M(\epsilon)$ and a significantly deteriorating SER improvement. Thus, the observed instances where the performance of the previous method [23], [24] is not as poor as that of the proposed method can be attributed to the previous method's lack of necessity to address the sampling range. In addition, the
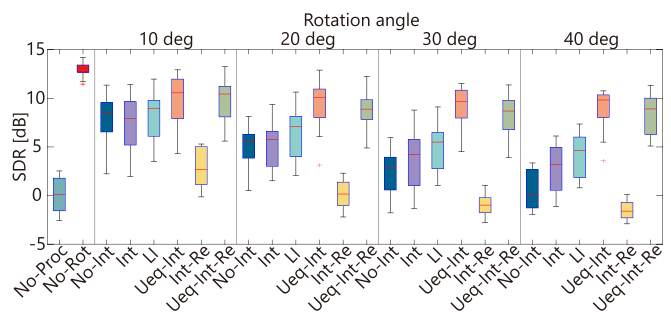


Fig. 14.    Boxplots of SDR obtained by MVDR beamformer in five situations: Unprocessed (**No-Proc**), no rotation of the CMA (**No-Rot**), without interpolation when the CMA rotates (**No-Int**), with ordinary SFI when the CMA rotates (**Int**), CSM-LI when the CMA rotates (**LI**), unes-SFI when the CMA rotates (**Ueq-Int**), re-estimation of the filter after ordinary SFI when CMA rotates (**Int-Re**), and re-estimation of the filter after unes-SFI when CMA rotates (**Ueq-Int-Re**).

previous method does not handle the ill-conditioned compensation matrix.

### C. Results of Source Enhancement With Batch Processing

In this experiment, we evaluate source enhancement performance using the MVDR beamformer. We fix the number of microphones at $M = 5$ and vary the rotation angle $\phi$ to $10°$, $20°$, $30°$, and $40°$. Firstly, we compute the filter weight $w$ for the MVDR beamformer using the RTF and the multichannel STFT spectrogram obtained from the unes-CMA at its original microphone position before any rotation. Then, the weight $w$ is applied to this spectrogram without rotation; this reference performance is denoted by **No-Rot**. This serves as the most favorable scenario as it uses true signals instead of interpolated signals for the MVDR beamformer.

We obtained the spectrograms of the unes-CMA at the positions after rotation, which we refer to as the spectrograms without being processed by any interpolation (**No-Int**). After the unes-CMA rotated, we estimated the interpolated spectrograms before rotation using the previous SFI method (**Int**) [23], [24], the CSM-LI method [25] which employs linear interpolation method only on the two neighboring microphones in the time domain (**LI**), and the proposed unes-SFI (**Ueq-Int**).
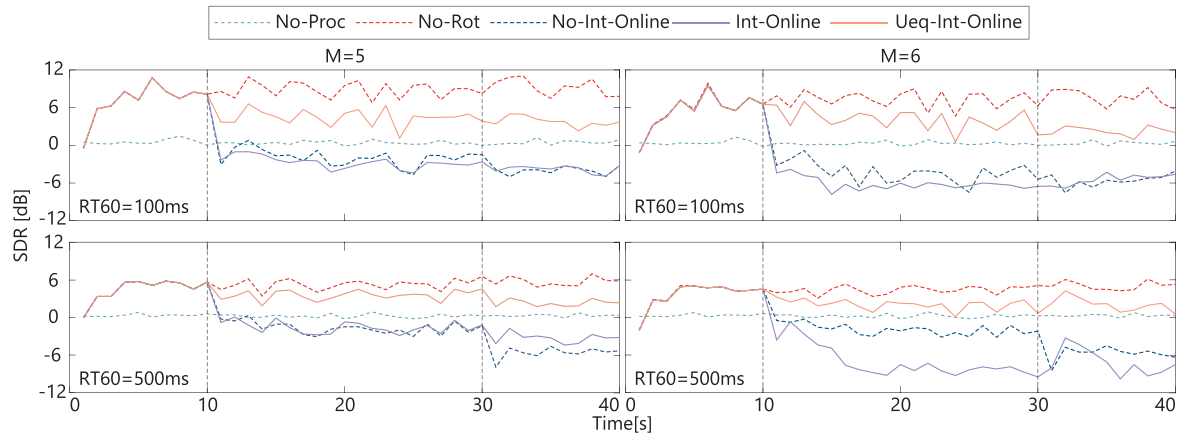
Fig. 15.   Segmental SDR every 1 s with $M = 5$ and $6$ and RT60 $= 100$ ms and $500$ ms, where the two vertical dashed lines indicate the time points when the rotation started: $0° \Rightarrow 20° \Rightarrow 40°$. **No-Proc** shows the mixture itself, **No-Rot** shows the case where rotation does not occur, **No-Int-Online** and **Int-Online** respectively show online processing without and with ordinary interpolation, and **Ueq-Int-Online** shows online processing with the unes-SFI.

Subsequently, we postprocessed these spectrograms using the MVDR beamformer's weight $w$ to generate the estimated target signal.

Additionally, we calculated a new MVDR beamformer using the same RTF before rotation and the interpolated spectrograms from **Int** and **Ueq-Int**, and applied this newly developed MVDR beamformer to the interpolated spectrograms; the results were denoted by **Int-Re** and **Ueq-Int-Re**, respectively. These results provide insights into the performance of online beamforming described in the next experiment. We used the unprocessed case (**No-Proc**), where the microphone signal was directly regarded as the target signal by mistake, and **No-Rot** as baselines for comparison.

The SDR results for different scenarios with a standard deviation of error set to $10°$ are presented in Fig. 14. As expected, **No-Proc** exhibits the lowest SDR, whereas **No-Rot** achieves the most significant source enhancement performance since the ATS remains time-invariant. Interestingly, the **Int** approach does not perform as well as anticipated. In most environments, **Int**'s SDR shows little difference from **No-Int**'s SDR, and in some circumstances, **Int**'s performance is even inferior to **No-Int**'s performance. These findings indicate that the previous method is ineffective when the CMA undergoes rotation owing to the non-uniformity of microphone spacing. In source enhancement using the MVDR beamformer, the previous interpolation provides only a slight improvement, and it is likely that a better source enhancement can be achieved without employing the previous method. In contrast, the proposed method (**Ueq-Int**) outperforms in both the case without interpolation (**No-Int**) and that with the previous interpolation technique (**Int** and **LI**), approaching the performance of the best-case scenario (**No-Rot**) regardless of the type of simulated environment used. The unequally spaced interpolation method demonstrates robustness to the non-uniform distribution of microphones on the CMA, significantly enhancing the array signal processing performance.

Furthermore, **Ueq-Int-Re** performs similarly to **Ueq-Int**, suggesting that unes-SFI can likely improve online processing as well, with only a slight decline due to the small mismatch between the covariance matrix estimated from the interpolated spectrogram and the pre-estimated RTF. Conversely, **Int-Re** exhibits poor source enhancement results, and is almost as ineffective as **No-Proc**. One of the main reasons for such degraded performance is that the previous method cannot precisely interpolate the spectrogram before rotation, resulting in a covariance matrix that entirely mismatched with the RTF.

### D.  Results of Source Enhancement With Online Processing

In this experiment, we propose the utilization of SFI for beamforming in an online processing scenario, taking into account continuous dynamic changes in the ATS. Online processing is designed to effectively handle minor variations in the ATS. To achieve this objective, we introduce a common smoothing factor denoted by $\alpha$ [46], [47], which enables the updating of spatial covariance during online processing. Additionally, we use the matrix inversion lemma, particularly the Sherman–Morrison formula [48], to alleviate the computational complexity associated with the covariance inversion in the MVDR formulation. By employing these techniques, we aim to increase the efficiency and effectiveness of the beamforming process in the presence of ATS variations.

It is noteworthy that the algorithm employed for online beamforming in this experiment bears a similarity to that utilized in the previous research [24]. The experimental conditions closely resemble those described in IV-A. However, there are some differences as follows. Two source signals were utilized, each with a duration of $40$ s. Additionally, we simulated the impulse response with reverberation times of $100$ ms and $500$ ms. The positions of the two sources were located at angles of $60°$ and $150°$, following the alignment shown in Fig. 3. The frame length was set to $256$ ms, and a segmental SDR with a length of $1$ s was employed to evaluate source enhancement performance. For the smoothing factor $\alpha$, we selected a value of $0.99$, which has been empirically validated to produce the highest segmental SDR. To initialize $\hat{V}_f^{-1}$, we used the inversion of the covariance matrix over the first $10$ frames. During the experiment, the unes-CMA

underwent two rotations: the first rotation commenced at $10\,\mathrm{s}$, progressing from $0°$ to $20°$, and the second rotation began at $30\,\mathrm{s}$, spanning from $20°$ to $40°$. Notably, the unes-CMA did not instantaneously rotate at $10\,\mathrm{s}$ or $30\,\mathrm{s}$ but rather underwent a gradual rotation at a uniform speed of $0.01°$ per time sample (equivalent to $160°$ per second). This rotation speed aligns with the typical average rotation speed for humans or humanoid robots. The observations in this experiment were generated by concatenating the observations of the unes-CMA after rotating at different angles in the simulation.

Fig. 15 presents the segmental SDR results obtained with $M = 5$ and 6, and reverberation times of $100\,\mathrm{ms}$ and $500\,\mathrm{ms}$. As shown, the **No-Rot** scenario, where the unes-CMA does not rotate, consistently achieves the most effective source enhancement performance. Surprisingly, unlike batch processing, the **No-Proc** method does not yield the lowest SDRs, contrary to initial expectations. The **Int-Online** method exhibits the poorest performance, sometimes even worse than the **No-Int-Online** method in some scenarios. These observations clearly indicate that the previous SFI technique [23], [24] is entirely ineffective when applied to an unes-CMA in online beamforming processing.

In comparison, our proposed method (**Ueq-Int-Online**) exhibits a significantly improved performance, achieving results closest to the highest performance (**No-Rot**) even in challenging environments with long reverberation time. This demonstrates the robustness and effectiveness of our approach in the context of online beamforming processing.

## V. CONCLUSION

In this paper, we introduced a novel framework for rotation-robust beamforming on an unes-CMA, building upon prior research. By extending the simple SFI, we developed an unes-SFI approach through the incorporation of a compensation matrix and modifications to the SFI technique. This framework effectively enables us to convert the time-variant ATS on an unes-CMA into a time-invariant ATS on an es-CMA. Consequently, we can estimate the signal of an unes-CMA before rotation, achieving rotation-robust beamforming. Furthermore, we conducted a detailed analysis of the compensation matrix's properties and the impact of the Nyqf component. Through a series of simulation experiments, we demonstrated that our newly presented system remains unaffected by non-uniform microphone distributions, effectively compensating for microphone positioning errors. Moreover, this method performs well in array signal processing scenarios, even during CMA rotation. However, our proposed method still requires further improvement, particularly when dealing with a small number of microphones on the CMA. Furthermore, in this study, we assumed knowledge of each microphone's position errors, but in practical situations, such information may not be available. Consequently, exploring the use of SFI on an unes-CMA without access to these known variables presents an interesting and worthwhile research direction, which we aim to investigate in our future work.

## 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 *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 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 *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 6855–6859.

[3] K. Yamaoka, N. Ono, and S. Makino, "Time-frequency-bin-wise linear combination of beamformers for distortionless signal enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3461–3475, 2021.

[4] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*, 2006, pp. 165–172.

[5] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. Independent Compon. Anal. Blind Signal Separation: 6th Int. Conf.*, 2006, pp. 601–608.

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

[7] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2011, pp. 189–192.

[8] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 550–563, Mar. 2010.

[9] R. Scheibler and N. Ono, "Fast and stable blind source separation with rank-1 updates," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 236–240.

[10] L. Pandey, A. Kumar, and V. Namboodiri, "Monoaural audio source separation using variational autoencoders," in *Proc. Interspeech*, 2018, pp. 3489–3493.

[11] E. Karamatli, A. T. Cemgil, and S. Kirbiz, "Audio source separation using variational autoencoders and weak class supervision," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1349–1353, Sep. 2019.

[12] 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, Sep. 2016.

[13] N. Makishima et al., "Independent deeply learned matrix analysis for determined audio source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 10, pp. 1601–1615, Oct. 2019.

[14] H. Kameoka, L. Li, S. Inoue, and S. Makino, "Semi-blind source separation with multichannel variational autoencoder," 2018, *arXiv:1808.00892*.

[15] S. Seki, H. Kameoka, L. Li, T. Toda, and K. Takeda, "Generalized multichannel variational autoencoder for underdetermined source separation," in *Proc. IEEE 27th Eur. Signal Process. Conf.*, 2019, pp. 1–5.

[16] H. Kameoka, L. Li, S. Inoue, and S. Makino, "Supervised determined source separation with multichannel variational autoencoder," *Neural Computation*, vol. 31, no. 9, pp. 1891–1914, 2019.

[17] J. Nikunen, A. Diment, and T. Virtanen, "Separation of moving sound sources using multichannel NMF and acoustic tracking," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 2, pp. 281–295, Feb. 2018.

[18] S. M. Naqvi, M. Yu, and J. A. Chambers, "Multimodal blind source separation for moving sources based on robust beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011, pp. 241–244.

[19] M. Taseska and E. A. P. Habets, "Blind source separation of moving sources using sparsity-based source detection and tracking," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 3, pp. 657–670, Mar. 2018.

[20] K. Nakadai, T. Matsui, H. G. Okuno, and H. Kitano, "Active audition system and humanoid exterior design," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2000, pp. 1453–1461.

[21] V. Tourbabin and B. Rafaely, "Direction of arrival estimation using microphone array processing for moving humanoid robots," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 2046–2058, Nov. 2015.

[22] J. Casebeer, J. Donley, D. Wong, B. Xu, and A. Kumar, "NICE-beam: Neural integrated covariance estimators for time-varying beamformers," 2021, *arXiv:2112.04613*.

[23] Y. Wakabayashi, K. Yamaoka, and N. Ono, "Rotation-robust beamforming based on sound field interpolation with regularly circular microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 771–775.

[24] Y. Wakabayashi, K. Yamaoka, and N. Ono, "Sound field interpolation for rotation-invariant multichannel array signal processing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 2286–2298, 2023.

[25] W. Ma, H. Bao, C. Zhang, and X. Liu, "Beamforming of phased microphone array for rotating sound source localization," *J. Sound Vib.*, vol. 467, 2020, Art. no. 115064.

[26] S. Luan, Y. Wakabayashi, and T. Toda, "Modified sound field interpolation method for rotation-robust beamforming with unequally spaced circular microphone array," in *Proc. IEEE 30th Eur. Signal Process. Conf.*, 2022, pp. 344–348.

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

[28] G. Lian, Y. Wakabayashi, T. Nakashima, and N. Ono, "Self-rotation angle estimation of circular microphone array based on sound field interpolation," in *Proc. IEEE Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2021, pp. 1016–1020.

[29] L. Wang, T.-K. Hon, J. D. Reiss, and A. Cavallaro, "Self-localization of ad-hoc arrays using time difference of arrivals," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 1018–1033, Feb. 2016.

[30] R. C. Felsheim, A. Brendel, P. A. Naylor, and W. Kellermann, "Head orientation estimation from multiple microphone arrays," in *Proc. IEEE 28th Eur. Signal Process. Conf.*, 2021, pp. 491–495.

[31] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A Practical Approach to Microarray Data Analysis*. Berlin, Germany: Springer, 2003, pp. 91–109.

[32] A. Falini, "A review on the selection criteria for the truncated SVD in data science applications," *J. Comput. Math. Data Sci.*, vol. 5, 2022, Art. no. 100064.

[33] S. Araki et al., "The 2011 signal separation evaluation campaign (SiSEC2011):-Audio source separation," in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, 2012, pp. 414–422.

[34] A. P. Habets, "Room impulse response (RIR) generator," 2008. [Online]. Available: https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator

[35] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[36] H.-L. Wei, S. A. Billings, Y. Zhao, and L. Guo, "An adaptive wavelet neural network for spatio-temporal system identification," *Neural Netw.*, vol. 23, no. 10, pp. 1286–1299, 2010.

[37] A. Wright, E.-P. Damskägg, L. Juvela, and V. Välimäki, "Real-time guitar amplifier emulation with deep learning," *Appl. Sci.*, vol. 10, no. 3, 2020, Art. no. 766.

[38] R. Abdelmalek, Z. Mnasri, and F. Benzarti, "Signal reconstruction based on the relationship between STFT magnitude and phase spectra," in *Proc. 8th Int. Conf. Sci. Electron., Technol. Inf. Telecommun.*, 2020, pp. 24–36.

[39] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. JPROC-57, no. 8, pp. 1408–1418, Aug. 1969.

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

[41] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.

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

[43] 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 Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[44] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.

[45] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Hoboken, NJ, USA: Wiley, 2005.

[46] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[47] T. Taniguchi, N. Ono, A. Kawamura, and S. Sagayama, "An auxiliary-function approach to online independent vector analysis for real-time blind source separation," in *Proc. IEEE 4th Joint Workshop Hands-free Speech Commun. Microphone Arrays*, 2014, pp. 107–111.

[48] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124–127, 1950.