Robust Beamforming for Multispeaker Audio Conferencing Under DOA Uncertainty

Gal Itzhak and Israel Cohen, *Fellow, IEEE*

***Abstract*—This paper presents a robust microphone array beamforming approach specifically designed for multispeaker audio conferencing, where the directions of arrival (DOAs) of the speakers can vary. First, we address the configuration of the array geometry. To achieve a consistent spatial response across all potential directions on the x-y plane, we propose a hybrid array geometry that combines a concentric circular array (CCA) on the x-y plane with a uniform linear array (ULA) along the z-axis. We discuss this geometry in detail and provide guidelines for selecting its parameters. Next, we focus on the beamforming strategy and introduce a method that directly controls the distortion level for signals originating from a specified near-end region (NER). In this context, we develop three high-directivity beamformers and evaluate their effectiveness based on white noise gain (WNG), directivity factor (DF), and beampattern characteristics. The proposed beamforming approach is experimentally assessed under various reverberant conditions, including both static and moving speech sources. Our results show that it outperforms existing methods in terms of perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) metrics. The improved performance is particularly pronounced in scenarios where the desired source direction differs from the nominal steering direction of the array, as well as when the reverberation level is mild.**

***Index Terms*—Microphone arrays, circularly symmetric arrays, audio conferencing, robust beamforming, direction of arrival (DOA).**

I. Introduction

SIGNAL-OF-INTEREST extraction from noisy observations is crucial in various fields. The use of microphone arrays, combined with beamformers, has been extensively researched and optimized. This has led to a diverse range of beamformer derivation methods and specific array designs, considering various criteria and design constraints [1], [2], [3], [4], [5], [6], [7]. Uniform Linear Arrays (ULAs) are traditionally favored for their simplicity in design and ease of analysis. However, this simplicity often results in limited design flexibility and suboptimal performance under real-world conditions. Challenges arise mainly when the desired signal's actual direction of arrival (DOA) differs from the assumed one or in reverberant environments [8], [9]. Additionally, ULAs require a larger physical size when the design includes more than a few microphones.

Developing two-dimensional array structures, such as rectangular arrays (RAs), has been a focus of recent research to enhance design adaptability. RAs demonstrate a reduced sensitivity to the DOA of the desired signal. They facilitate simultaneous optimization across multiple design criteria, as demonstrated in [10]. These structures are particularly effective in differential settings, where microphones are positioned nearby [11], [12], and are also beneficial for tasks involving DOA estimation [13], [14]. However, the inherent limited symmetry in RAs introduces a degree of DOA dependency, with a tendency to favor signals arriving parallel to one of the array's axes. Circular arrays (CAs) and uniform circular arrays (UCAs) offer greater symmetry in the x-y plane, which enables more flexible beamforming designs. This symmetry allows for enhanced control over desired performance measures and the potential for a frequency-independent spatial response alongside comprehensive array steering capabilities [15], [16], [17], [18]. Concentric circular arrays (CCAs) and uniform CCAs (UCCAs) further extend these geometries. They provide a consistent mainlobe beamwidth concerning azimuth and elevation angles, or both [19], [20], [21]. UCCAs also possess practical advantages such as compact array size and effectively handling white noise [22].

On top of the standard array geometries, arbitrarily shaped planar arrays have been investigated and optimized. For example, differential planar arrays were shown to enable steering of the mainlobe either by utilizing the Jacobi-Anger expansion [23], [24], [25] or by considering high-order beampattern design [26], [27]. Other studies attempted to leverage a planar geometry to attain a frequency-invariant beampattern [28] or a constant mainlobe beamwidth [29]. Nevertheless, these approaches were either not fully steerable, imposed limitations on the array's interelement spacings, or required complex beamforming design procedures that included heuristics. Additionally, their associated beamformers exhibited limited mainlobe beamwidths confined by the planar array's size and shape.

In many real-world situations, multiple speakers of interest may often move during or between periods of speech. A typical example is a conference room setting where participants sit around a table and engage in a discussion, often moving from their seats or actively presenting. In these scenarios, communication devices with built-in microphone arrays and loudspeakers, typically located nearby, are frequently used to extend the meeting to online participants. To effectively capture the desired

Received 23 June 2024; revised 26 October 2024 and 27 November 2024; accepted 2 December 2024. Date of publication 5 December 2024; date of current version 24 January 2025. This work was supported in part by Israel Science Foundation under Grant 1449/23 and in part by Pazy Research Foundation. The associate editor coordinating the review of this article and approving it for publication was Dr. Jens Ahrens. *(Corresponding author: Gal Itzhak.)*

The authors are with the Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: galitz@technion.ac.il; icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/TASLP.2024.3512358

("near-end") signal from all speakers, continuous estimation and tracking of the dynamic properties of the scenario are essential. Current methods have addressed this challenge by estimating the instantaneous acoustic transfer function [30], echo paths [31], or steering vector [32], and have proven effective when the number of speakers is small and the scenario dynamics are limited. However, these approaches struggle to accommodate scenarios with multiple or rapidly changing speakers, especially when dispersed over a large spatial area.

Alternative strategies have focused on dynamically localizing the signal of interest or reducing the effects of discrepancies between the actual and estimated DOAs through robust beamforming techniques. Several methods, as referenced in [33], [34], [35], [36], propose an adaptive mechanism. This process involves estimating and tracking the true DOA before applying the beamforming. While effective in static or slowly varying environments, these methods fall short in rapidly changing scenarios involving moving speakers. Additionally, they often rely on substantial prior knowledge or assumptions about the nature of the desired signal. On the other hand, alternative robust approaches circumvent the direct estimation of the actual steering vector. Instead, they aim to maintain a nearly distortionless response even with minor DOA mismatches [37], [38], [39]. However, the efficacy of these methods tends to decline when faced with significant DOA mismatches or in scenarios involving multiple speakers of interest.

Recent research has explored beamforming design with a focus on a continuous region in space, referred to as the region-of-interest (ROI), from which a desired signal may arrive at the array [40], [41], [42], [43]. In [40] (and later on in [41]), an optimization of linear array topology was proposed, aiming to maximize broadband array directivity for a continuous ROI while keeping the WNG above a specified level. This method demonstrated effectiveness for small ROIs but required prior knowledge of the precise DOA of the desired signal, a challenge when dealing with a strong far-end (FE) signal or a moving speaker. Additionally, it lacked direct control over desired array characteristics, such as the mainlobe beamwidth.

The study in [42] utilized an RA topology, optimizing a uniform structure along one axis and a non-uniform structure along the other. This approach offered more flexibility and ensured a constant mainlobe beamwidth from a particular threshold frequency. However, the limited rectangular symmetry and the signal model used meant that it was suited to relatively small ROIs and did not allow direct control over distortion levels. In [43], a new beamforming method based on optimizing the geometry of a sparse concentric circular array was introduced. This approach showed improved performance over previous methods for wider ROIs. However, the choice of CAs and CCAs as the primary geometric structures was not thoroughly justified, and the design method for beamformer taps did not directly consider the desired ROI, indicating potential areas for further significant improvement. Moreover, none of the existing studies enabled a uniform spatial response across a range of azimuthal directions.

This paper introduces a robust approach to beamforming for multispeaker audio conferencing. Our approach focuses on controlling distortion and considers a near-end region (NER) in space from which the speech signals from different conference participants may originate. We independently address the geometrical layout selection and beamformer optimization. Our proposed array geometry consists of a CCA placed on the x-y plane and a ULA placed on the z-axis. This geometry can exhibit a constant spatial response considering all possible directions on the x-y plane. We demonstrate the advantages of employing this geometry and discuss ways to select the array-structure parameters. We propose three high-directivity beamformers and analyze them based on standard performance measures: WNG, directivity factor (DF), and corresponding beampatterns. Finally, we present simulations in various noisy and reverberant scenarios while considering static and moving speech signal sources. We compare our proposed beamformers to known beamformers from the literature and show that our approach is superior when considering desired speech source directions that considerably deviate from the nominal steering direction of the array and when the reverberation level is mild.

The rest of the paper is organized as follows. Section II presents the signal model and the relevant mathematical notation. In Section III, we discuss the configuration of the array geometry. Section IV formulates the proposed distortion-controlled beamforming approach and introduces appropriate performance measures. In Section V, we present three high-directivity beamformers. Section VI includes extensive simulations that analyze the proposed geometry considering its design parameters, evaluate the proposed beamformers regarding the derived performance measures, and compare them with existing beamformers. This section ends with multiple scenarios of speech signal simulations in noisy and reverberant environments while considering both static and moving speech signal sources.

II. Signal Model and Problem Formulation

Consider a signal of interest propagating from the farfield in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, in an elevation angle θ and an azimuth angle ϕ, and impinges on a three-dimensional (3-D) microphone array. By using the polar coordinate system, we can define the steering vector that represents the phase differences between the array microphones and a reference point [5]:

$$\begin{aligned}\mathbf{d}_{\theta,\phi}(f) &= \left[D^{(1)}_{\theta,\phi}(f) \quad \cdots \quad D^{(M)}_{\theta,\phi}(f)\right]^T \\ &= \left[e^{\jmath 2\pi f \Delta^{(\theta,\phi)}_{\text{ref},1}/c} \quad \cdots \quad e^{\jmath 2\pi f \Delta^{(\theta,\phi)}_{\text{ref},M}/c}\right]^T, \end{aligned} \tag{1}$$

where the superscript ^{T} denotes the transpose operator, $\jmath = \sqrt{-1}$ is the imaginary unit, $f > 0$ is the temporal frequency, M denotes the number of array microphones and $\Delta^{(\theta,\phi)}_{\text{ref},i}$ is the relative Euclidean travel distance between the i-th microphone $(i = 1, \ldots, M)$ and the reference point considering a farfield signal originating from (θ, ϕ). Then, the observed signal vector in the frequency domain may be expressed as [8]

$$\begin{aligned}\mathbf{y}(f) &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}_{\theta,\phi}(f)X(f) + \mathbf{v}(f), \end{aligned} \tag{2}$$

where $X(f)$ is the zero-mean desired signal as received by the reference microphone, and $\mathbf{v}(f)$ is the zero-mean additive noise signal vector.

Denoting the desired source incident angle by (θ_0, ϕ_0) and dropping the explicit dependence on f, the covariance matrix of $\mathbf{y}$ is given by

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}} &= E\left(\mathbf{y}\mathbf{y}^H\right) \\ &= p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}_{\theta_0,\phi_0}^H + \mathbf{\Phi}_{\mathbf{v}},\end{aligned} \tag{3}$$

where the superscript ^{H} is the conjugate-transpose operator, $p_X = E(|X|^2)$ is the variance of X, and $\mathbf{\Phi}_{\mathbf{v}} = E(\mathbf{v}\mathbf{v}^H)$ is the covariance matrix of $\mathbf{v}$. Assuming that the variance of the noise is approximately the same at all sensors, we can express (3) as

$$\mathbf{\Phi}_{\mathbf{y}} = p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}_{\theta_0,\phi_0}^H + p_V \mathbf{\Gamma}_{\mathbf{v}}, \tag{4}$$

where p_V is the variance of the noise at a reference microphone (e.g., the microphone on the origin of the coordinate system) and $\mathbf{\Gamma}_{\mathbf{v}} = \mathbf{\Phi}_{\mathbf{v}}/p_V$ is the pseudo-coherence matrix of the noise. From (4), we deduce that the input signal-to-noise ratio (SNR) is

$$\text{iSNR} = \frac{\text{tr}\left(p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}^H\right)}{\text{tr}\left(p_V \mathbf{\Gamma}_{\mathbf{v}}\right)} = \frac{p_X}{p_V}, \tag{5}$$

where $\text{tr}(\cdot)$ denotes the trace of a square matrix.

Traditionally, when a single and static desired source is considered, (θ_0, ϕ_0) is assumed to be known or may be reliably estimated. In such cases, it is common to apply a complex beamformer $\mathbf{h}$ to $\mathbf{y}$ to generate an estimate of the desired speech signal by

$$\begin{aligned}\hat{X} &= \mathbf{h}^H\mathbf{y} \\ &= X\mathbf{h}^H\mathbf{d}_{\theta_0,\phi_0} + \mathbf{h}^H\mathbf{v}.\end{aligned} \tag{6}$$

Then, describing the spatial response of the beamformer by considering all possible spatial angles, the beampattern of $\mathbf{h}$ is given by

$$\mathcal{B}_{\theta,\phi}(\mathbf{h}) = \mathbf{h}^H\mathbf{d}_{\theta,\phi}. \tag{7}$$

Unfortunately, when multiple desired and potentially dynamic speakers are considered, the assumption does not hold, implying that a reliable estimate of the true DOA (θ_0, ϕ_0) is often inapplicable. In particular, this is a common issue in conference meeting scenarios in which multiple desired speakers sit around a table, some of which may physically be moving while speaking or actively presenting. An illustration of such scenarios is depicted in Fig. 1. Typically, in such scenarios, a full-duplex communication device is placed on the conference table, constructed of a microphone array and an adjacent loudspeaker to allow seamless communication with remote participants. That is, when the far-end (FE) loudspeaker signal is silent, the near-end (NE) signal associated with the desired in-room speakers may be reliably localized and tracked over time, implying that (6) may be used directly. However, when the FE signal is present, reliable localization of the NE signal is practically impossible as the former tends to be more powerful than the latter by a few orders of magnitude, and (2) turns into

$$\begin{aligned}\mathbf{y} &= X\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}} + \mathbf{v} \\ &= X\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}} + \mathbf{v}_{\text{NE}} + \mathbf{v}_{\text{FE}},\end{aligned} \tag{8}$$

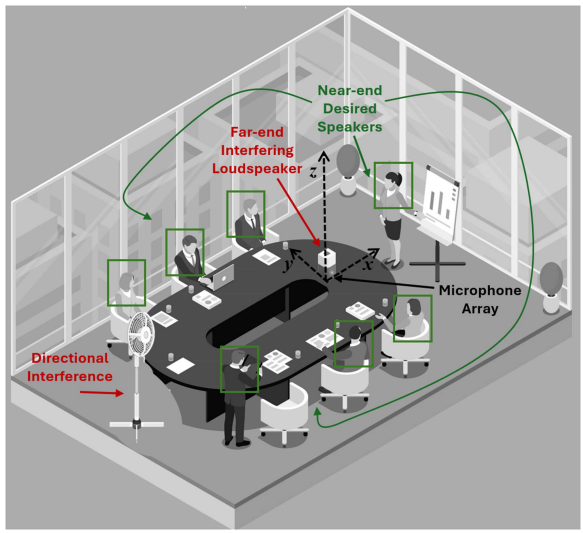


Fig. 1. Illustration of the sort of problems addressed in this paper. Multiple desired speakers sit in a noisy room around a conference table where a loudspeaker-plus-microphone-array full-duplex communication device is placed. The loudspeaker and the microphone array embedded in the communication device are closely spaced.

where $\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}$ indicates the unknown time-varying steering vector associated with the instantaneous desired speaker (assumed for simplicity to be unique per time frame), $\mathbf{v}_{\text{NE}}$ represents the common acoustic noise fields (e.g., white thermal noise and directional interferences), and $\mathbf{v}_{\text{FE}}$ represents the FE loudspeaker signal.

In the remainder of this paper, we address the challenge above and introduce a novel beamforming methodology tailored for conference room environments. This method balances between distortion and quality of sound capture. Our approach includes two key aspects. First, we configure a microphone array geometry that ensures a uniform response across all directions associated with the NE signal. Second, we design complementary beamformers that are characterized by maximum directivity and a controlled level of distortion, taking into account entire regions in space rather than focusing on just a single direction.

III. Array Geometry Configuration

This section describes the microphone array geometry designed for scenarios with multiple dynamic speakers situated across a wide range of directions from the listener. This can occur, for example, in multispeaker conference meetings. In such scenarios, we need the microphone array to have a consistent spatial response over a wide range of azimuth angles, but only a limited range of consistent spatial response in terms of the elevation angle. This is because the desired speakers are not expected to be located at extreme elevation angles relative to the microphone array; all speakers are expected to be on roughly the same plane. Therefore, the beampattern $\mathcal{B}_{\theta,\phi}(\mathbf{h})$ can be written as:

$$\begin{aligned}\mathcal{B}_{\theta,\phi}(\mathbf{h}) &= \mathbf{h}^H\mathbf{d}_{\theta,\phi} \\ &= F(\theta),\end{aligned} \tag{9}$$

where $F(\theta)$ is a function dependent only on θ. As in this section, we focus on the array geometry configuration and leave the beamformer design for Section IV, we may conveniently assume that all elements of $\mathbf{h}$ equal to 1. Note that, albeit specific, this choice is sufficient to demonstrate the advantage of the proposed geometry. In contrast, the ultimate beampattern would depend on the proposed beamformer discussed in the next section. In addition, while it is essentially possible to obtain a constant spatial response using a variety of array geometry and matching beamformer combinations, assigning similar weights to all array elements is likely to result in a fully steerable geometry that exhibits no preference toward specific directions. Thus, (9) becomes

$$\begin{aligned}\mathcal{B}_{\theta,\phi}(\mathbf{1}) &= \sum_{i=1}^{M} D_{\theta,\phi}^{(i)} \\ &= F(\theta), \end{aligned} \tag{10}$$

with $\mathbf{1}$ being a vector of all ones.

The function F may be selected arbitrarily according to any design heuristic. For our case, we let F follow the zero-order Bessel function of the first kind, $J_0(x)$, which is well associated with the azimuth-angle-independence circular symmetry [44], [45]. The idea is to leverage this symmetry-inspired heuristic to obtain a uniform spatial response across all angles. As we require the argument of $J_0(x)$ to be merely a function of θ, (10) turns into

$$\begin{aligned}\mathcal{B}_{\theta,\phi}(\mathbf{1}) &= \sum_{i=1}^{M} D_{\theta,\phi}^{(i)} \\ &\propto J_0[G(\theta)] \\ &= \int_{-\pi}^{\pi} e^{-\jmath G(\theta)\sin(\psi)} \mathrm{d}\psi \\ &= \int_{-\pi}^{\pi} e^{\jmath G(\theta)\cos(\psi-\psi_0+\pi/2)} \mathrm{d}\psi, \end{aligned} \tag{11}$$

where $G(\theta)$ is a function of θ that is independent of ϕ, ψ_0 is a real constant to which selection $\mathcal{B}_{\theta,\phi}(\mathbf{1})$ is invariant, and ψ is the internal integration variable of $J_0(x)$. Note that both measures embody degrees of freedom that may be set arbitrarily. Incorporating the azimuth angle ϕ by setting $\psi_0 = \phi + \pi/2$, and accounting for the frequency and the elevation angle by setting $G(\theta) = 2\pi f r \sin(\theta)/c$ with r and c being arbitrary real parameters, we obtain

$$\begin{aligned}\mathcal{B}_{\theta,\phi}(\mathbf{1}) &= \sum_{i=1}^{M} D_{\theta,\phi}^{(i)} \\ &\propto J_0(2\pi f r \sin(\theta)/c) \\ &= \int_{-\pi}^{\pi} e^{\jmath \frac{2\pi f r}{c}\cos(\phi-\psi)\sin\theta} \mathrm{d}\psi \\ &\approx \sum_{i=1}^{M} e^{\jmath \frac{2\pi f r}{c}\cos(\phi-\psi_i)\sin\theta} \Delta_{i,i+1} \\ &\approx \Delta \sum_{i=1}^{M} e^{\jmath \frac{2\pi f r}{c}\cos(\phi-\psi_i)\sin\theta}, \end{aligned} \tag{12}$$

where $\Delta_{i,i+1}$ is the Euclidean distance between ψ_i and ψ_{i+1} (with the distance between ψ_M and ψ_1 considered when $i = M$), and Δ is the average of all values of $\Delta_{i,i+1}$. We can identify the last row of (12) as the sum of a circular-array steering vector elements of radius r considering its center as the reference point. Note that as required, $\mathcal{B}_{\theta,\phi}(\mathbf{1})$ is independent of ϕ and that the discrete approximation of the integral in the third row is independent of the microphone placements on the circular array circumference provided they are dense enough. Finally, notice that the obtained geometry enables full array steering capabilities if only a uniform spatial response is desired, considering a subset of azimuthal directions. Further analysis of the array geometry configuration is discussed in Section VI.

As (12) remains valid for any value of r, it is clear that it may be generalized by replacing $J_0(2\pi f r \sin(\theta)/c)$ by $\sum_{n=1}^{N} J_0(2\pi f r_n \sin(\theta)/c)$, which corresponds to the element sum of an N-ring CCA steering vector laying in the x-y plane, with r_n being the radius of the n-th ring. Then, denoting the number of microphones on the latter by M_n, we may properly define its corresponding steering vector as [46]

$$\begin{aligned}\mathbf{a}_{n;\theta,\phi}(f) = \Big[& e^{\jmath \frac{2\pi f r_n}{c}\cos(\phi-\psi_{n,1})\sin\theta} \\ & \cdots \quad e^{\jmath \frac{2\pi f r_n}{c}\cos(\phi-\psi_{n,M_n})\sin\theta} \Big]^T, \end{aligned} \tag{13}$$

with $\psi_{n,i}$ being the angle between the i-th microphone on the n-th ring ($i = 1, \ldots, M_n$) and the positive x-axis direction. In the special case of uniformly-spaced microphones on the ring, we have

$$\psi_{n,i} = \frac{2\pi(i-1)}{M_n}. \tag{14}$$

Stacking the steering vectors of all rings, we obtain the full steering vector of the CCA by

$$\mathbf{a}_{\theta,\phi} = \left[\mathbf{a}_{1;\theta,\phi}^T \quad \mathbf{a}_{2;\theta,\phi}^T \quad \cdots \quad \mathbf{a}_{N;\theta,\phi}^T\right]^T. \tag{15}$$

On top of the CCA, let us consider another circularly symmetric ϕ-independent microphone layout: a ULA laying on the z-axis composed of $2P+1$ omnidirectional microphones. This layout is invariant to azimuth-angle rotations and has already been shown effective in controlling the elevation angle [47]. Thus, combining the CCA mentioned above with the proposed ULA would enable a flexible design that considers both angles. Denoting the interelement spacing of the ULA by δ_{z}, its corresponding steering vector is given by [46]

$$\begin{aligned}\mathbf{d}_{\theta;\mathsf{z}} = \Big[& e^{\jmath 2\pi f P \delta_{\mathsf{z}} \cos\theta/c} \quad \cdots \quad 1 \\ & \cdots \quad e^{-\jmath 2\pi f P \delta_{\mathsf{z}} \cos\theta/c} \Big]^T. \end{aligned} \tag{16}$$

With the steering vectors at hand, we can define the joint steering vector by

$$\mathbf{d}_{\theta,\phi} = \left[\mathbf{d}_{\theta;\mathsf{z}}^T \quad \mathbf{a}_{\theta,\phi}^T\right]^T, \tag{17}$$

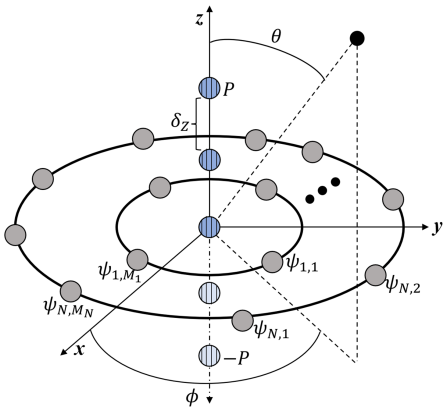


Fig. 2. Illustration of the proposed array geometry. Solid gray circles indicate the CCA microphones placed on the x-y plane, and striped blue-shaded circles indicate the ULA microphones placed on the z-axis on, beneath, and above the x-y plane.

in which the first $2P+1$ elements correspond to the microphones located on the z-axis and the remaining $M=\sum_n M_n$ elements correspond to the CCA microphones located on the x-y plane. The proposed array geometry is illustrated in Fig. 2.

Finally, with our proposed geometry at hand and considering the unknown instantaneous direction $(\theta_{\text{NE}},\phi_{\text{NE}})$, we may rewrite (8) as

$$\begin{aligned}\mathbf{y}&=\begin{bmatrix}\mathbf{y}_{\text{z}}^T & \mathbf{y}_{\text{x-y}}^T\end{bmatrix}^T\\&=X\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}+\mathbf{v}\\&=X\begin{bmatrix}\mathbf{d}_{\theta_{\text{NE}};\text{z}}^T & \mathbf{a}_{\theta_{\text{NE}},\phi_{\text{NE}}}^T\end{bmatrix}^T+\mathbf{v},\end{aligned}\tag{18}$$

where $\mathbf{y}_{\text{z}}$ denotes the observation vector corresponding to the microphones located on the z-axis and $\mathbf{y}_{\text{x-y}}$ is the observation vector corresponding to the CCA located on the x-y plane.

IV. Distortion-Controlled Beamforming

In this section, we describe the beamforming design approach. We aim to develop a method that maintains a constant response over a spatial region covering the DOAs of all desired speakers. This ensures uniform signal capture quality and optimal audio performance in environments with dynamic and varied speaker locations.

Considering (8), we may define a complex beamformer $\mathbf{h}$ of length $2P+M+1$ that aims to estimate the desired signal X by

$$\begin{aligned}\hat{X}&=\mathbf{h}^H\mathbf{y}\\&=X\mathbf{h}^H\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}+\mathbf{h}^H\mathbf{v}.\end{aligned}\tag{19}$$

Clearly, if $\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}$ was static and known, a distortionless response constraint, given by

$$\mathbf{h}^H\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}=1,\tag{20}$$

could directly be exploited to derive $\mathbf{h}$. However, as stated above, $\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}$ is dynamic and unknown, implying that an alternative approach has to be taken instead.

Let us start by considering the performance measures. We begin with the output SNR. From (19) with any arbitrary DOA (θ,ϕ), we have

$$\text{oSNR}\left(\mathbf{h}\right)=\frac{p_X}{p_V}\times\frac{|\mathbf{h}^H\mathbf{d}_{\theta,\phi}|^2}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}},\tag{21}$$

which implies that the SNR gain is given by

$$\mathcal{G}\left(\mathbf{h}\right)=\frac{\text{oSNR}\left(\mathbf{h}\right)}{\text{iSNR}}=\frac{|\mathbf{h}^H\mathbf{d}_{\theta,\phi}|^2}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}}.\tag{22}$$

Consequently, the WNG is given by

$$\mathcal{W}\left(\mathbf{h}\right)=\frac{|\mathbf{h}^H\mathbf{d}_{\theta,\phi}|^2}{\mathbf{h}^H\mathbf{h}},\tag{23}$$

and the DF is

$$\mathcal{D}\left(\mathbf{h}\right)=\frac{|\mathbf{h}^H\mathbf{d}_{\theta,\phi}|^2}{\mathbf{h}^H\mathbf{\Gamma}_{\text{d}}\mathbf{h}},\tag{24}$$

where $\mathbf{\Gamma}_{\text{d}}$ is the pseudo-coherence matrix of the spherically isotropic (diffuse) noise field [5] defined by

$$[\mathbf{\Gamma}_{\text{d}}]_{i_1,i_2}=\text{sinc}(2\pi f\Delta_{i_1,i_2}/c),\tag{25}$$

where $i_1,i_2=1,2,\ldots,(2P+M+1)$ are the microphone indices of the beamformer $\mathbf{h}$, Δ_{i_1,i_2} is the Euclidean distance between the i_1th and i_2th microphones, and $\text{sinc}(x)=\sin(x)/x$.

The DOA is unknown, and therefore, the performance measures in (21)–(24) are unsuitable for the optimal derivation of $\mathbf{h}$. Consequently, we consider the DOA to be a random variable characterized by some probability density function $p_{\mathbf{d}}(\theta,\phi)$ defined over the NER: $\theta\in\Theta_{\text{NER}},\phi\in\Phi_{\text{NER}}$. Then, we may define

$$\begin{aligned}\bar{\mathbf{x}}&=XE[\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}]\\&=X\int_{\theta\in\Theta_{\text{NER}}}\int_{\phi\in\Phi_{\text{NER}}}p_{\mathbf{d}}(\theta,\phi)\mathbf{d}_{\theta,\phi}\sin\theta\text{d}\phi\text{d}\theta,\end{aligned}\tag{26}$$

which constitutes a weighted sum over all possible directions of $(\theta_{\text{NE}},\phi_{\text{NE}})$ considering their corresponding probabilities. Note that this formulation greatly differs from the approaches taken in previous studies in the context of NER (or ROI) [40], [42], [43]. In addition, while $\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}$ is unknown, $p_{\mathbf{d}}(\theta,\phi)$ or more generally Θ_{NER} and Φ_{NER} may be reliably estimated and updated during silent periods of the FE signal as they are typically static given a specific scenario. In future work, we may focus on their implementation and optimization, considering different acoustic environments.

Substituting (26) into (21), we may define the output SNR over the entire NER as

$$\begin{aligned}\text{oSNR}_{\text{NER}}\left(\mathbf{h}\right)&=\frac{1}{p_V}\times\frac{E|\mathbf{h}^H\bar{\mathbf{x}}|^2}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}}\\&=\frac{p_X}{p_V}\times\frac{1}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}}\\&\quad\times\frac{|\mathbf{h}^H\int_{\theta\in\Theta_{\text{NER}}}\int_{\phi\in\Phi_{\text{NER}}}\mathbf{d}_{\theta,\phi}\sin\theta\text{d}\phi\text{d}\theta|^2}{|\int_{\theta\in\Theta_{\text{NER}}}\int_{\phi\in\Phi_{\text{NER}}}\sin\theta\text{d}\phi\text{d}\theta|^2}\\&=\frac{p_X}{p_V}\times\frac{1}{\Omega_{\text{NER}}^2}\times\frac{|\mathbf{h}^H\mathbf{b}_{\text{NER}}|^2}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}},\end{aligned}\tag{27}$$

where $\mathbf{b}_{\text{NER}}=\int_{\theta\in\Theta_{\text{NER}}}\int_{\phi\in\Phi_{\text{NER}}}\mathbf{d}_{\theta,\phi}\sin\theta\text{d}\phi\text{d}\theta$ may be regarded as the NER steering vector, having the integration

bounds set according to the desirable NER, and $\Omega_{\text{NER}} = \int_{\theta\in\Theta_{\text{NER}}}\int_{\phi\in\Phi_{\text{NER}}}\sin\theta \mathrm{d}\phi\mathrm{d}\theta$ is the spatial angle associated with that region. In practice, both can be easily evaluated using standard numerical analysis techniques.

It is evident that if the vector $\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}$ is known, (27) simplifies to the form presented in (21). It is important to note that the assumption of a uniform distribution $p_{\mathbf{d}}(\theta,\phi)$ for $\mathbf{d}_{\theta_{\text{NE}},\phi_{\text{NE}}}$ across the NER is made in the second equality of (27). This assumption is particularly practical in scenarios where obtaining a reliable estimation of $p_{\mathbf{d}}(\theta,\phi)$ is challenging. However, this model is flexible, and alternative distributions can be readily integrated into the framework. Such adaptations would be beneficial if certain directions within the NER were more likely than others, allowing for a more tailored and precise representation of the spatial characteristics in the beamforming design.

In the proposed approach, as the NER may potentially be wide, we do not require a distortionless response. Such a response could result in degradation inflicted by undesirable reverberations and spatial interferences. Instead, we allow a controlled desired signal distortion within the NER. Therefore, we replace the distortionless constraint of (20) with the following distortion-controlled constraint

$$\mathbf{h}^H\mathbf{b}_{\text{NER}} = 1. \tag{28}$$

Consequently, the SNR gain is given by

$$\begin{aligned}\mathcal{G}_{\text{NER}}(\mathbf{h}) &= \frac{\text{oSNR}_{\text{NER}}(\mathbf{h})}{\text{iSNR}}\\ &= \frac{1}{\Omega_{\text{NER}}^2}\times\frac{|\mathbf{h}^H\mathbf{b}_{\text{NER}}|^2}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}}\\ &= \frac{1}{\Omega_{\text{NER}}^2}\times\frac{1}{\mathbf{h}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{h}}.\end{aligned} \tag{29}$$

This formula shows that the SNR gain is inversely proportional to the square of the size of the NER, Ω_{NER}^2. This relationship emphasizes the significance of balancing the NER's extent and the acceptable level of distortion in the desired signal. This balance plays a significant role in optimizing the performance of a beamformer, particularly in environments with spatially diverse acoustic characteristics.

The WNG over the entire NER is obtained by

$$\mathcal{W}_{\text{NER}}(\mathbf{h}) = \frac{1}{\Omega_{\text{NER}}^2}\times\frac{1}{\mathbf{h}^H\mathbf{h}}, \tag{30}$$

and the DF over the entire NER is

$$\mathcal{D}_{\text{NER}}(\mathbf{h}) = \frac{1}{\Omega_{\text{NER}}^2}\times\frac{1}{\mathbf{h}^H\mathbf{\Gamma}_{\text{d}}\mathbf{h}}. \tag{31}$$

It is worth noting that our method shares some similarities with the coherently scattered source scenario examined in the study by Wang et al. [48]. However, our approach and its underlying rationale differ significantly. In particular, it is assumed in [48] that the desired signal is scattered spatially, with a known and stable distribution over time. In contrast, our model assumes that the signal source is not localized and may have time-varying characteristics.

To conclude this section, it should be noted that the framework we have developed here is versatile and can be adapted to any array geometry, as no further assumptions were considered on top of the basic signal model described in (1). However, in the current work, we specifically integrate this formulation with the array geometry outlined in Section III. Using this geometry, we enable full array steering capabilities even when a desired spatial response is only required concerning a subset of azimuthal directions. This comes in handy when there is no a priori information on the array's pointing direction concerning the NER or when the scenario evolves (e.g., when desired speakers are added to or removed from the conference room). Nevertheless, we recognize the potential for future research to explore and apply this methodology to other array configurations, including joint optimization of the array geometry and beamformer taps, requiring prior knowledge of the array's location in the room, favoring certain areas in space over others.

V. Proposed Beamformers

In this section, we combine the array geometry proposed in Section III and the distortion-controlled beamforming suggested in Section IV to derive three high-directivity beamformers.

The first beamformer we propose is obtained upon maximizing the array directivity subject to the distortion-controlled constraint. The optimization problem for maximum directivity may be expressed as

$$\min_{\mathbf{h}}\ \mathbf{h}^H\mathbf{\Gamma}_{\text{d}}\mathbf{h}\ \ \text{s.t.}\ \ \mathbf{h}^H\mathbf{b}_{\text{NER}} = 1, \tag{32}$$

whose solution yields the following NER beamformer

$$\mathbf{h}_{\text{NER}} = \frac{\mathbf{\Gamma}_{\text{d}}^{-1}\mathbf{b}_{\text{NER}}}{\mathbf{b}_{\text{NER}}^H\mathbf{\Gamma}_{\text{d}}^{-1}\mathbf{b}_{\text{NER}}}. \tag{33}$$

For the second proposed beamformer, we require further spatial constraints on the received beampattern. For example, we may impose null constraints which set the beampattern's zeros considering an undesirable source (or sources) impinging on the array from some $\mathbf{d}_{\theta_{\text{null}},\phi_{\text{null}}}$ direction, which must be static and known. In this case, the optimal beamformer is given by solving

$$\min_{\mathbf{h}}\ \mathbf{h}^H\mathbf{\Gamma}_{\text{d}}\mathbf{h}\ \ \text{s.t.}\ \ \mathbf{h}^H\mathbf{C} = \begin{bmatrix}1\\0\end{bmatrix}, \tag{34}$$

where the constraints matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \begin{bmatrix}\mathbf{b}_{\text{NER}} & \mathbf{d}_{\theta_{\text{null}},\phi_{\text{null}}}\end{bmatrix}, \tag{35}$$

and $\mathbf{d}_{\theta_{\text{null}},\phi_{\text{null}}}$ is the steering vector for the null direction $(\theta_{\text{null}},\phi_{\text{null}})$. The solution of (34) yields the second beamformer we propose in the study, which is given by

$$\mathbf{h}_{\text{NER/NS}} = \mathbf{\Gamma}_{\text{d}}^{-1}\mathbf{C}\left[\mathbf{C}^H\mathbf{\Gamma}_{\text{d}}^{-1}\mathbf{C}\right]^{-1}\begin{bmatrix}1\\0\end{bmatrix}, \tag{36}$$

and referred to as the NER-NS beamformer.

For the third beamformer proposed in this work, we utilize the suggested approach to define a continuous far-end region (FER) in which the beampattern may be attenuated by a predefined value ϵ_{FER}. This could be useful when the FE signal is located within a known space, but its exact direction is unknown. To put

it simply, similar to $\mathbf{b}_{\text{NER}}$, we define

$$\mathbf{b}_{\text{FER}} = \int_{\theta\in\Theta_{\text{FER}}}\int_{\phi\in\Phi_{\text{FER}}} \mathbf{d}_{\theta,\phi}\sin\theta \mathrm{d}\phi\mathrm{d}\theta, \tag{37}$$

having the integral computed over the FER: $\theta\in\Theta_{\text{FER}}, \phi\in\Phi_{\text{FER}}$. Then, it is straightforward to adapt $\mathbf{C}$ to

$$\mathbf{C} = \begin{bmatrix}\mathbf{b}_{\text{NER}} & \mathbf{b}_{\text{FER}}\end{bmatrix}, \tag{38}$$

whereas the expression for the third beamformer proposed in this work, namely the NER-FER beamformer $\mathbf{h}_{\text{NER/FER}}$, is given by

$$\mathbf{h}_{\text{NER/FER}} = \mathbf{\Gamma}_{\text{d}}^{-1}\mathbf{C}\left[\mathbf{C}^H\mathbf{\Gamma}_{\text{d}}^{-1}\mathbf{C}\right]^{-1}\begin{bmatrix}1\\ \epsilon_{\text{FER}}\,\Omega_{\text{FER}}\end{bmatrix}, \tag{39}$$

with $\Omega_{\text{FER}} = \int_{\theta\in\Theta_{\text{FER}}}\int_{\phi\in\Phi_{\text{FER}}}\sin\theta\mathrm{d}\phi\mathrm{d}\theta$ being the spatial angle associated with the FER.

VI. Experimental Results

A. Array Geometry Analysis

The microphone array geometry consists of a ULA along the z-axis and a CCA located on the x-y plane. In general, we may optimize the microphone locations on the CCA considering the desirable NER (for example, as suggested in [40] and [42] with linear and rectangular arrays, respectively); here, we focus on a UCCA structure and leave such geometry optimizations for future research.

The number of microphones in the UCCA is given by

$$M = \sum_{n=1}^{N} M_n = NM_1. \tag{40}$$

In the following, we assess the impact of the array-structure design parameters on the spatial response of the proposed beamformer: the number of microphones in a single side of the ULA- P, the number of UCCA rings- N, and the number of evenly-spaced microphones on a single ring- M_1.

Let us define $\mathbf{e}_{i;\theta,\phi}$ as the steering vector corresponding to all cross-ring microphones located on the imaginary line whose azimuth distance to the positive x-axis is ψ_i (note that the dependency on n is dropped due to the UCCA structure):

$$\mathbf{e}_{i;\theta,\phi} = \Big[e^{\jmath\frac{2\pi f r_1}{c}\cos(\phi-\psi_i)\sin\theta} \quad e^{\jmath\frac{2\pi f r_2}{c}\cos(\phi-\psi_i)\sin\theta} \quad \dots \quad e^{\jmath\frac{2\pi f r_N}{c}\cos(\phi-\psi_i)\sin\theta}\Big]^T. \tag{41}$$

Notice that the elements of $\mathbf{e}_{i;\theta,\phi}$ are taken from the CCA's steering vector $\mathbf{a}_{\theta,\phi}$ defined in (15) according to their corresponding microphone locations along the same imaginary radial line. Then, reformulating (7), we have

$$\begin{aligned}\mathcal{B}_{\theta,\phi} &= \mathbf{h}^H\mathbf{d}_{\theta,\phi}\\ &= \begin{bmatrix}\mathbf{h}_{\text{z}}^H & \mathbf{h}_{\text{UCCA}}^H\end{bmatrix}\begin{bmatrix}\mathbf{d}_{\theta;\text{z}}^T & \mathbf{a}_{\theta,\phi}^T\end{bmatrix}^T\\ &= \mathcal{B}_{\theta,\text{z}} + \sum_n \mathcal{B}_{\mathbf{a}_{n;\theta,\phi}}\\ &= \mathcal{B}_{\theta,\text{z}} + \sum_i \mathcal{B}_{\mathbf{e}_{i;\theta,\phi}},\end{aligned} \tag{42}$$

where $\mathbf{h}_{\text{z}}$ and $\mathbf{h}_{\text{UCCA}}$ are the beamformers corresponding to the ULA and the UCCA, respectively, and $\mathbf{a}_{n;\theta,\phi}$ is the steering vector corresponding to the n-th ring defined in (13). In addition, $\mathcal{B}_{\mathbf{a}_{n;\theta,\phi}}$ and $\mathcal{B}_{\mathbf{e}_{i;\theta,\phi}}$ are defined by the inner product of the appropriate elements of $\mathbf{h}$ and their corresponding steering vectors $\mathbf{a}_{n;\theta,\phi}$ and $\mathbf{e}_{i;\theta,\phi}$. For example, the beamformer corresponding to the inner UCCA ring is given by

$$\mathcal{B}_{\mathbf{a}_{1;\theta,\phi}} = \mathbf{h}_{2P+2:2P+1+M_1}^H\mathbf{a}_{1;\theta,\phi}. \tag{43}$$

Consequently, we deduce that the beampattern of the proposed array is a sum of the beampattern of the ULA along the z-axis and the beampatterns of the M_1 ULAs of size N, each pointing to a distinct azimuth angle ψ_i.

Let us assume that the NER is centered around the end-fire direction on the x-y plane with Θ_{NER} being narrow, and Φ_{NER} considered broad. This setup is particularly suitable for a conference room setting where a group of speakers is situated around a table, using a communication device that is strategically positioned to focus on a presenting speaker. Since the NER is oriented towards the broadside relative to the ULA, $\mathbf{h}_{\text{z}}$, it can be effectively conceptualized as a delay-and-sum (DS) beamformer [8]. This DS beamformer is characterized by almost uniform real-valued taps, a frequency-dependent spatial response that exhibits symmetry about the x-y plane, and a constant response concerning the azimuth angle ϕ. Additionally, the parameter P, which is associated with the aperture of $\mathbf{h}_{\text{z}}$, plays a critical role in influencing the directivity of the array. As the value of P increases, it effectively narrows the mainlobe beamwidth about the elevation angle θ, thereby enhancing the directivity of the array.

Next, we consider the parameters M_1 and N, which set the spatial response considering ϕ. It is well known that microphones on a UCA yield an approximately invariant spatial response provided they are dense enough [49]. Therefore, we should set M_1 high enough to satisfy this approximation. Moreover, considering the term $\sum_i \mathcal{B}_{\mathbf{e}_{i;\theta,\phi}}$ from the last row of (42), we may regard each ψ_i-pointing ULA of length N as a small high-directivity beamformer whose directivity in the ψ_i direction (and hence the directivity of $\mathbf{h}$) improves as N increases.

To illustrate, Fig. 3 depicts the spatial responses (or beampatterns) of $\mathbf{h}_{\text{NER}}$ in five distinct sets of the discussed parameters (M_1, N, P) concerning both θ and ϕ, and with a total of 17 array microphones. Additionally, we set $\delta_{\text{z}} = 2\,\text{cm}$ and $r_m = 1\,\text{cm}$, and define the NER by $\Phi_{\text{NER}} = [-90^o, 90^o]$ and $\Theta_{\text{NER}} = 90^o$, implying we are interested in the right half of the x-y plane. It can be observed that a higher value of P leads to a narrower mainlobe beamwidth about θ, along with more significantly attenuated sidelobes. This directly affects the array's directivity. When considering M_1, we note that the beampatterns with respect to ϕ vary within the desirable NER when its value is low, such as when it equals 4. As for N, we observe that the directivity improves when it is higher than 1, meaning that more than a single array ring exists. In this case, the mainlobe is strictly confined to the desirable Φ_{NER}, and the sidelobes are attenuated. This is valid as long as M_1 is high enough to provide appropriate angular resolution.

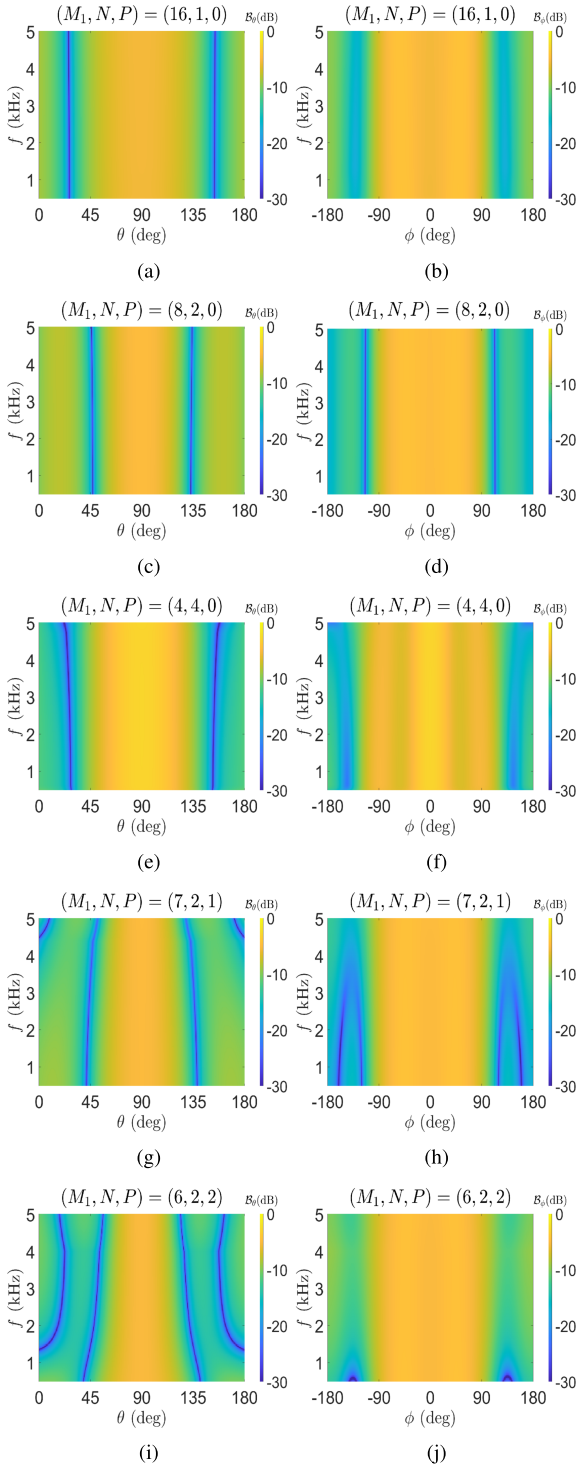


Fig. 3. Beampatterns of $\mathbf{h}_{\text{NER}}$ with five distinct sets of the array-structure design parameters (M_1, N, P). The design parameters are elaborated at the top of each figure. The elevation beampatterns are depicted on the left column, and the azimuth beampatterns are on the right column.

Fig. 4 completes the geometry analysis from a different perspective by addressing the WNG and DF over the entire NER. It is evident that P is the most prominent parameter in $\mathcal{D}_{\text{NER}}$. Moreover, increasing N improves the array directivity only when the value of M_1 is not set too low. Interestingly,

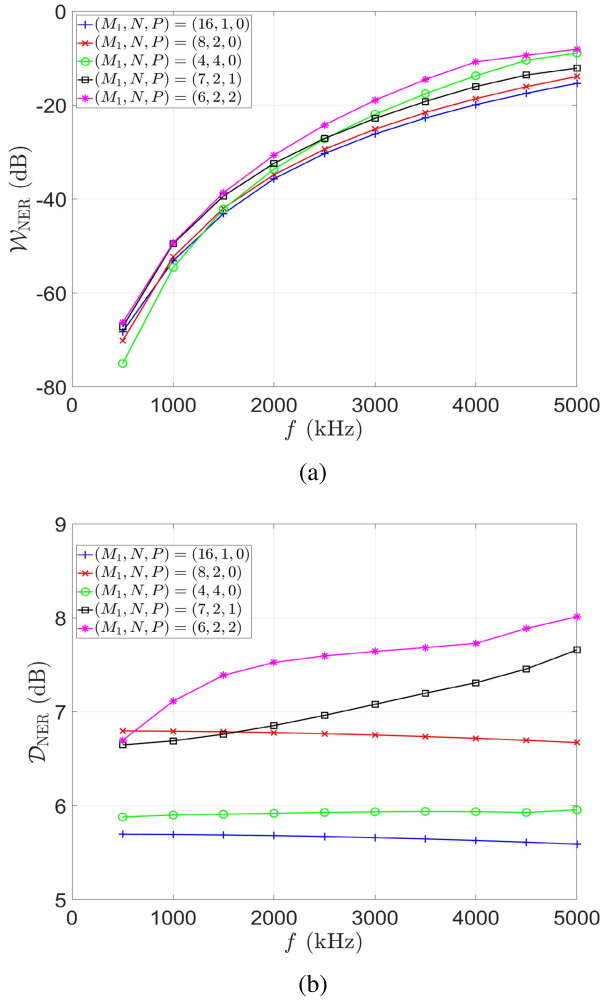


Fig. 4. WNG and DF measures of $\mathbf{h}_{\text{NER}}$ considering the entire NER. (a) $\mathcal{W}_{\text{NER}}$ and (b) $\mathcal{D}_{\text{NER}}$.

this also leads to superior performance in terms of $\mathcal{W}_{\text{NER}}$, as can be seen with $(M_1, N, P) = (6, 2, 2)$. Therefore, in practice, the design process should be done as follows. The parameter M_1 should be set high enough to allow an appropriate angular resolution (around $7 - 8$). Then, given a constraint on the total number of array microphones and the z-axis aperture, P and N are set, controlling the spatial response considering the elevation angle θ the azimuth angle ϕ, respectively.

B. Performance Evaluation

In this section, we conduct an in-depth evaluation of the three beamformers we have developed, focusing on their WNG, DF, and three-dimensional beampatterns performance. Their effectiveness is benchmarked against a range of well-established and recently introduced beamformers from the literature. We utilize an array configuration of 23 microphones for our beamformers, with parameters set as $(M_1, N, P) = (8, 2, 3)$. For $\mathbf{h}_{\text{NER/NS}}$, the null direction is set to $(\theta_{\text{null}}, \phi_{\text{null}}) = (90^o, 125^o)$, and for $\mathbf{h}_{\text{NER/FER}}$, the angular ranges are $\Theta_{\text{FER}} = [67^o, 113^o]$ and $\Phi_{\text{FER}} = [90^o, 160^o]$.

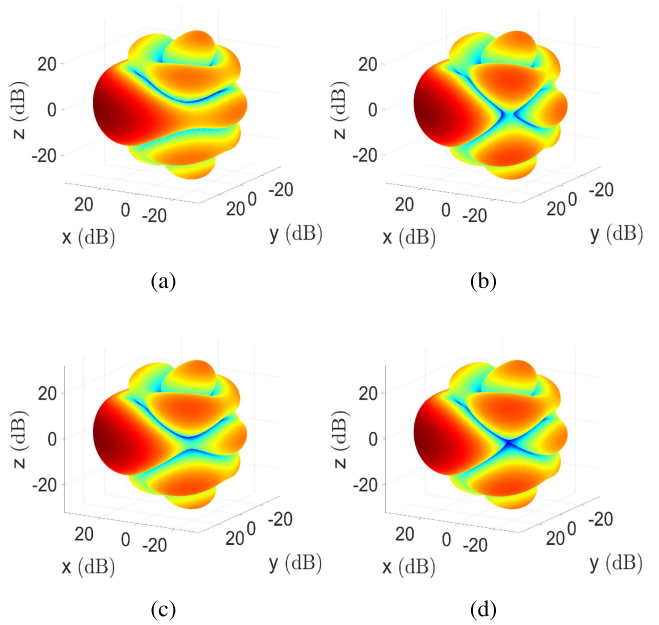


Fig. 5. 3-D beampatterns of the three proposed beamformers. (a) $\mathbf{h}_{\mathrm{NER}}$, (b) $\mathbf{h}_{\mathrm{NER/NS}}$, (c) $\mathbf{h}_{\mathrm{NER/FER}}$ with $\epsilon_{\mathrm{FER}} = -20$ dB, and (d) $\mathbf{h}_{\mathrm{NER/FER}}$ with $\epsilon_{\mathrm{FER}} = -30$ dB.

The 3-D beampatterns are shown in Fig. 5, where $\mathbf{h}_{\mathrm{NER/FER}}$ is particularly showcased with two different values of ϵ_{FER}: -20 dB and -30 dB. It is observed that all beamformers exhibit a similar beampattern considering the desirable NER and any direction distant from the FER. However, we note that while the spatial response of $\mathbf{h}_{\mathrm{NER}}$ is significant for the FER, it is attenuated with the rest: with $\mathbf{h}_{\mathrm{NER/NS}}$ a null is placed in the center of the FER effectively attenuating the adjacent directions in an uncontrolled manner. In contrast, with $\mathbf{h}_{\mathrm{NER/FER}}$, the continuous FER is attenuated according to the value of ϵ_{FER}. Naturally, a lower value of ϵ_{FER} attenuates the response in the FER to a greater extent.

Fig. 6 compares the WNG and DF measures of $\mathbf{h}_{\mathrm{NER/FER}}$ with the parameters described above to two common beamformers from the literature. Note that here we refer to $\mathcal{W}$ and $\mathcal{D}$ from (23) and (24), respectively, which are a function of both the frequency and the desired signal DOA. We simulate a superdirective UCCA (SD-UCCA) beamformer denoted by $\mathbf{h}_{\mathrm{SD;UCCA}}$ and the rectangular ROI-oriented CB beamformer of [42] denoted by $\mathbf{f}_{\mathrm{ROI/CB}}$. The former exhibits 24 microphones composed of 3 uniform rings with 8 uniformly distributed microphones each. In contrast, the latter exhibits 28 microphones consisting of 7 uniformly-distributed microphones along the y-axis and 4 optimally-placed microphones along the x-axis.

We observe that while $\mathbf{h}_{\mathrm{SD;UCCA}}$ and $\mathbf{f}_{\mathrm{ROI/CB}}$ outperform $\mathbf{h}_{\mathrm{NER/FER}}$ in terms of the DF when the desired signal does not deviate much from the endfire direction, their performances dramatically drop outside of the $\phi \in [-20^o, 20^o]$ and $\phi \in [-30^o, 30^o]$ regions, respectively. In contrast, $\mathbf{h}_{\mathrm{NER/FER}}$ exhibits a roughly constant DF performance considering the entire NER and frequency range. Considering the WNG measure, we note that $\mathbf{h}_{\mathrm{NER/FER}}$ is superior as long as $\phi \notin [-20^o, 20^o]$,

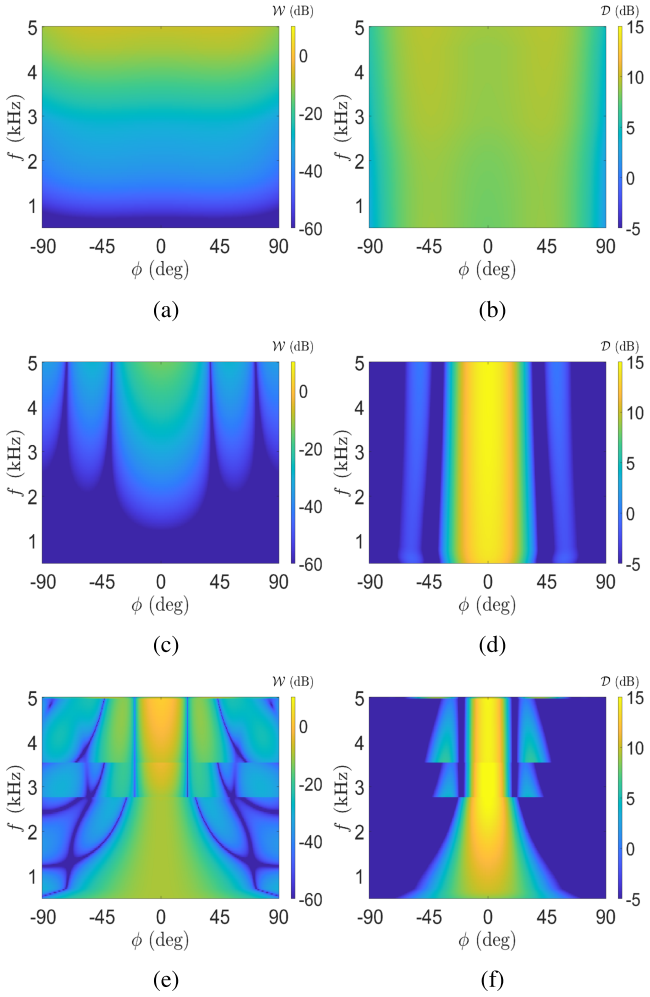


Fig. 6. WNG and DF measures for the proposed $\mathbf{h}_{\mathrm{NER/FER}}$ and two existing beamformers from the literature. (a) WNG for $\mathbf{h}_{\mathrm{NER/FER}}$, (b) DF for $\mathbf{h}_{\mathrm{NER/FER}}$, (c) WNG for $\mathbf{h}_{\mathrm{SD;UCCA}}$, (d) DF for $\mathbf{h}_{\mathrm{SD;UCCA}}$, (e) WNG for $\mathbf{f}_{\mathrm{ROI/CB}}$, and (f) DF for $\mathbf{f}_{\mathrm{ROI/CB}}$.

and is independent of ϕ . The performance gap is particularly noticeable for high frequencies.

C. Speech-Signal Simulations in Reverberant Environments

In this section, we analyze the performance of the proposed beamformers through simulations of noisy speech signals in reverberant environments in four distinct scenarios. The simulations are performed as follows. We use a room impulse response (RIR) generator [50] based on the image method [51] to simulate the reverberant noise-free signal received in the three proposed beamformers (that is, $\mathbf{h}_{\mathrm{NER}}$, $\mathbf{h}_{\mathrm{NER/NS}}$ and $\mathbf{h}_{\mathrm{NER/FER}}$), and the proposed array geometry with the same settings described in the previous section with $(M_1, N, P) = (8, 2, 3)$ and a total of 23 microphones. In addition, we simulate the aforementioned $\mathbf{h}_{\mathrm{SD;UCCA}}$ and $\mathbf{f}_{\mathrm{ROI/CB}}$ beamformers with 24 and 28 microphones, respectively, as well as a DS beamformer $\mathbf{h}_{\mathrm{DS;z}}$ located on the z-axis that consists of 23 microphones with an interelement spacing of 5 mm.

TABLE I
PESQ SCORES OF THE NOISY AND ENHANCED SIGNALS WITH THE THREE PROPOSED BEAMFORMERS, THE SD-UCCA BEAMFORMER, THE DS-ULA BEAMFORMER ALONG THE z-AXIS, AND THE RECTANGULAR ROI-ORIENTED CB BEAMFORMER

	Scen. (a)	Scen. (b)	Scen. (c)	Scen. (d)
Noisy	1.17	1.22	1.47	1.17
$\mathbf{h}_{\mathrm{NER}}$ {23}	**2.02**	**1.93**	**1.86**	**1.88**
$\mathbf{h}_{\mathrm{NER/NS}}$ {23}	**2.00**	**1.91**	**1.82**	**1.85**
$\mathbf{h}_{\mathrm{NER/FER}}$ {23}	**2.01**	**1.91**	**1.83**	**1.86**
$\mathbf{h}_{\mathrm{SD;UCCA}}$ {24}	**2.02**	1.66	1.70	1.42
$\mathbf{h}_{\mathrm{DS;z}}$ {23}	1.35	1.41	1.77	1.27
$\mathbf{f}_{\mathrm{ROI/CB}}$ [42] {28}	1.86	1.43	1.38	1.30

The curly brackets indicate the total number of array microphone.

TABLE II
STOI SCORES OF THE NOISY AND ENHANCED SIGNALS WITH THE THREE PROPOSED BEAMFORMERS, THE SD-UCCA BEAMFORMER, THE DS-ULA BEAMFORMER ALONG THE z-AXIS, AND THE RECTANGULAR ROI-ORIENTED CB BEAMFORMER

	Scen. (a)	Scen. (b)	Scen. (c)	Scen. (d)
Noisy	0.88	0.83	0.88	0.81
$\mathbf{h}_{\mathrm{NER}}$ {23}	0.92	**0.86**	0.85	**0.84**
$\mathbf{h}_{\mathrm{NER/NS}}$ {23}	0.92	**0.86**	0.86	**0.84**
$\mathbf{h}_{\mathrm{NER/FER}}$ {23}	0.92	**0.86**	0.86	**0.84**
$\mathbf{h}_{\mathrm{SD;UCCA}}$ {24}	**0.95**	0.83	0.83	0.72
$\mathbf{h}_{\mathrm{DS;z}}$ {23}	0.91	0.85	**0.90**	0.83
$\mathbf{f}_{\mathrm{ROI/CB}}$ [42] {28}	0.93	**0.86**	0.82	0.83

The curly brackets indicate the total number of array microphone.

Each of the arrays is centered around the location $(x, y, z) = (4, 2, 1)\mathrm{m}$ in an $8 \times 6 \times 3\,\mathrm{m}$ room and pointing towards the endfire direction on the x-y plane. Denoting the true DOA of the desired speaker concerning the azimuth angle by ϕ_0 (and assuming $\theta_0 = 90^o$), we simulate the four following scenarios embodying desired speakers located in distinct respective deviations from the array's pointing direction

- Scen. (a): $\phi_0 = 0^o$;
- Scen. (b): $\phi_0 = 40^o$;
- Scen. (c): $\phi_0 = 80^o$;
- Scen. (d): $\phi_0 = 40^o$, $(\theta_{\mathrm{I}_1}, \phi_{\mathrm{I}_1}) = (45^o, 45^o)$, $(\theta_{\mathrm{I}_2}, \phi_{\mathrm{I}_2}) = (90^o, 143^o)$;

where in Scen. (d), $(\theta_{\mathrm{I}_1}, \phi_{\mathrm{I}_1})$ and $(\theta_{\mathrm{I}_2}, \phi_{\mathrm{I}_2})$ are the DOAs of two directional interferences. In all scenarios, we set $T_{60} = 200$ msec, where T_{60} is defined by Sabin-Franklin's formula [52] and add two simulated noise fields: a white thermal Gaussian noise that naturally exists in the microphones and a spherically-isotropic diffuse noise that models practical noise commonly found in acoustic environments. The latter is set to being 30 dB stronger than the former.

In Scen. (d), the two spatial interferences have equal power, each being 10 dB weaker than the diffuse noise. The input SNR is set to iSNR $= 7$ dB. The desired speech signal, $x(t)$, is a concatenation of 24 speech signals (12 speech signals per gender) with varying dialects that are taken from the TIMIT database [53]. It is sampled at a sampling rate of $f_{\mathrm{s}} = 1/T_{\mathrm{s}} = 16$ kHz within the signal duration T. The speech signal enhancement is performed in the short-time Fourier transform (STFT) domain using 75% overlapping time frames and a Hamming analysis window of length 256 (16 msec), with no additional zero padding. That is, the fast Fourier transform (FFT) applied to each time frame is of length 256.

We analyze and compare the perceptual evaluation of speech quality (PESQ) [54] and short-time objective intelligibility (STOI) [55] scores of the time-domain enhanced signals in each of the four scenarios and with each of the six beamformers above. The results are shown in Tables I and II. We note that in Scen. (a), when the desired speech signal impinges on the arrays from $\phi_0 = 0^o$, all beamformers enhance the observed speech in terms of both scores. Considering the STOI score, all beamformers exhibit a roughly similar performance with $\mathbf{h}_{\mathrm{SD;UCCA}}$ slightly outperforming the others. Considering the PESQ score, $\mathbf{h}_{\mathrm{SD;UCCA}}$ and the three proposed beamformers are shown to be significantly superior to the $\mathbf{h}_{\mathrm{DS;z}}$ and $\mathbf{f}_{\mathrm{ROI/CB}}$. In contrast, when ϕ_0 significantly deviates from the 0^o direction, that is, in Scens. (b) to (d), the proposed beamformers are shown to be highly preferable regarding the PESQ score. This significant performance gap can be associated with the nature of the proposed beamformers and their explicit continuous NER support. Considering the STOI scores in these scenarios, the proposed beamformers outperform the existing beamformers in the challenging Scen. (d), but perform worse than others with no DOA deviation, as in Scen. (a) and from $\mathbf{h}_{\mathrm{DS;z}}$ in the absence of directional interferences as in Scen. (c). Nevertheless, the performance gap between the proposed and existing approaches regarding the STOI scores is small.

D. Speech-Signal Simulations With a Moving Source

In this section, we analyze and compare the performance of the proposed beamformers through speech-signal simulations with a moving desired source in a set of highly-reverberant and noisy scenarios. More specifically, we maintain the same simulation procedure presented in the former section, including the arrays, beamforming approaches, and processing methods. Moreover, to demonstrate the importance of the proposed beamforming approach on top of the proposed array geometry, we simulate a superdirective beamformer applied to the same geometry of the three proposed beamformers and denoted by $\mathbf{h}_{\mathrm{SD}}$. In contrast to the former section, we employ a dynamic speaker that moves inside the room and effectively generates a time-varying DOA from -90^o to 90^o. This is attained by concatenating the reverberant signals associated with the desired source in 31 evenly spaced static scenarios generated using the image method. Additionally, we keep the two interferences presented in the previous Scen. (d) but lower the input SNR to iSNR $= 0$ dB. Considering the reverberation properties of the room, we simulate three scenarios with varying values of T_{60}: 300 msec, 600 msec, and 900 msec. The results, evaluating the average and standard deviation of the PESQ and STOI scores of the time-domain enhanced signals, are described in III and IV, respectively. We observe that, as with the static source cases, the results considering the PESQ scores are significantly preferable with the proposed approach, having both the array geometry and beamforming method applied as suggested in this paper. Specifically, $\mathbf{h}_{\mathrm{NER}}$ is shown to be superior in all three scenarios. Note that $\mathbf{h}_{\mathrm{NER/FER}}$ is preferable to

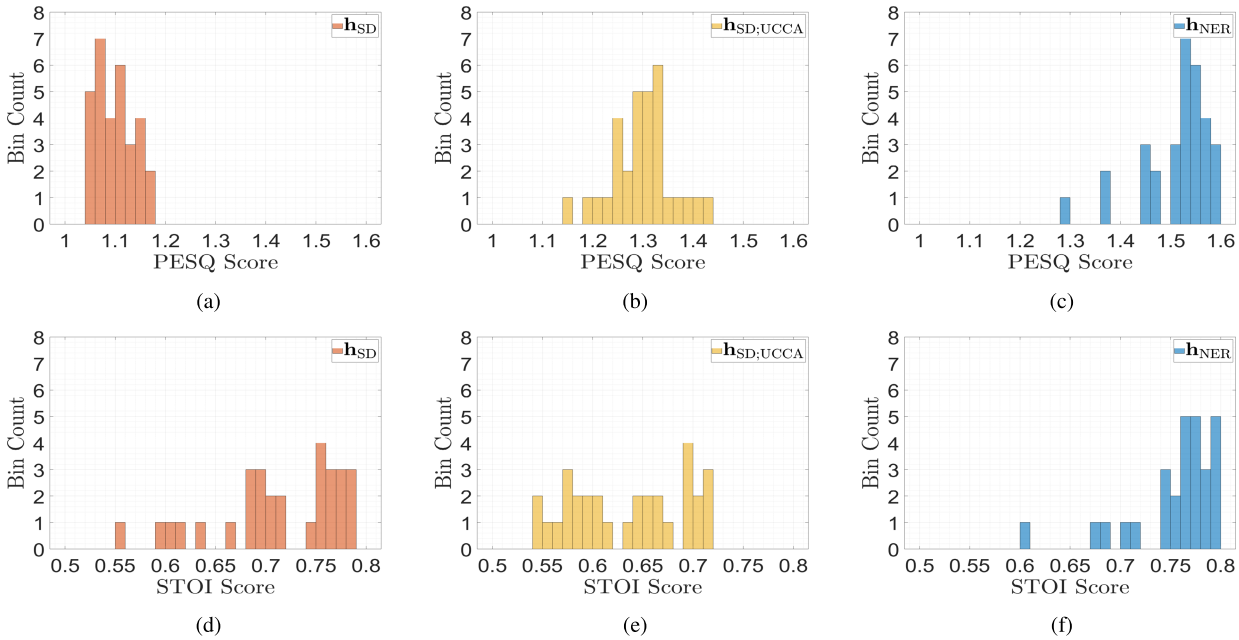


Fig. 7. PESQ and STOI scores histograms of the enhanced signals in the moving-source scenario with the proposed and two existing approaches. (a) PESQ with $\mathbf{h}_{\mathrm{SD}}$, (b) PESQ with $\mathbf{h}_{\mathrm{SD;UCCA}}$, (c) PESQ with $\mathbf{h}_{\mathrm{NER}}$, (d) STOI with $\mathbf{h}_{\mathrm{SD}}$, (e) STOI with $\mathbf{h}_{\mathrm{SD;UCCA}}$, and (f) STOI with $\mathbf{h}_{\mathrm{NER}}$. The value of T_{60} is set to 300 msec.

TABLE III
AVERAGE AND STANDARD DEVIATION OF THE PESQ SCORES OF THE NOISY AND ENHANCED SIGNALS CONSIDERING A MOVING DESIRED SPEECH SOURCE AND VARYING VALUES OF T_{60}

T_{60} =	**300 msec**	**600 msec**	**900 msec**
Noisy	1.04	1.03	1.02
$\mathbf{h}_{\mathrm{NER}}$ {23}	**1.51 ±0.07**	**1.33 ±0.04**	**1.28 ±0.03**
$\mathbf{h}_{\mathrm{NER/NS}}$ {23}	**1.48 ±0.09**	**1.32 ±0.06**	**1.27 ±0.05**
$\mathbf{h}_{\mathrm{NER/FER}}$ {23}	**1.49 ±0.08**	**1.32 ±0.05**	**1.24 ±0.05**
$\mathbf{h}_{\mathrm{SD}}$ {23}	1.10 ±0.04	1.07 ±0.02	1.06 ±0.02
$\mathbf{h}_{\mathrm{SD;UCCA}}$ {24}	1.30 ±0.06	1.26 ±0.05	**1.22 ±0.04**
$\mathbf{h}_{\mathrm{DS;z}}$ {23}	1.05 ±0.02	1.04 ±0.03	1.04 ±0.03
$\mathbf{f}_{\mathrm{ROI/CB}}$ [42] {28}	1.06 ±0.05	1.05 ±0.03	1.05 ±0.02

The curly brackets indicate the total number of array microphone.

TABLE IV
AVERAGE AND STANDARD DEVIATION OF THE STOI SCORES OF THE NOISY AND ENHANCED SIGNALS CONSIDERING A MOVING DESIRED SPEECH SOURCE AND VARYING VALUES OF T_{60}

T_{60} =	**300 msec**	**600 msec**	**900 msec**
Noisy	0.67	0.60	0.57
$\mathbf{h}_{\mathrm{NER}}$ {23}	**0.76 ±0.04**	**0.66 ±0.04**	**0.60 ±0.04**
$\mathbf{h}_{\mathrm{NER/NS}}$ {23}	**0.76 ±0.06**	**0.66 ±0.06**	**0.60 ±0.06**
$\mathbf{h}_{\mathrm{NER/FER}}$ {23}	**0.76 ±0.05**	**0.66 ±0.05**	**0.60 ±0.05**
$\mathbf{h}_{\mathrm{SD}}$ {23}	0.71 ±0.06	**0.62 ±0.06**	**0.58 ±0.06**
$\mathbf{h}_{\mathrm{SD;UCCA}}$ {24}	0.63 ±0.06	0.56 ±0.05	0.54 ±0.04
$\mathbf{h}_{\mathrm{DS;z}}$ {23}	0.70 ±0.01	0.61 ±0.01	**0.57 ±0.02**
$\mathbf{f}_{\mathrm{ROI/CB}}$ [42] {28}	0.65 ±0.09	0.57 ±0.09	0.53 ±0.08

The curly brackets indicate the total number of array microphone.

$\mathbf{h}_{\mathrm{NER/NS}}$ when $T_{60} = 300$ msec and that the higher the value of T_{60} the smaller the performance gap between the proposed and existing approaches. This is an artifact of the NER, which, on the one hand, maintains a controlled desired signal distortion over an entire range of directions, but on the other hand, is susceptible to reverberations impinging on the array from the same directions. Considering the STOI scores, it is evident that the proposed approach outperforms the existing approaches by a great deal. Note that this is in contrast to the static scenarios discussed in the previous section and that $\mathbf{h}_{\mathrm{NER/NS}}$ and $\mathbf{h}_{\mathrm{NER/FER}}$ exhibit minor performance improvements concerning $\mathbf{h}_{\mathrm{NER}}$. Finally, as with the PESQ scores, the higher the value of T_{60}, the smaller the performance gap between the proposed and existing approaches.

To conclude this section, Fig. 7 depicts the histograms of the PESQ and STOI scores associated with $\mathbf{h}_{\mathrm{SD}}$, $\mathbf{h}_{\mathrm{SD;UCCA}}$ and $\mathbf{h}_{\mathrm{NER}}$, respectively, addressing the 31 static scenarios used to simulate the dynamic scenarios described above with $T_{60} = 300$ msec and varying deviations of the desired speaker. Considering the PESQ score histograms, the variance of the occupied bins is comparable with all beamformers; however, the occupied bins with $\mathbf{h}_{\mathrm{NER}}$ are centered around the highest mean value, underlying its superiority. Considering the STOI score histograms, both $\mathbf{h}_{\mathrm{SD}}$ and $\mathbf{h}_{\mathrm{SD;UCCA}}$ exhibit higher variance and lower mean value than $\mathbf{h}_{\mathrm{NER}}$, indicating that the latter is preferable. Nevertheless, the reduced differences between the mean values of these histograms indicate a smaller performance gap in favor of $\mathbf{h}_{\mathrm{NER}}$ compared to the PESQ score performance.

VII. CONCLUSION

We have developed a robust beamforming methodology designed to enhance multispeaker audio conferencing by controlling distortion and utilizing the characteristics of array geometry and beamformer design. Our primary focus was on array geometry, which is essential for ensuring a uniform spatial response

across all feasible directions on the x-y plane. To achieve this, we employed a configuration consisting of a circular concentric array on the x-y plane and a uniform linear array along the z-axis. Additionally, we devised a distortion-controlled approach for beamformer design, enabling us to adjust signal distortion levels across a desired NER in space. This method demonstrated the advantages of our approach and provided practical guidelines for selecting parameters for array structure design. For UCCAs, we found that the number of microphones on a single UCCA ring must be sufficiently large to ensure adequate angular resolution. We identified two critical parameters, P and N, for controlling the spatial response and array directivity concerning the elevation angle θ and the azimuth angle ϕ. Based on these insights, we designed and assessed three high-directivity beamformers, analyzing their WNG, DF, and spatial response performance. We also compared these beamformers with both conventional and recently proposed ones found in the current literature. Finally, we conducted simulations using speech signals in various noisy and reverberant conditions, with both static and moving speech sources. For static sources, the proposed beamformers outperformed the others in terms of PESQ scores, particularly when there were significant deviations of the desired speech source from the array's nominal steering direction. However, all beamformers exhibited similar performance regarding STOI scores. For dynamic sources, the proposed beamformers were more effective, as illustrated by their PESQ and STOI scores, especially in conditions with mild reverberation.

REFERENCES

[1] T. Gerkmann, M. Krawczyk-Becker, and J. L. Roux, "Phase processing for single-channel speech enhancement: History and recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 55–66, Mar. 2015.

[2] M. Parchami, W. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time Fourier transform domain," *IEEE Circuits Syst. Mag.*, vol. 16, no. 3, pp. 45–77, thirdquarter 2016.

[3] Q. Zhang, A. Nicolson, M. Wang, K. K. Paliwal, and C. Wang, "DeepMMSE: A deep learning approach to MMSE-based noise power spectral density estimation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1404–1415, 2020.

[4] G. Itzhak, J. Benesty, and I. Cohen, "Quadratic approach for single-channel noise reduction," *EURASIP J. Audio, Speech, Music Process.*, vol. 2020, pp. 1–14, 2020.

[5] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. New York, NY, USA: Simon and Schuster, 1992.

[6] G. Richard et al., "Audio signal processing in the 21st century: The important outcomes of the past 25 years," *IEEE Signal Process. Mag.*, vol. 40, no. 5, pp. 12–26, Jul. 2023.

[7] A. M. Elbir, K. V. Mishra, S. A. Vorobyov, and R. W. Heath, "Twenty-five years of advances in beamforming: From convex and nonconvex optimization to learning techniques," *IEEE Signal Process. Mag.*, vol. 40, no. 4, pp. 118–131, Jun. 2023.

[8] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. New York, NY, USA: Wiley, 2018.

[9] J. Jin, G. Huang, X. Wang, J. Chen, J. Benesty, and I. Cohen, "Steering study of linear differential microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 158–170, 2021.

[10] G. Itzhak and I. Cohen, "Differential and constant-beamwidth beamforming with uniform rectangular arrays," in *Proc. 17th Int. Workshop Acoustic Signal Enhancement*, Sep. 2022, pp. 1–5.

[11] G. Itzhak, I. Cohen, and J. Benesty, "Robust differential beamforming with rectangular arrays," in *Proc. 29th Eur. Signal Process. Conf.*, 2021, pp. 246–250.

[12] G. Itzhak, J. Benesty, and I. Cohen, "Multistage approach for steerable differential beamforming with rectangular arrays," *Speech Commun.*, vol. 142, pp. 61–76, 2022.

[13] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316–328, Feb. 1996.

[14] P. Heidenreich, A. M. Zoubir, and M. Rubsamen, "Joint 2-D DOA estimation and phase calibration for uniform rectangular arrays," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4683–4693, Sep. 2012.

[15] J. Benesty, J. Chen, and I. Cohen, *Design of Circular Differential Microphone Arrays*. Cham, Switzerland: Springer, 2017.

[16] Y. Buchris, I. Cohen, and J. Benesty, "Frequency-domain design of asymmetric circular differential microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 4, pp. 760–773, Apr. 2018.

[17] G. Huang, J. Chen, and J. Benesty, "On the design of robust steerable frequency-invariant beampatterns with concentric circular microphone arrays," in *Proc. 2018 IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 506–510.

[18] X. Wang, G. Huang, I. Cohen, J. Benesty, and J. Chen, "Robust steerable differential beamformers with null constraints for concentric circular microphone arrays," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 4465–4469.

[19] R. Sharma, I. Cohen, and B. Berdugo, "Controlling elevation and azimuth beamwidths with concentric circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1491–1502, 2021.

[20] A. Kleiman, I. Cohen, and B. Berdugo, "Constant-beamwidth beamforming with concentric ring arrays sensors, special issue on sensors in indoor positioning systems," *Sensors*, vol. 21, no. 21, pp. 7253–7271, Nov. 2021.

[21] A. Kleiman, I. Cohen, and B. Berdugo, "Constant-beamwidth beamforming with nonuniform concentric ring arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 1952–1962, 2022.

[22] G. Huang, J. Benesty, and J. Chen, "Design of robust concentric circular differential microphone arrays," *J. Acoust. Soc. America*, vol. 141, no. 5, pp. 3236–3249, 2017.

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

[24] F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "Efficient implementations of first-order steerable differential microphone arrays with arbitrary planar geometry," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1755–1766, 2020.

[25] F. Borra, A. B. I. Bertuletti, F. Antonacci, and A. Sarti, "Arrays of first-order steerable differential microphones," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 751–755.

[26] A. Bernardini, M. D'Aria, R. Sannino, and A. Sarti, "Efficient continuous beam steering for planar arrays of differential microphones," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 794–798, Jun. 2017.

[27] G. Huang, J. Chen, J. Benesty, I. Cohen, and X. Zhao, "Steerable differential beamformers with planar microphone arrays," *EURASIP J. Audio, Speech, Music Process.*, vol. 2020, 2020, Art. no. 15.

[28] P. E. Son, "On the design of sparse arrays with frequency-invariant beam pattern," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 226–238, 2021.

[29] A. Frank and I. Cohen, "Constant-beamwidth kronecker product beamforming with nonuniform planar arrays," *Front. Signal Process.*, vol. 2, 2022, Art. no. 829463.

[30] J. Park and J. H. Chang, "State-space microphone array nonlinear acoustic echo cancellation using multi-microphone near-end speech covariance," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 10, pp. 1520–1534, 2019.

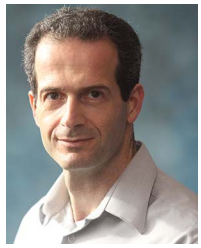
[31] M. M. Halimeh and W. Kellermann, "Frequency-domain mimo acoustic echo cancellation based on a kronecker product approximation," in *Proc. 2022 Int. Workshop Acoustic Signal Enhancement*, 2022, pp. 1–5.

[32] Y. Konforti, I. Cohen, and B. Berdugo, "Multichannel acoustic echo cancellation with beamforming in dynamic environments," *IEEE Open J. Signal Process.*, vol. 4, pp. 479–488, 2023.

[33] K. L. Bell, Y. Ephraim, and H. L. V. Trees, "Robust adaptive beamforming under uncertainty in source direction-of-arrival," in *Proc. 8th Workshop Stat. Signal Array Process.*, pp. 546–549, 1996.

[34] K. L. Bell, Y. Ephraim, and H. L. V. Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.
[35] A. Khabbazibasmenj, S. A. Vorobyov, and A. Hassanien, "Robust adaptive beamforming based on steering vector estimation with as little as possible prior information," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2974–2987, Jun. 2012.
[36] Y. Huang, M. Zhou, and S. A. Vorobyov, "New designs on MVDR robust adaptive beamforming based on optimal steering vector estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3624–3638, Jul. 2019.
[37] J. Li, P. Stoica, and W. Zhisong, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.
[38] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.
[39] C. Y. Chen and P. P. Vaidyanathan, "Quadratically constrained beamforming robust against direction-of-arrival mismatch," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4139–4150, Aug. 2007.
[40] Y. Konforti, I. Cohen, and B. Berdugo, "Array geometry optimization for region-of-interest broadband beamforming," in *Proc. 17th Int. Workshop Acoustic Signal Enhancement*, Sep. 2022, pp. 1–5.
[41] R. Moisseev, G. Itzhak, and I. Cohen, "Array geometry optimization for region-of-interest near-field beamforming," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2024, pp. 576–580.
[42] G. Itzhak and I. Cohen, "Region-of-interest oriented constant-beamwidth beamforming with rectangular arrays," in *Proc. 2023 IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2023, pp. 1–5.
[43] G. Itzhak and I. Cohen, "Kronecker-product beamforming with sparse concentric circular arrays," *IEEE Open J. Signal Process.*, vol. 5, pp. 64–72, 2024.
[44] B. Spain and M. G. Smith, *Functions of Mathematical Physics*. New York, NY, USA: Van Nostrand Reinhold Company, 1970.
[45] F. W. J. Olver, W. L. Daniel, R. F. Boisvert, and C. W. Clark, *The NIST Handbook of Mathematical Functions*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[46] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2004.
[47] G. Itzhak and I. Cohen, "Differential constant-beamwidth beamforming with cube arrays," *Speech Commun.*, vol. 149, pp. 98–107, 2023.
[48] X. Wang, I. Cohen, J. Chen, and J. Benesty, "On robust and high directive beamforming with small-spacing microphone arrays for scattered sources," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 4, pp. 842–852, 2019.
[49] C. T. Molloy, "Calculation of the directivity index for various types of radiators," *J. Acoust. Soc. America*, vol. 20, no. 4, pp. 387–405, 1948.
[50] E. A. P. Habets, "Room impulse response (RIR) generator," 2008. [Online]. Available: https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator
[51] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.
[52] A. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*. Cham, Switzerland: Springer, 2019.
[53] "Darpa timit acoustic phonetic continuous speech corpus CDROM," 1993.
[54] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proc. 2001 IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 2001, vol. 2, pp. 749–752.
[55] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

Gal Itzhak received the B.Sc. degrees in electrical engineering and physics in 2012, and the Ph.D. degree (direct track) in electrical engineering in 2021 from the Technion - Israel Institute of Technology, Haifa, Israel. Between 2012 and 2018, he was an Algorithm Engineer, and a Researcher and Consultant with the Ministry of Defense, primarily focusing on digital communication and highspeed logic design. From 2018 to 2021, he was a Senior Researcher and an Advisor with the private sector, specializing in statistical and machine-learning based models for wireless-device fingerprinting. Since 2021, he has been the Principal Scientist and a Research Manager with cybersecurity domain, focusing on the design of large-scale data models for security threats detection, prevention and prioritization.He has held various research and supervisory positions in academia. His research interests include array signal processing, speech enhancement, machine learning, and cybersecurity.



Israel Cohen (Fellow, IEEE) received the B.Sc. (*summa cum laude*), M.Sc. and Ph.D. degrees in electrical engineering from the Technion – Israel Institute of Technology, Haifa, Israel, in 1990, 1993 and 1998, respectively. From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT, USA. In 2001, he joined the Electrical Engineering Department of the Technion. He is currently the Louis and Samuel Seidan Professor of electrical and computer engineering with the Technion – Israel Institute of Technology. He is the Co-Editor of the Multichannel Speech Processing Section of the *Springer Handbook of Speech Processing* (Springer, 2008), and the co-author of *Fundamentals of Signal Enhancement and Array Signal Processing* (Wiley-IEEE Press, 2018). His research interests include array processing, statistical signal processing, deep learning, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification and adaptive filtering. Dr. Cohen was the recipient of Honorary Doctorate from the Karunya Institute of Technology and Sciences, Coimbatore, India, in 2023, Norman Seiden Prize for Academic Excellence in 2017, SPS Signal Processing Letters Best Paper Award in 2014, Alexander Goldberg Prize for Excellence in Research in 2010, and Muriel and David Jacknow Award for Excellence in Teaching in 2009. He was an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS, a member of the IEEE Audio and Acoustic Signal Processing Technical Committee and the IEEE Speech and Language Processing Technical Committee, and a Distinguished Lecturer of the IEEE Signal Processing Society.