

Least-Distortion Maximum Gain Beamformer for Time-Domain Region-of-Interest Beamforming

Ariel Frank  and Israel Cohen , *Fellow, IEEE*

Abstract—Region-of-interest (ROI) beamformers are very useful in cases where precise information about the source position is unavailable, such as situations involving estimation errors or source movement. This paper presents an approach to ROI beamforming using convolutive filters in the time domain. The proposed beamformer can focus on a specific spatial region of interest while suppressing interference and noise from other directions. We formulate the signal model, considering a desired source signal propagating from a particular ROI and an array of sensors capturing a convolved version of the signal in some noise field. Appropriate performance measures are introduced to derive and analyze ROI-centric beamformers. The ROI beamformer is designed to maximize the gain in signal-to-noise ratio under a constraint on the signal distortion in the spatial region of interest. Additional parameters are introduced to balance the beamformer’s gain, distortion, and robustness. Simulations demonstrate the effectiveness of the proposed method in ROI beamforming and interference suppression.

Index Terms—Broadband beamforming, maximum SNR beamforming, microphone array, minimum variance distortionless response (MVDR) beamformer, region-of-interest beamforming.

I. INTRODUCTION

BEAMFORMING has a wide range of applications in various fields such as communications, sonar, and speech enhancement [1], [2], [3]. The use of microphone arrays has become increasingly popular in recent times, especially in smart homes for audio analysis, conference rooms to include remote participants, and even eyeglasses frames to improve hearing for individuals with hearing difficulties [4], [5], [6]. Capon’s minimum variance distortionless response (MVDR) [7] beamformer is the optimal choice for maximizing the array gain under the distortionless constraint when the direction-of-arrival (DOA) of the source is precisely known [8]. However, in practical scenarios, accurate knowledge of the DOA is often unavailable due to factors such as limited prior information, source mobility, switching sources, and estimation errors due to background noise, interfering sources, and reverberations. A broad framework for handling such uncertainties is region-of-interest (ROI)

beamforming, where the beamformer focuses on a spatial region likely to contain the source. This framework aims to enhance the desired signal while suppressing any interference and noise originating outside the ROI.

A common strategy to handle DOA uncertainty is to design a beamformer with a frequency-invariant beampattern [9], [10], [11], [12], ensuring the same attenuation across all frequencies for a given direction, thereby preventing signal distortion. However, as the signal’s direction deviates further from the center of the mainlobe, the attenuation increases, resulting in a reduction in array gain.

This limitation is addressed in constant beamwidth beamforming [13], [14], [15], where a beamformer is designed to maintain a prescribed half-power beamwidth across all frequencies, ensuring minimal attenuation for signals within the beamwidth. However, the attenuation amount might vary slightly at different frequencies, resulting in some signal distortion. A different approach for compensating for imprecise DOA can be achieved by considering the sources as scattered over a range of directions [16], [17], [18].

In addition to the previously discussed methods, further ROI beamforming techniques have been developed and utilized in various domains such as time, frequency, and time-frequency. Time-domain methods operate directly on the raw waveform signals, offering accurate signal models for linear time-invariant (LTI) systems and effectively capturing transient and broadband signal characteristics. However, they typically involve large covariance matrices, resulting in high computational complexity, especially for large arrays or high sampling rates. In contrast, frequency-domain methods utilize the Fourier transform to decompose signals into multiple frequency bins, significantly reducing matrix dimensions due to separate, smaller covariance matrices at each frequency bin. This decomposition simplifies computational demands and facilitates frequency-specific spatial filtering. Nonetheless, frequency-domain methods using the multiplicative transfer function (MTF) approximation inherently assume that the analysis window length exceeds the duration of the acoustic impulse response [19]. Such assumptions can compromise accuracy, particularly by neglecting crossband interactions [20].

Time-frequency domain methods explicitly leverage the joint time-frequency representation provided by the short-time Fourier transform (STFT). While they also operate on smaller matrices per frequency bin, akin to frequency-domain methods, the explicit time-frequency representation allows for improved adaptability to non-stationary signals. However, even

Received 17 October 2024; revised 21 April 2025; accepted 30 May 2025. Date of publication 6 June 2025; date of current version 18 June 2025. This work was supported in part by Israel Science Foundation under Grant 1449/23 and in part by the Pazy Research Foundation. The associate editor coordinating the review of this article and approving it for publication was Prof. Daniele Salvati. (Corresponding author: Ariel Frank.)

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: arielfrank@campus.technion.ac.il; icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/TASLPRO.2025.3577388

STFT-based models employing the more accurate convolutive transfer function (CTF) approximation [21] still omit crossband terms, potentially limiting precise modeling in scenarios with substantial crossband interactions. Thus, while frequency and STFT-based approaches offer computational efficiency at the expense of model approximations, time-domain methods excel in precise temporal modeling of broadband signals at the cost of higher complexity.

Examples of ROI beamformers in the time domain include [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. ROI beamformers in the frequency domain are discussed in [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42]. In the time-frequency domain, examples are presented in [43], [44], [45]. A straightforward solution to create an ROI beamformer is to use directional microphones that concentrate on the ROI [31]. Another simple approach is to combine the outputs of several beamformers, each pointing to a different direction within the ROI [39], [40]. Hoshuyama et al. [24] demonstrate that by restricting the coefficients' magnitude of the generalized sidelobe canceller (GSC), the desired signal's portion that leaks into the GSC's blocking matrices is not canceled when the DOA is incorrect. Grbić and Nordholm [33] and Davis et al. [43] provide a Wiener solution, while [25], [27], [28], [37] design the beamformer by approximating a desired beampattern using the least squares method. Martinez et al. [38] maximized a weighted average array gain over the ROI, but did not incorporate a distortionless constraint.

Several methods have been proposed to constrain the beampattern response of signals originating in the ROI to limit distortion. Zhang and Thng [26] suggested constraints to slightly widen the beamwidth. Zhang et al. [45] made the beamformer robust to estimation errors by imposing two distortionless constraints – toward the assumed and estimated directions. Takao et al. [22] focused the beampattern on a region using the distortionless constraint for two directions. Lorenz and Boyd [30] constrained the beampattern response to be greater than one over the entire ROI. Itzhak and Cohen [42] constrained the average of the beampattern response over the ROI to equal one. Grenier [23] and Zheng et al. [29] suggested constraining the beamformer's response to unity over the entire ROI and relaxed the constraints using singular value decomposition to circumvent the insufficient degrees of freedom. Setting the rank of the relaxed constraints facilitates a tradeoff between noise reduction and signal distortion. Other time-domain eigenvalue methods include [27], [28], but they did not consider the noise covariance matrix. Frequency-domain generalized eigenvalue methods considering the noise covariance matrix include [35], [36], [41], [45].

Unlike the previous time-domain methods [22], [23], [24], [25], [26], [27], [28], [29], [30], which relied on the acoustic transfer function [23], assumed the near-field propagation model [25], [28], [29], or assumed far-field propagation in an anechoic environment [22], [24], [26], [27], [30], our method leverages relative impulse responses, which are often employed in the frequency domain [41], [44], [45]. Relative impulse responses capture the inter-sensor acoustic differences relative to a reference sensor and typically exhibit significantly shorter durations than full transfer functions. This reduces filter lengths,

computational complexity, and memory requirements. It also simplifies estimation, as fewer parameters need to be calculated or calibrated, thereby improving robustness. Moreover, by avoiding assumptions such as far-field or near-field propagation, relative impulse responses enable more accurate modeling in reverberant environments, making them particularly well suited for time-domain beamforming under realistic acoustic conditions.

Recently, neural network solutions have gained significant attention due to their remarkable performance, including in challenging acoustic environments. Approaches have even explored extracting audio signals from an ROI while suppressing sounds originating outside of the region [46], [47], [48], [49], [50]. These approaches are particularly effective when trained on large, diverse datasets that reflect the intended application environment. Nevertheless, classical beamforming techniques remain advantageous when interpretability, computational efficiency, robustness to unseen scenarios, or limited training data availability is crucial. Therefore, the choice between neural network solutions and classical beamformers largely depends on the specific application context, desired computational constraints, and available data resources. Given these considerations, this paper focuses on classical methods, introducing a robust, optimal time-domain beamformer designed specifically for ROI beamforming. The proposed beamformer maximizes the signal-to-noise ratio (SNR) gain while minimizing the average distortion over the entire ROI.

The main contributions of this paper are as follows. First, we formulate a time-domain ROI beamforming framework that introduces a novel distortion constraint and performance measures tailored specifically for spatial regions. Second, we derive an optimal beamformer by analytically solving a constrained optimization problem via joint diagonalization of the appropriate ROI and noise covariance matrices. This yields a parameter that enables a tradeoff between the array gain and the average distortion. Additional hyperparameters further enhance the beamformer's robustness to spatially white and diffuse noises. Third, we outline a practical approach for computing the beamformer that incorporates the acoustic characteristics of the microphone array, including early reflections and nonideal microphone responses. This approach ensures more accurate and robust ROI beamforming in real acoustic environments. Finally, simulations showcase the beamformer's performance and compare it with robust variants of traditional beamformers and a state-of-the-art ROI beamformer.

The remainder of this paper is organized as follows. Section II presents the signal model. Section III establishes the appropriate performance measures for analyzing ROI beamformers. Section IV introduces a distortion constraint for a specific ROI and presents the beamformer that maximizes the array gain under this constraint. Section V includes simulation results in a reverberant environment. Finally, Section VI concludes.

II. SIGNAL MODEL AND PROBLEM FORMULATION

Consider a desired source signal, $x(t)$, originating from a far-field source in the angular direction (θ_d, ϕ_d) in a spherical coordinate system (see Fig. 1), where $\theta_d \in [0, \pi]$ denotes the

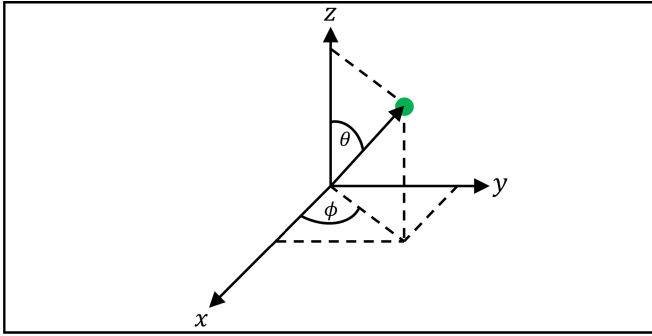


Fig. 1. Spherical coordinate system representation. The far-field source is marked with a green circle at the angular direction (θ, ϕ) . The polar angle θ is measured from the z -axis, and the azimuthal angle ϕ is measured counterclockwise from the positive x -axis in the XY plane.

polar angle (measured as the inclination from the z -axis) and $\phi_d \in [-\pi, \pi]$ the azimuthal angle (measured counterclockwise in the XY plane from the positive x -axis). Consider an array of M sensors with an arbitrary geometry that captures a convolved version of the desired signal in some noise field. At discrete-time index t , the received signals are described as

$$\begin{aligned} y_m(t) &= g_m(t) * x(t) + v_m(t) \\ &= x_m(t) + v_m(t), \quad m = 1, 2, \dots, M, \end{aligned} \quad (1)$$

where $g_m(t)$ represents the impulse response from the unknown source location to the m th sensor, $*$ denotes linear convolution, and $v_m(t)$ represents the additive noise at sensor m . We assume that the signals $x_m(t) = g_m(t) * x(t)$ and $v_m(t)$ are statistically independent, have zero mean, and are stationary, real, and broadband. By definition, the convolved signals $x_m(t)$, $m = 1, 2, \dots, M$, exhibit coherence with each other, while the noise terms $v_m(t)$, $m = 1, 2, \dots, M$, typically possess only partial coherence. Most of the expressions in this paper have a functional dependence on the source's angular direction. For legibility, the dependency of (θ_d, ϕ_d) is not explicitly written.

We can choose an arbitrary sensor, e.g., Sensor 1, as the reference sensor, and write (1) as

$$y_m(t) = d_m(t) * x_1(t) + v_m(t), \quad m = 1, 2, \dots, M, \quad (2)$$

where $d_m(t)$ can be seen as the relative impulse response from the desired source location to the m th sensor, relative to the reference sensor, satisfying

$$d_m(t) * x_1(t) = g_m(t) * x(t), \quad m = 2, \dots, M. \quad (3)$$

For the reference sensor, we have

$$d_1(t) = \delta(t), \quad (4)$$

where $\delta(t)$ is the Kronecker delta. Approximating the relative impulse response as a noncausal finite-length filter of length L_d , we can express (2) as

$$y_m(t) = \sum_{k=-\Delta}^{L_d-1-\Delta} d_m(k) x_1(t-k) + v_m(t), \quad (5)$$

where Δ is the number of noncausal time samples utilized for the approximation in (5), assuming that the samples of $d_m(k)$

outside the range $k \in \{-\Delta, \dots, L_d - 1 - \Delta\}$ are negligible for all $m = 2, \dots, M$. This signal model combines aspects of the models presented in Section 4.2.3 of [51] and in [8]. The former approximates the relative impulse response as a causal finite impulse response (FIR), while the latter uses a noncausal FIR under the assumption of an anechoic environment. In contrast, our model approximates the relative impulse responses as noncausal FIRs without restricting them to anechoic conditions. This model offers a more accurate representation, as the relative impulse response is inherently noncausal, as further discussed in the results (see Fig. 3).

Writing (5) in vector form, we have

$$y_m(t) = \mathbf{d}_m^T(\theta_d, \phi_d) \mathbf{x}'_1(t + \Delta) + v_m(t), \quad (6)$$

where the superscript T is the transpose operator,

$$\begin{aligned} \mathbf{x}'_1(t + \Delta) &= \\ &[x_1(t + \Delta) \quad x_1(t + \Delta - 1) \quad \dots \quad x_1(t + \Delta - L_d + 1)]^T \end{aligned} \quad (7)$$

denotes a vector that contains L_d samples of $x_1(t)$, and

$$\begin{aligned} \mathbf{d}_m(\theta_d, \phi_d) &= \\ &[d_m(-\Delta) \quad d_m(1 - \Delta) \quad \dots \quad d_m(L_d - 1 - \Delta)]^T \end{aligned} \quad (8)$$

denotes an L_d -dimensional vector that represents a FIR approximation of $d_m(t)$. Note that the functional dependence of $d_m(t)$ on the direction of the desired source is emphasized by adding the variables (θ_d, ϕ_d) to \mathbf{d}_m . From (4), the vector $\mathbf{d}_1(\theta_d, \phi_d)$ is a 1-sparse vector whose $(\Delta + 1)$ th component is equal to 1:

$$\mathbf{d}_1(\theta_d, \phi_d) = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0]^T. \quad (9)$$

By examining L_y consecutive time samples of the signal from the m th sensor, we have

$$\begin{aligned} \mathbf{y}_m(t) &= [y_m(t) \quad y_m(t-1) \quad \dots \quad y_m(t-L_y+1)]^T \\ &= \mathbf{D}_m(\theta_d, \phi_d) \bar{\mathbf{x}}_1(t + \Delta) + \mathbf{v}_m(t) \\ &= \mathbf{x}_m(t) + \mathbf{v}_m(t), \end{aligned} \quad (10)$$

where

$$\begin{aligned} \mathbf{D}_m(\theta_d, \phi_d) &= \\ &= \begin{bmatrix} \mathbf{d}_m^T(\theta_d, \phi_d) & 0 & 0 & \dots & 0 \\ 0 & \mathbf{d}_m^T(\theta_d, \phi_d) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{d}_m^T(\theta_d, \phi_d) \end{bmatrix} \end{aligned} \quad (11)$$

is a Toeplitz matrix of size $L_y \times L$, with $L = L_d + L_y - 1$,

$$\begin{aligned} \bar{\mathbf{x}}_1(t + \Delta) &= \\ &= [x_1(t + \Delta) \quad x_1(t + \Delta - 1) \quad \dots \quad x_1(t + \Delta - L + 1)]^T \end{aligned} \quad (12)$$

is a vector that contains L samples of $x_1(t)$, and

$$\begin{aligned} \mathbf{x}_m(t) &= [x_m(t) \quad x_m(t-1) \quad \dots \quad x_m(t-L_y+1)]^T \\ &= \mathbf{D}_m(\theta_d, \phi_d) \bar{\mathbf{x}}_1(t + \Delta), \end{aligned} \quad (13)$$

$$\mathbf{v}_m(t) = [v_m(t) \ v_m(t-1) \ \cdots \ v_m(t-L_y+1)]^T \quad (14)$$

are vectors of length L_y . By concatenating the observations from the M sensors, we obtain the observation signal vector of length ML_y :

$$\begin{aligned} \underline{\mathbf{y}}(t) &= [\mathbf{y}_1^T(t) \ \mathbf{y}_2^T(t) \ \cdots \ \mathbf{y}_M^T(t)]^T \\ &= \underline{\mathbf{D}}(\theta_d, \phi_d) \bar{\mathbf{x}}_1(t + \Delta) + \underline{\mathbf{v}}(t) \\ &= \underline{\mathbf{x}}(t) + \underline{\mathbf{v}}(t), \end{aligned} \quad (15)$$

where

$$\underline{\mathbf{D}}(\theta_d, \phi_d) = \begin{bmatrix} \mathbf{D}_1(\theta_d, \phi_d) \\ \mathbf{D}_2(\theta_d, \phi_d) \\ \vdots \\ \mathbf{D}_M(\theta_d, \phi_d) \end{bmatrix} \quad (16)$$

is a matrix of size $ML_y \times L$ and

$$\begin{aligned} \underline{\mathbf{x}}(t) &= [\mathbf{x}_1^T(t) \ \mathbf{x}_2^T(t) \ \cdots \ \mathbf{x}_M^T(t)]^T \\ &= \underline{\mathbf{D}}(\theta_d, \phi_d) \bar{\mathbf{x}}_1(t + \Delta), \end{aligned} \quad (17)$$

$$\underline{\mathbf{v}}(t) = [\mathbf{v}_1^T(t) \ \mathbf{v}_2^T(t) \ \cdots \ \mathbf{v}_M^T(t)]^T \quad (18)$$

are vectors of length ML_y . From (15), we deduce that the covariance matrix (of size $ML_y \times ML_y$) of $\underline{\mathbf{y}}(t)$ is

$$\begin{aligned} \mathbf{R}_{\underline{\mathbf{y}}} &= E[\underline{\mathbf{y}}(t)\underline{\mathbf{y}}^T(t)] \\ &= \mathbf{R}_{\underline{\mathbf{x}}} + \mathbf{R}_{\underline{\mathbf{v}}} \\ &= \underline{\mathbf{D}}(\theta_d, \phi_d) \mathbf{R}_{\bar{\mathbf{x}}_1} \underline{\mathbf{D}}^T(\theta_d, \phi_d) + \mathbf{R}_{\underline{\mathbf{v}}}, \end{aligned} \quad (19)$$

where $E(\cdot)$ is the expectation operator, $\mathbf{R}_{\underline{\mathbf{x}}}$ and $\mathbf{R}_{\underline{\mathbf{v}}}$ are the covariance matrices (of size $ML_y \times ML_y$) of $\underline{\mathbf{x}}(t)$ and $\underline{\mathbf{v}}(t)$, respectively, and $\mathbf{R}_{\bar{\mathbf{x}}_1}$ is the covariance matrix (of size $L \times L$) of $\bar{\mathbf{x}}_1(t + \Delta)$.

We aim to design an optimal time-domain or broadband beamformer for a given ROI, Ω , with a real-valued spatiotemporal filter. The ROI, Ω , represents a set of angular directions in spherical coordinates and includes the desired angular direction (θ_d, ϕ_d) . By applying the spatiotemporal filter

$$\underline{\mathbf{h}} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \ \cdots \ \mathbf{h}_M^T]^T \quad (20)$$

of length ML_y , where \mathbf{h}_m , $m = 1, 2, \dots, M$, are temporal convolutive filters of length L_y , to the observation signal vector, we obtain the broadband beamformer's output:

$$\begin{aligned} z(t) &= \sum_{m=1}^M \mathbf{h}_m^T \mathbf{y}_m(t) \\ &= \underline{\mathbf{h}}^T \underline{\mathbf{y}}(t) \\ &= x_{\text{fd}}(t) + v_m(t), \end{aligned} \quad (21)$$

where

$$\begin{aligned} x_{\text{fd}}(t) &= \sum_{m=1}^M \mathbf{h}_m^T \mathbf{D}_m(\theta_d, \phi_d) \bar{\mathbf{x}}_1(t + \Delta) \\ &= \underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta_d, \phi_d) \bar{\mathbf{x}}_1(t + \Delta) \end{aligned} \quad (22)$$

is the filtered desired signal and

$$\begin{aligned} v_m(t) &= \sum_{m=1}^M \mathbf{h}_m^T \mathbf{v}_m(t) \\ &= \underline{\mathbf{h}}^T \underline{\mathbf{v}}(t) \end{aligned} \quad (23)$$

is the residual noise. We deduce that the variance of $z(t)$ is

$$\begin{aligned} \sigma_z^2 &= \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{y}}} \underline{\mathbf{h}} \\ &= \sigma_{x_{\text{fd}}}^2 + \sigma_{v_m}^2, \end{aligned} \quad (24)$$

where

$$\begin{aligned} \sigma_{x_{\text{fd}}}^2 &= \underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta_d, \phi_d) \mathbf{R}_{\bar{\mathbf{x}}_1} \underline{\mathbf{D}}^T(\theta_d, \phi_d) \underline{\mathbf{h}} \\ &= \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{x}}} \underline{\mathbf{h}} \end{aligned} \quad (25)$$

is the variance of $x_{\text{fd}}(t)$ and

$$\sigma_{v_m}^2 = \underline{\mathbf{h}}^T \mathbf{R}_{\underline{\mathbf{v}}} \underline{\mathbf{h}} \quad (26)$$

is the variance of $v_m(t)$.

In principle, any element of the vector $\bar{\mathbf{x}}_1(t + \Delta)$ can be considered as the desired signal. Therefore, from (22), we see that for a given DOA (θ_d, ϕ_d) , the distortionless constraint is

$$\underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta_d, \phi_d) = \mathbf{i}_l^T, \quad (27)$$

where \mathbf{i}_l is the l th column of the $L \times L$ identity matrix, \mathbf{I}_L . For a given ROI Ω , we would like to minimize the average distortion over the entire ROI:

$$\begin{aligned} J_{d,\Omega}(\underline{\mathbf{h}}) &= \frac{1}{|\Omega|} \iint_{(\theta,\phi) \in \Omega} [\underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}} - \mathbf{i}_l]^T [\underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}} - \mathbf{i}_l] \sin \theta d\phi d\theta, \end{aligned} \quad (28)$$

where

$$|\Omega| = \iint_{(\theta,\phi) \in \Omega} \sin \theta d\phi d\theta. \quad (29)$$

Itzhak and Cohen [42] considered a spatial probability density function for the source position within the ROI. If such a model is available, it can be incorporated into (28). To streamline the analysis, we assume a uniform spatial distribution over the ROI throughout this paper.

In practical scenarios, the performance and latency of the beamformer are influenced by the choice of the parameter l (where $1 \leq l \leq L$). The beamformer uses $x_1(t + \Delta)$ to estimate $x_1(t + \Delta - l + 1)$ at its output, meaning the beamformer's latency is $l - 1$.

It follows from (15) that the observation signal vector $\underline{\mathbf{y}}(t)$ is correlated with the L samples of $\bar{\mathbf{x}}_1(t + \Delta)$ shown in (12). Generally, for speech signals, the correlation between samples increases with their temporal proximity. Therefore, the middle sample, $x_1(t + \Delta - \lfloor L/2 \rfloor)$, exhibits the strongest correlation with the other samples of $\bar{\mathbf{x}}_1(t + \Delta)$, where $\lfloor \cdot \rfloor$ denotes the floor function. Consequently, this sample can be estimated most accurately. Hence, to optimize the performance of the beamformer and minimize distortion in the output signal, we select $l = \lfloor L/2 \rfloor + 1$. If we choose $L_d = 2\Delta + 1$, then the latency of

the beamformer becomes $l - 1 = \Delta + \lfloor L_y/2 \rfloor$. Decreasing the value of l reduces latency but may also harm performance and cause distortion.

III. PERFORMANCE MEASURES

This section establishes the appropriate performance measures for calculating and studying ROI beamformers with convolutive filters in the time domain. We assume that the desired signal received by the reference sensor, $x_1(t)$, is white to accommodate scenarios where the statistics of the desired signals are unknown. This assumption enables us to deal with desired broadband signals, ensuring that our methods are applicable even if we do not have specific statistical information about the desired signals. Our formulation extends the performance measures proposed in [8], which assume a desired signal arriving from a known DOA.

The input SNR is calculated from the observations at the reference sensor, $\mathbf{y}_1(t) = \mathbf{x}_1(t) + \mathbf{v}_1(t)$. We easily find that

$$\begin{aligned} \text{iSNR} &= \frac{\text{tr}(\mathbf{R}_{\mathbf{x}_1})}{\text{tr}(\mathbf{R}_{\mathbf{v}_1})} \\ &= \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2}, \end{aligned} \quad (30)$$

where $\text{tr}(\cdot)$ denotes the trace operation, $\mathbf{R}_{\mathbf{x}_1}$ and $\mathbf{R}_{\mathbf{v}_1}$ are the covariance matrices (of size $L_y \times L_y$) of $\mathbf{x}_1(t)$ and $\mathbf{v}_1(t)$, respectively, and $\sigma_{x_1}^2$ and $\sigma_{v_1}^2$ are the variances of $x_1(t)$ and $v_1(t)$, respectively.

The output SNR for a given DOA (θ, ϕ) is obtained from (24). It is given by

$$\begin{aligned} \text{oSNR}(\underline{\mathbf{h}}, \theta, \phi) &= \frac{\sigma_{x_{\text{id}}}^2}{\sigma_{v_{\text{m}}}^2} \\ &= \frac{\underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta, \phi) \mathbf{R}_{\bar{\mathbf{x}}_1} \underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \mathbf{R}_{\mathbf{v}} \underline{\mathbf{h}}} \\ &= \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2} \times \frac{\underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}} \underline{\mathbf{h}}}, \end{aligned} \quad (31)$$

where

$$\underline{\Gamma}_{\mathbf{v}} = \frac{\mathbf{R}_{\mathbf{v}}}{\sigma_{v_1}^2} \quad (32)$$

is the pseudo-correlation matrix (of size $ML_y \times ML_y$) of $\mathbf{v}(t)$. The third equality in (31) is valid, assuming $x_1(t)$ is white. We see from (31) that the array gain for (θ, ϕ) is

$$\begin{aligned} \mathcal{G}(\underline{\mathbf{h}}, \theta, \phi) &= \frac{\text{oSNR}(\underline{\mathbf{h}}, \theta, \phi)}{\text{iSNR}} \\ &= \frac{\underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}} \underline{\mathbf{h}}}. \end{aligned} \quad (33)$$

The average output SNR for a given ROI Ω is given by

$$\begin{aligned} \text{oSNR}_{\Omega}(\underline{\mathbf{h}}) &= \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2} \times \frac{\frac{1}{|\Omega|} \iint_{(\theta, \phi) \in \Omega} \underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}} \sin \theta d\phi d\theta}{\underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}} \underline{\mathbf{h}}} \end{aligned}$$

$$= \frac{\sigma_{x_1}^2}{\sigma_{v_1}^2} \times \frac{\underline{\mathbf{h}}^T \underline{\Gamma}_{\underline{\mathbf{D}}, \Omega} \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}} \underline{\mathbf{h}}}, \quad (34)$$

where

$$\underline{\Gamma}_{\underline{\mathbf{D}}, \Omega} = \frac{1}{|\Omega|} \iint_{(\theta, \phi) \in \Omega} \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \sin \theta d\phi d\theta \quad (35)$$

is a matrix of size $ML_y \times ML_y$. Hence, the average array gain for Ω is

$$\begin{aligned} \mathcal{G}_{\Omega}(\underline{\mathbf{h}}) &= \frac{\text{oSNR}_{\Omega}(\underline{\mathbf{h}})}{\text{iSNR}} \\ &= \frac{\underline{\mathbf{h}}^T \underline{\Gamma}_{\underline{\mathbf{D}}, \Omega} \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}} \underline{\mathbf{h}}}. \end{aligned} \quad (36)$$

The white noise gain (WNG) is the array gain for spatiotemporal white noise. Hence, it is obtained by substituting $\underline{\Gamma}_{\mathbf{v}} = \mathbf{I}_{ML_y}$ in (36), where \mathbf{I}_{ML_y} is the $ML_y \times ML_y$ identity matrix, i.e.,

$$\mathcal{W}_{\Omega}(\underline{\mathbf{h}}) = \frac{\underline{\mathbf{h}}^T \underline{\Gamma}_{\underline{\mathbf{D}}, \Omega} \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \underline{\mathbf{h}}}. \quad (37)$$

The directivity factor (DF) is the array gain for a spatially diffuse noise field. The pseudo-correlation matrix (of size $ML_y \times ML_y$) of diffuse noise is given by

$$\underline{\Gamma}_0 = \frac{1}{4\pi} \int_0^{\pi} \int_0^{2\pi} \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \sin \theta d\phi d\theta. \quad (38)$$

Hence, the DF is defined as

$$\mathcal{D}_{\Omega}(\underline{\mathbf{h}}) = \frac{\underline{\mathbf{h}}^T \underline{\Gamma}_{\underline{\mathbf{D}}, \Omega} \underline{\mathbf{h}}}{\underline{\mathbf{h}}^T \underline{\Gamma}_0 \underline{\mathbf{h}}}. \quad (39)$$

The beampattern conveys how a beamformer reacts to a source signal from a specific direction (θ, ϕ) . Each beamformer has a unique directional response pattern, implying that it has different sensitivities to signals from various directions. The broadband power beampattern is defined as

$$|\mathcal{B}(\underline{\mathbf{h}}, \theta, \phi)|^2 = \underline{\mathbf{h}}^T \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \underline{\mathbf{h}}, \quad (40)$$

and quantifies the power response of the beamformer to signals arriving from the specified direction. The DF is related to the broadband power beampattern by

$$\mathcal{D}_{\Omega}(\underline{\mathbf{h}}) = \frac{\frac{1}{|\Omega|} \iint_{(\theta, \phi) \in \Omega} |\mathcal{B}(\underline{\mathbf{h}}, \theta, \phi)|^2 \sin \theta d\phi d\theta}{\frac{1}{4\pi} \int_0^{\pi} \int_0^{2\pi} |\mathcal{B}(\underline{\mathbf{h}}, \theta, \phi)|^2 \sin \theta d\phi d\theta}. \quad (41)$$

The noise reduction factor measures the extent of noise rejection achieved by the beamformer. It is defined as the ratio of the variance of the noise at the reference sensor to the variance of the residual noise:

$$\begin{aligned} \xi_n(\underline{\mathbf{h}}) &= \frac{\sigma_{v_1}^2}{\underline{\mathbf{h}}^T \mathbf{R}_{\mathbf{v}} \underline{\mathbf{h}}} \\ &= \frac{1}{\underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}} \underline{\mathbf{h}}}. \end{aligned} \quad (42)$$

The noise reduction factor must be at least 1 to prevent noise amplification. A higher value indicates more noise rejection.

The process of beamforming can cause a reduction in both the desired signal and the noise. This reduction in the desired signal

can lead to distortion. To measure the reduction in the desired signal, we use a factor called the desired signal reduction factor. It is defined as the ratio of the power of the desired signal at the reference sensor to the power of the filtered desired signal. For a specific DOA (θ, ϕ) , the desired signal reduction factor is expressed as

$$\begin{aligned} \xi_d(\mathbf{h}, \theta, \phi) &= \frac{\sigma_{x_1}^2}{\mathbf{h}^T \underline{\mathbf{D}}(\theta, \phi) \mathbf{R}_{\mathbf{x}_1} \underline{\mathbf{D}}^T(\theta, \phi) \mathbf{h}} \\ &= \frac{1}{\mathbf{h}^T \underline{\mathbf{D}}(\theta, \phi) \underline{\mathbf{D}}^T(\theta, \phi) \mathbf{h}}. \end{aligned} \quad (43)$$

For a given ROI Ω , the average desired signal reduction is given by

$$\xi_{d,\Omega}(\mathbf{h}) = \frac{1}{\mathbf{h}^T \underline{\Gamma}_{\underline{\mathbf{D}},\Omega} \mathbf{h}}. \quad (44)$$

As the value of $\xi_{d,\Omega}(\mathbf{h})$ approaches 1, the desired signal undergoes less distortion.

A fundamental relation is established as

$$\frac{\text{oSNR}_\Omega(\mathbf{h})}{\text{iSNR}} = \frac{\xi_n(\mathbf{h})}{\xi_{d,\Omega}(\mathbf{h})}. \quad (45)$$

This expression highlights the equivalence between gain/loss in SNR, noise reduction, and distortion of the desired signal.

IV. OPTIMAL ROI BEAMFORMER

This section explains how to obtain optimal convolutive filters in the time domain for ROI beamforming. We introduce a distortion constraint for a specific region of interest and illustrate how to maximize the array gain under this constraint.

A. ROI-Distortion Constraint

The ROI-distortion constraint is obtained by taking the gradient of the average distortion over the entire ROI, $J_{d,\Omega}(\mathbf{h})$, with respect to \mathbf{h} and equating the result to zero. We can rewrite $J_{d,\Omega}(\mathbf{h})$ from (28) as

$$J_{d,\Omega}(\mathbf{h}) = \mathbf{h}^T \underline{\Gamma}_{\underline{\mathbf{D}},\Omega} \mathbf{h} - \mathbf{h}^T \underline{\mathbf{D}}_\Omega \mathbf{i}_l - \mathbf{i}_l^T \underline{\mathbf{D}}_\Omega^T \mathbf{h} + 1, \quad (46)$$

where

$$\underline{\mathbf{D}}_\Omega = \frac{1}{|\Omega|} \iint_{(\theta,\phi) \in \Omega} \underline{\mathbf{D}}(\theta, \phi) \sin \theta d\theta d\phi \quad (47)$$

is a matrix of size $ML_y \times L$. Therefore, the distortion constraint for an ROI Ω is given by

$$\underline{\Gamma}_{\underline{\mathbf{D}},\Omega} \mathbf{h} = \underline{\mathbf{D}}_\Omega \mathbf{i}_l. \quad (48)$$

In the traditional approach, a distortionless constraint is applied to a specific DOA (θ_d, ϕ_d) , as given in (27). This constraint ensures that the desired signal is perfectly preserved at the beamformer output when the DOA of the desired source is accurately known. However, deviations from the presumed DOA can lead to significant distortion of the signal at the beamformer output.

Our proposed methodology does not assume a known DOA for the desired source. Instead, we consider the source to be

located within a specified ROI. Under this assumption, the minimal average distortion, derived by substituting (48) into (46), is nonzero. Consequently, we refer to (48) as an ‘‘ROI-distortion constraint’’ rather than a traditional ‘‘distortionless constraint.’’ The key benefit of our approach lies in its robustness: even when the actual direction of the desired source diverges from the anticipated DOA, the resultant distortion at the beamformer output is significantly reduced compared with the conventional method. Considering an ROI makes our approach more adaptable and reliable in scenarios where the precise direction of the desired source is uncertain or variable.

Note that, generally, it is not possible to obtain explicit expressions for $\underline{\Gamma}_{\underline{\mathbf{D}},\Omega}$ as defined in (35), $\underline{\mathbf{D}}_\Omega$ as defined in (47), and $\underline{\Gamma}_0$ as defined in (38). However, we can obtain $\underline{\mathbf{D}}(\theta, \phi)$ using (16), (11), and (8), and then compute $\underline{\Gamma}_{\underline{\mathbf{D}},\Omega}$, $\underline{\mathbf{D}}_\Omega$, and $\underline{\Gamma}_0$ by numerical integration.

In practical applications, it is possible to play a broadband white stationary sound as the source signal $x(t)$ and record the sound signals received in the microphone array. If the source direction (θ, ϕ) is fixed, then according to (17), we can calculate $\underline{\mathbf{D}}(\theta, \phi)$ from the sample cross-covariance matrix of $\mathbf{x}(t)$ and $\mathbf{x}_1^T(t + \Delta)$:

$$\underline{\mathbf{D}}(\theta, \phi) = \frac{1}{\sigma_{x_1}^2 \tau} \sum_{t=1}^{\tau} \mathbf{x}(t) \mathbf{x}_1^T(t + \Delta), \quad (49)$$

where τ denotes the total number of samples employed for the averaging process. In this context, the sample covariance matrix of $\mathbf{x}_1(t + \Delta)$ has been approximated as $\sigma_{x_1}^2 \mathbf{I}_L$.

If we rotate the microphone array relative to the source [52], [53], [54] or rotate the source relative to the array [55], [56], [57], we can record sound in a way that the source direction $[\theta(t), \phi(t)]$ slowly varies over time. Doing so can make the source direction uniformly distributed over the ROI Ω . As a result, the signal $\mathbf{x}(t)$ becomes a function of $[\theta(t), \phi(t)] \in \Omega$. Therefore, when calculating the sample mean, it inherently includes averaging over the ROI, which yields

$$\underline{\mathbf{D}}_\Omega = \frac{\sum_t \mathbf{x}(t, \theta, \phi) \mathbf{x}_1^T(t + \Delta) \sin \theta(t)}{\sigma_{x_1}^2 \sum_t \sin \theta(t)}, \quad (50)$$

$$\underline{\Gamma}_{\underline{\mathbf{D}},\Omega} = \frac{\sum_t \mathbf{x}(t, \theta, \phi) \mathbf{x}^T(t, \theta, \phi) \sin \theta(t)}{\sigma_{x_1}^2 \sum_t \sin \theta(t)}. \quad (51)$$

If the source direction is uniformly distributed over the entire sphere $(\theta, \phi) \in [0, \pi] \times [0, 2\pi]$, then the sample mean yields

$$\underline{\Gamma}_0 = \frac{\sum_t \mathbf{x}(t, \theta, \phi) \mathbf{x}^T(t, \theta, \phi) \sin \theta(t)}{\sigma_{x_1}^2 \sum_t \sin \theta(t)}. \quad (52)$$

The methodology outlined above for computing $\underline{\Gamma}_{\underline{\mathbf{D}},\Omega}$, $\underline{\mathbf{D}}_\Omega$, and $\underline{\Gamma}_0$ is a practical approach that enables to incorporate the acoustic characteristics of the microphone array. This includes considerations of early reflections originating from the device housing the array. Additionally, the method accounts for the actual sensitivities of the microphones and their spectral and directional response profiles. This approach ensures more accurate and realistic designs of the ROI beamformers in various acoustic environments.

B. Robust Least-Distortion Maximum Gain Beamformer

The least-distortion maximum gain (LDMG) beamformer in the time domain is derived by maximizing the average array gain over the ROI, $\mathcal{G}_\Omega(\mathbf{h})$ in (36), subject to the ROI-distortion constraint in (48), i.e.,

$$\max_{\mathbf{h}} \frac{\mathbf{h}^T \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{h}}{\mathbf{h}^T \mathbf{\Gamma}_{\mathbf{v}} \mathbf{h}} \quad \text{subject to} \quad \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{h} = \mathbf{D}_\Omega \mathbf{i}_l. \quad (53)$$

To make the LDMG robust and enable a compromise between noise reduction, DF, and WNG, we replace $\mathbf{\Gamma}_{\mathbf{v}}$ with

$$\mathbf{\Gamma}_{\mathbf{v},\epsilon} = (1 - \epsilon_0 - \epsilon_1) \mathbf{\Gamma}_{\mathbf{v}} + \epsilon_0 \mathbf{\Gamma}_0 + \epsilon_1 \mathbf{I}_{ML_y}, \quad (54)$$

where the parameters ϵ_0 ($0 \leq \epsilon_0 \leq 1$) and ϵ_1 ($0 \leq \epsilon_1 \leq 1 - \epsilon_0$) control the tradeoff between noise reduction at the beamformer output, DF, and WNG. The matrix $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ of size $ML_y \times ML_y$ is a superposition of the pseudo-correlation matrices of the noise signal $\mathbf{v}(t)$, a spatially diffuse noise field, and a spatially white noise field. In many practical situations, accurately estimating $\mathbf{\Gamma}_{\mathbf{v}}$ can be challenging, which further justifies the need to artificially incorporate diffuse and white noise.

Hence, the optimization problem of the robust LDMG can be formulated as

$$\max_{\mathbf{h}} \frac{\mathbf{h}^T \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{h}}{\mathbf{h}^T \mathbf{\Gamma}_{\mathbf{v},\epsilon} \mathbf{h}} \quad \text{subject to} \quad \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{h} = \mathbf{D}_\Omega \mathbf{i}_l, \quad (55)$$

which is equivalent to

$$\min_{\mathbf{h}} \left[\frac{1 - \epsilon_0 - \epsilon_1}{\mathcal{G}_\Omega(\mathbf{h})} + \frac{\epsilon_0}{\mathcal{D}_\Omega(\mathbf{h})} + \frac{\epsilon_1}{\mathcal{W}_\Omega(\mathbf{h})} \right] \quad \text{s.t.} \quad \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{h} = \mathbf{D}_\Omega \mathbf{i}_l, \quad (56)$$

by substituting (36), (37), (39), and (54), and using the fact that $\arg\max f(\mathbf{h}) = \arg\min(1/f(\mathbf{h}))$ for any strictly positive function $f(\mathbf{h})$, which holds since the matrices $\mathbf{\Gamma}_{\mathbf{D},\Omega}$ and $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ are positive semi-definite. Increasing ϵ_0 (or ϵ_1) results in a larger DF (respectively, WNG) at the cost of a lower array gain (less noise reduction). To ensure that $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ has full rank, ϵ_1 can be adjusted.

To solve this problem, we follow the methodology in [58], which derives an eigen-decomposition-based beamformer. While we adopt the general approach, the objective function and the constraint are tailored specifically to our ROI formulation. Let us assume that $\text{rank}(\mathbf{\Gamma}_{\mathbf{D},\Omega}) = P \leq ML_y$. Since $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ has full rank, the two symmetric matrices $\mathbf{\Gamma}_{\mathbf{D},\Omega}$ and $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ can be jointly diagonalized via the generalized eigenvalue decomposition as follows [59]:

$$\mathbf{T}^T \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{T} = \mathbf{\Lambda}, \quad (57)$$

$$\mathbf{T}^T \mathbf{\Gamma}_{\mathbf{v},\epsilon} \mathbf{T} = \mathbf{I}_{ML_y}, \quad (58)$$

where

$$\mathbf{T} = [\mathbf{t}_1 \quad \mathbf{t}_2 \quad \cdots \quad \mathbf{t}_{ML_y}] \quad (59)$$

is a full-rank matrix (of size $ML_y \times ML_y$) where the columns are the generalized eigenvectors and

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{ML_y}) \quad (60)$$

is a diagonal matrix (of size $ML_y \times ML_y$) with real and non-negative generalized eigenvalues. Furthermore, $\mathbf{\Lambda}$ and \mathbf{T} are the eigenvalue and eigenvector matrices, respectively, of $\mathbf{\Gamma}_{\mathbf{v},\epsilon}^{-1} \mathbf{\Gamma}_{\mathbf{D},\Omega}$, i.e.,

$$\mathbf{\Gamma}_{\mathbf{v},\epsilon}^{-1} \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{T} = \mathbf{T} \mathbf{\Lambda}. \quad (61)$$

The eigenvalues of $\mathbf{\Gamma}_{\mathbf{v},\epsilon}^{-1} \mathbf{\Gamma}_{\mathbf{D},\Omega}$ can be ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P > \lambda_{P+1} = \cdots = \lambda_{ML_y} = 0$. We also denote by $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{ML_y}$, the corresponding eigenvectors. It is convenient to break the matrix \mathbf{T} into two parts, corresponding to the nonzero and the zero eigenvalues. Let

$$\mathbf{T} = [\mathbf{T}_1 \quad \mathbf{T}_2], \quad (62)$$

where the $ML_y \times P$ matrix \mathbf{T}_1 contains the eigenvectors corresponding to the nonzero eigenvalues of $\mathbf{\Gamma}_{\mathbf{v},\epsilon}^{-1} \mathbf{\Gamma}_{\mathbf{D},\Omega}$, and the $ML_y \times (ML_y - P)$ matrix \mathbf{T}_2 contains the eigenvectors corresponding to the null eigenvalues of $\mathbf{\Gamma}_{\mathbf{v},\epsilon}^{-1} \mathbf{\Gamma}_{\mathbf{D},\Omega}$.

We can express \mathbf{h} as

$$\begin{aligned} \mathbf{h} &= \mathbf{T} \mathbf{a} \\ &= \mathbf{T}_1 \mathbf{a}_1 + \mathbf{T}_2 \mathbf{a}_2, \end{aligned} \quad (63)$$

where $\mathbf{a} = [\mathbf{a}_1^T \quad \mathbf{a}_2^T]^T$ is the transformed beamformer of length ML_y , \mathbf{a}_1 contains the first P elements of \mathbf{a} , and \mathbf{a}_2 contains the remaining $ML_y - P$ elements of \mathbf{a} . Instead of optimizing \mathbf{h} directly, we can, equivalently, optimize \mathbf{a} since \mathbf{T} is an invertible matrix ($\mathbf{T}^{-1} = \mathbf{T}^T \mathbf{\Gamma}_{\mathbf{v},\epsilon}$). So when \mathbf{a} is obtained, we can easily find \mathbf{h} from (63).

Using (57), (58), and (63), we can write the optimization problem of the robust LDMG beamformer given in (55) as

$$\max_{\mathbf{a}} \frac{\mathbf{a}^T \mathbf{\Lambda} \mathbf{a}}{\mathbf{a}^T \mathbf{a}} \quad \text{subject to} \quad \mathbf{\Lambda} \mathbf{a} = \mathbf{T}^T \mathbf{D}_\Omega \mathbf{i}_l, \quad (64)$$

or equivalently as

$$\max_{\mathbf{a}} \frac{\mathbf{a}_1^T \mathbf{\Lambda}_1 \mathbf{a}_1}{\mathbf{a}_1^T \mathbf{a}_1 + \mathbf{a}_2^T \mathbf{a}_2} \quad \text{subject to} \quad \mathbf{\Lambda}_1 \mathbf{a}_1 = \mathbf{T}_1^T \mathbf{D}_\Omega \mathbf{i}_l, \quad (65)$$

where

$$\mathbf{\Lambda}_1 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_P) \quad (66)$$

is a diagonal matrix of size $P \times P$. Since $\mathbf{\Lambda}_1$ is invertible, the constraint in (65) imposes

$$\mathbf{a}_1 = \mathbf{\Lambda}_1^{-1} \mathbf{T}_1^T \mathbf{D}_\Omega \mathbf{i}_l, \quad (67)$$

and the expression being maximized in (65) under this constraint is obtained for $\mathbf{a}_2 = \mathbf{0}$. Hence, the robust LDMG beamformer is

$$\begin{aligned} \mathbf{h}_{\text{rLDMG},\epsilon} &= \mathbf{T}_1 \mathbf{\Lambda}_1^{-1} \mathbf{T}_1^T \mathbf{D}_\Omega \mathbf{i}_l \\ &= \left[\sum_{p=1}^P \frac{\mathbf{t}_p \mathbf{t}_p^T}{\lambda_p} \right] \mathbf{D}_\Omega \mathbf{i}_l. \end{aligned} \quad (68)$$

From (68) and the relation $\mathbf{T}_1^T \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{T}_1 = \mathbf{\Lambda}_1$ (cf. eq (57)), we deduce that the corresponding array gain, DF, and WNG over

the ROI are

$$\mathcal{G}_\Omega(\mathbf{h}_{\text{FLDMG},\epsilon}) = \frac{\mathbf{i}_l^T \mathbf{D}_\Omega^T \mathbf{T}_1 \mathbf{A}_1^{-1} \mathbf{T}_1^T \mathbf{D}_\Omega \mathbf{i}_l}{\mathbf{h}_{\text{FLDMG},\epsilon}^T \mathbf{\Gamma}_{\mathbf{v}} \mathbf{h}_{\text{FLDMG},\epsilon}}, \quad (69)$$

$$\mathcal{D}_\Omega(\mathbf{h}_{\text{FLDMG},\epsilon}) = \frac{\mathbf{i}_l^T \mathbf{D}_\Omega^T \mathbf{T}_1 \mathbf{A}_1^{-1} \mathbf{T}_1^T \mathbf{D}_\Omega \mathbf{i}_l}{\mathbf{h}_{\text{FLDMG},\epsilon}^T \mathbf{\Gamma}_0 \mathbf{h}_{\text{FLDMG},\epsilon}}, \quad (70)$$

$$\mathcal{W}_\Omega(\mathbf{h}_{\text{FLDMG},\epsilon}) = \frac{\mathbf{i}_l^T \mathbf{D}_\Omega^T \mathbf{T}_1 \mathbf{A}_1^{-1} \mathbf{T}_1^T \mathbf{D}_\Omega \mathbf{i}_l}{\mathbf{h}_{\text{FLDMG},\epsilon}^T \mathbf{h}_{\text{FLDMG},\epsilon}}. \quad (71)$$

A practical method for selecting ϵ_0 and ϵ_1 without relying on exhaustive search techniques to attain a desired average WNG and DF is outlined below. First, set $\epsilon_0 = \epsilon_1 = 0$. Gradually increase ϵ_1 until the attained WNG exceeds the desired threshold. Then, increase ϵ_0 until the DF surpasses its threshold. This step will likely reduce the WNG, so ϵ_1 should be further increased to restore the WNG to its target level. Since the DF is generally less sensitive to changes in ϵ_0 and ϵ_1 than the WNG, this iterative process can now be stopped with only a minor reduction in DF.

To further enhance the robustness of the LDMG beamformer $\mathbf{h}_{\text{FLDMG},\epsilon}$, and to achieve a better balance between noise reduction, DF, and WNG, we modify the expression in (68). This modification is done by limiting the summation and adjusting the eigenvalues using the diagonal loading technique [60]. Specifically, each eigenvalue is increased by a positive constant $\mu\lambda_1$:

$$\mathbf{h}_{\text{FLDMG},K,\epsilon,\mu} = \left[\sum_{p=1}^K \frac{\mathbf{t}_p \mathbf{t}_p^T}{\lambda_p + \mu\lambda_1} \right] \mathbf{D}_\Omega \mathbf{i}_l, \quad (72)$$

where $1 \leq K \leq P$, and $\mu \geq 0$. The parameter μ is chosen based on the noise level, the required robustness level, or through empirical methods. The constant $\mu\lambda_1$ acts as a regularization term, making the beamforming solution more robust to situations like limited data samples. The goal is to ensure the stability and performance of the beamformer, even when the estimated covariance matrix may be ill-conditioned or have estimation errors.

It is well known that the rank K of the approximated signal covariance matrix provides a tradeoff between the array gain and the signal distortion [58], [61], and our formulation extends this principle to the case of region-of-interest beamforming. The noise reduction achieved by using the beamformer $\mathbf{h}_{\text{FLDMG},K,\epsilon,\mu}$ increases as the parameter K decreases, but this comes with the tradeoff of increasing the average distortion over the ROI Ω , i.e., $J_{d,\Omega}(\mathbf{h}_{\text{FLDMG},K,\epsilon,\mu})$. Increasing the value of μ also leads to a higher noise reduction but also results in a higher average distortion over Ω . On the other hand, increasing the value of ϵ_0 or ϵ_1 results in a higher DF or higher WNG, respectively, but at the cost of lower noise reduction.

Applying a spatiotemporal filter to the measurements according to (21) requires ML_y multiplications. The spatiotemporal filter's weights can be precomputed. For computing the weights for the robust LDMG beamformer (72), the two most computationally demanding steps include computing the matrix $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ and jointly diagonalizing the matrices $\mathbf{\Gamma}_{\mathbf{D},\Omega}$ and $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$. The calculation of $\mathbf{\Gamma}_{\mathbf{v},\epsilon}$ is dominated by the computation of $\mathbf{\Gamma}_0$. Specifically, evaluating the integrand $\mathbf{D}(\theta, \phi) \mathbf{D}^T(\theta, \phi)$ requires $L(ML_y)^2$ multiplications per angular direction. When

this integral is approximated by summation over $|\Omega_d|$ discrete angular directions, the resulting complexity for computing $\mathbf{\Gamma}_0$ is $\mathcal{O}[|\Omega_d|L(ML_y)^2]$. The subsequent joint diagonalization of matrices with dimensions $ML_y \times ML_y$ incurs an additional complexity of $\mathcal{O}[(ML_y)^3]$ [62].

C. Unconstrained Maximum Gain Least-Distortion Beamformer

The maximum average array gain over the ROI that can be obtained without imposing the ROI-distortion constraint is given by

$$\max_{\mathbf{h}} \frac{\mathbf{h}^T \mathbf{\Gamma}_{\mathbf{D},\Omega} \mathbf{h}}{\mathbf{h}^T \mathbf{\Gamma}_{\mathbf{v},\epsilon} \mathbf{h}}, \quad (73)$$

for $\epsilon_0 = \epsilon_1 = 0$. Hence, the unconstrained maximum gain beamformer is

$$\mathbf{h}_{\text{max}} = \varsigma \mathbf{t}_1, \quad (74)$$

where $\varsigma \neq 0$ is an arbitrary real number, and the maximum average array gain is

$$\mathcal{G}_\Omega(\mathbf{h}_{\text{max}}) = \lambda_1. \quad (75)$$

Substituting (74) into (46), taking the derivative of $J_{d,\Omega}(\mathbf{h}_{\text{max}})$ with respect to ς and equating the result to zero yields the unconstrained maximum gain beamformer that minimizes the average distortion over Ω :

$$\mathbf{h}_{\text{max},1} = \frac{\mathbf{t}_1 \mathbf{t}_1^T \mathbf{D}_\Omega \mathbf{i}_l}{\lambda_1}. \quad (76)$$

We see that $\mathbf{h}_{\text{FLDMG},P,\epsilon,0} = \mathbf{h}_{\text{FLDMG},\epsilon}$, and for $K=1$ and $\epsilon_0 = \epsilon_1 = \mu = 0$, we obtain the unconstrained maximum gain beamformer that minimizes the average distortion over Ω :

$$\mathbf{h}_{\text{FLDMG},1,0,0} = \mathbf{h}_{\text{max},1}. \quad (77)$$

D. Contrast With the MVDR Beamformer

It is interesting to compare the LDMG beamformer with the traditional MVDR beamformer. The LDMG beamformer is optimized explicitly for a designated ROI Ω , with its design focusing on maximizing the average array gain across the ROI, as denoted by $\mathcal{G}_\Omega(\mathbf{h})$ in (36). This optimization is subject to the ROI-distortion constraint in (48). On the other hand, the MVDR beamformer targets minimizing the variance of the residual noise at the output, $\sigma_{v_m}^2$ in (26), while adhering to a distortionless constraint, which ensures that the desired signal remains undistorted, as stated in (27).

When the ROI Ω is reduced to a single DOA (θ_d, ϕ_d) , the optimization problem in (53) simplifies to

$$\min_{\mathbf{h}} \mathbf{h}^T \mathbf{\Gamma}_{\mathbf{v}} \mathbf{h} \quad \text{subject to} \quad \mathbf{D}^T(\theta_d, \phi_d) \mathbf{h} = \mathbf{i}_l, \quad (78)$$

and the LDMG beamformer reduces to the MVDR beamformer:

$$\begin{aligned} & \mathbf{h}_{\text{MVDR}}(\theta_d, \phi_d) \\ &= \mathbf{\Gamma}_{\mathbf{v}}^{-1} \mathbf{D}(\theta_d, \phi_d) \left[\mathbf{D}^T(\theta_d, \phi_d) \mathbf{\Gamma}_{\mathbf{v}}^{-1} \mathbf{D}(\theta_d, \phi_d) \right]^{-1} \mathbf{i}_l. \end{aligned} \quad (79)$$

By replacing $\Gamma_{\mathbf{v}}$ by $\Gamma_{\mathbf{v},\epsilon}$ from (54), the robust MVDR beamformer can be expressed as

$$\begin{aligned} & \mathbf{h}_{\text{rMVDR}}(\theta_d, \phi_d) \\ &= \Gamma_{\mathbf{v},\epsilon}^{-1} \mathbf{D}(\theta_d, \phi_d) \left[\mathbf{D}^T(\theta_d, \phi_d) \Gamma_{\mathbf{v},\epsilon}^{-1} \mathbf{D}(\theta_d, \phi_d) \right]^{-1} \mathbf{i}_l. \end{aligned} \quad (80)$$

This comparison shows the different optimization criteria of the LDMG and MVDR beamformers, highlighting their respective advantages in various beamforming scenarios. The LDMG beamformer's ROI-centric approach provides greater flexibility and effectiveness in broader regions, while the MVDR and its robust variant perform better in situations with well-defined and stationary DOAs. A comparative evaluation of these beamformers is provided in Section V-C.

The robust MVDR beamformer has a computational complexity similar to that of the robust LDMG beamformer. Both methods involve computing the matrix $\Gamma_{\mathbf{v},\epsilon}$, which is one of the most computationally demanding steps. Additionally, the robust MVDR beamformer (80) involves matrix multiplications and inversions, resulting in a computational complexity of $\mathcal{O}[(ML_y)^3]$.

V. EXPERIMENTAL RESULTS

A. Simulation Setup

Consider a device equipped with microphones and loudspeakers. Such a device is suitable for professional meeting rooms. For simplicity, assume the device comprises a symmetric linear array of $M = 11$ omnidirectional microphones with interelement spacings of $[7.5, 7.5, 10, 5, 5, 5, 5, 10, 7.5, 7.5]$ cm. In addition to the microphone array, a loudspeaker is positioned on each side of the array between the two pairs of microphones with the interelement spacing of 10 cm.

For simulating a reverberant environment, we consider a scenario inside a room with width, length, and height dimensions of $5 \times 4 \times 6$ m. The device is placed along the y -axis with the center microphone, designated as the reference sensor, in the center of the room. To simulate the room impulse responses (RIR) from varying room positions to each microphone, we use the RIR generator [63]. The room reverberation time is set to $\text{RT60} = 0.4$ seconds, the sound propagation speed is $c = 340$ m/s, and the sample rate is $f_s = 16$ kHz. The generated RIRs, $g_m(t)$, have a length of $L_{\text{RIR}} = 3200$ samples.

We assume that the DOAs of all signals are from a polar angle of $\theta = 90^\circ$ and from azimuthal directions $\phi \in [-90^\circ, 90^\circ]$, where the origin of the spherical coordinate system is at the center of the room. The desired source signal is assumed to be located $r = 1$ meter from the reference microphone and its azimuthal direction is $\phi_d \in \Omega = [-15^\circ, 15^\circ]$. The scenario is illustrated in Fig. 2.

The choice of $\Omega = [-15^\circ, 15^\circ]$ reflects a scenario for professional meeting rooms where the source is expected to be within a moderate angular range. This range allows the beamformer to focus on a practical region where participants may typically sit. Scenarios with other ranges for the ROI, Ω , are considered at the end of Section V-B.

We now present a guideline for selecting the values of Δ , L_d , L_y , and l . In an anechoic environment, the value of Δ

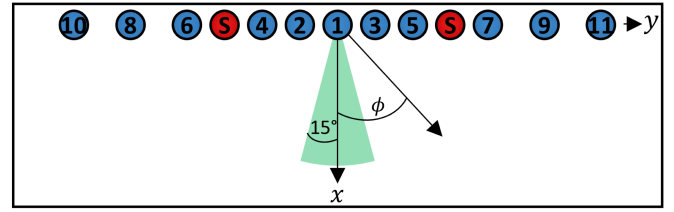


Fig. 2. Illustration of the device equipped with a symmetric array of 11 microphones (blue numbered circles) and two loudspeakers (red circles labeled 'S'). The green region indicates the region of interest. A beamformer is designed to suppress sound from the background and the loudspeakers while preserving signals from the region of interest.

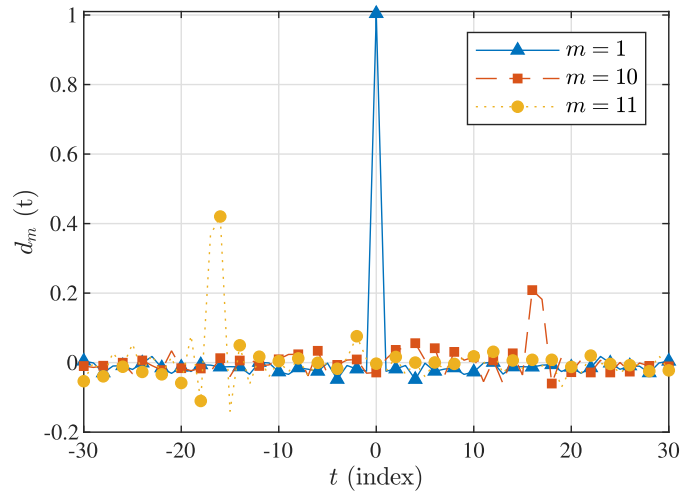


Fig. 3. Estimated relative impulse response of the center and outermost microphones for a source in the end-fire direction for $\Delta = 30$ and $L_d = 61$.

must be at least the sample delay it takes for a sound wave to propagate to the reference microphone from the farthest from it. This motivates us to choose the center microphone as the reference microphone. Therefore, $\Delta \geq \lceil \frac{0.35}{c} f_s \rceil = 17$. Due to the symmetry of the model (it takes the same time to propagate from (to) the reference microphone to (from) the farthest away microphone), it is sufficient to set the length of the relative impulse response to $L_d = 2\Delta + 1$. In a reverberative environment, the relative impulse response is a noncausal infinite-length filter. Therefore, we took a margin relative to the anechoic environment scenario and chose $\Delta = 30$ and $L_d = 2\Delta + 1 = 61$.

To ensure that Δ and L_d are large enough, one can verify that the estimated relative impulse responses converge to zero at the beginning and end of the vectors given in (8) for all m and for every angular direction. Relative impulse responses for a source positioned at $(r, \theta, \phi) = (1, 90^\circ, 90^\circ)$ are shown in Fig. 3. The response of the 11th sensor demonstrates the need for noncausal coefficients ($\Delta > 0$) to accommodate sound waves impinging the reference microphone after they have already impinged on other microphones. For source directions $\phi \in (0^\circ, 90^\circ)$ ($\phi \in [-90^\circ, 0^\circ)$) the sound wave impinges the reference microphone after it has already impinged on the odd (even) sensors.

According to the signal model (5) and assuming $x_1(t)$ is white, the $(l - L_d + 1)$ th component of the vector $\mathbf{y}_m(t)$ is

the first sample of the vector to correlate with the desired reference signal $x_1(t + \Delta - l + 1)$. For the first component to correlate, we should have $l = L_d$. The l th component of the vector $\mathbf{y}_m(t)$ is the last sample of the vector to correlate with the desired reference signal $x_1(t + \Delta - l + 1)$. For the last component to correlate, we should have $l = L_y$. Therefore, we choose $l = L_y = L_d = 61$. Indeed, with these values, we maintain $l = \lfloor L/2 \rfloor + 1$ as suggested in Section II.

The matrices required for the robust LDMG beamformer are estimated by their sample means as outlined in Section IV-A. We now detail the estimation process for the simulated scenario, which simulates using an external loudspeaker to play a broadband white stationary sound as the source signal. For simplicity, the microphones and loudspeakers are simulated to have an isotropic response, yet the procedure is the same if they have directivity. To simulate a broadband signal, we generated a white Gaussian noise signal of 1 s duration. For each azimuthal angle $\phi \in \Omega$, with a resolution of 1° , we generated the RIR from the position $(r, \theta, \phi) = (1, 90^\circ, \phi)$ to each microphone in the array. The signal at each microphone was then obtained by convolving the white Gaussian noise signal with the corresponding RIR. After simulating the white Gaussian noise signal radiating from all positions within the ROI, the matrices $\underline{\mathbf{D}}_\Omega$ and $\underline{\mathbf{\Gamma}}_{\underline{\mathbf{D}},\Omega}$ were computed according to (50) and (51), respectively. Similarly, the pseudo-correlation matrix of the diffuse noise, $\underline{\mathbf{\Gamma}}_0$, was calculated by simulating the white Gaussian noise signal to emit from the positions $(r, \theta, \phi) = (1, 90^\circ, \phi)$ for $\phi \in [-90^\circ, 90^\circ]$. Then, $\underline{\mathbf{\Gamma}}_0$ was computed according to (52). The pseudo-correlation matrix of the interference, $\underline{\mathbf{\Gamma}}_i$, was obtained by simulating the white Gaussian noise signal to emit from the positions of the two loudspeakers and then computed by its sample mean, similar to (51). In the simulated scenario, the pseudo-correlation matrix of the noise, $\underline{\mathbf{\Gamma}}_v$, was modeled as a weighted sum of the pseudo-correlation matrices of the interference and white noise, i.e., $\underline{\mathbf{\Gamma}}_v = 0.99\underline{\mathbf{\Gamma}}_i + 0.01\underline{\mathbf{I}}_{ML_y}$.

B. Performance Evaluation

Fig. 4 shows the performance of the robust LDMG beamformer as a function of K for several values of ϵ_0 , ϵ_1 , and μ . It demonstrates the tradeoff between array gain and distortion controlled by varying the parameter K . Decreasing K improves the average array gain but at the cost of degrading the average distortion; for $K = 1$ and $\epsilon_0 = \epsilon_1 = \mu = 0$ the highest array gain is achieved via the unconstrained maximum gain beamformer that minimizes the average distortion over Ω . The value of $\mu \in \{0, 0.01\}$ does not significantly influence the beamformer's performance, as long as K is not too high. This is because $\mu\lambda_{-1}$ is negligible compared to the largest eigenvalues. However, for high values of K , increasing μ improves the average array gain, WNG, and DF, but at the cost of degrading the average distortion; this is because $\mu\lambda_{-1}$ acts as a regularization term in (72). For most values of K , increasing ϵ_0 or ϵ_1 has the desired effect of improving the average DF or WNG and the average distortion. However, this improvement comes at the expense of the average array gain, which can be attributed to the weighting assigned to

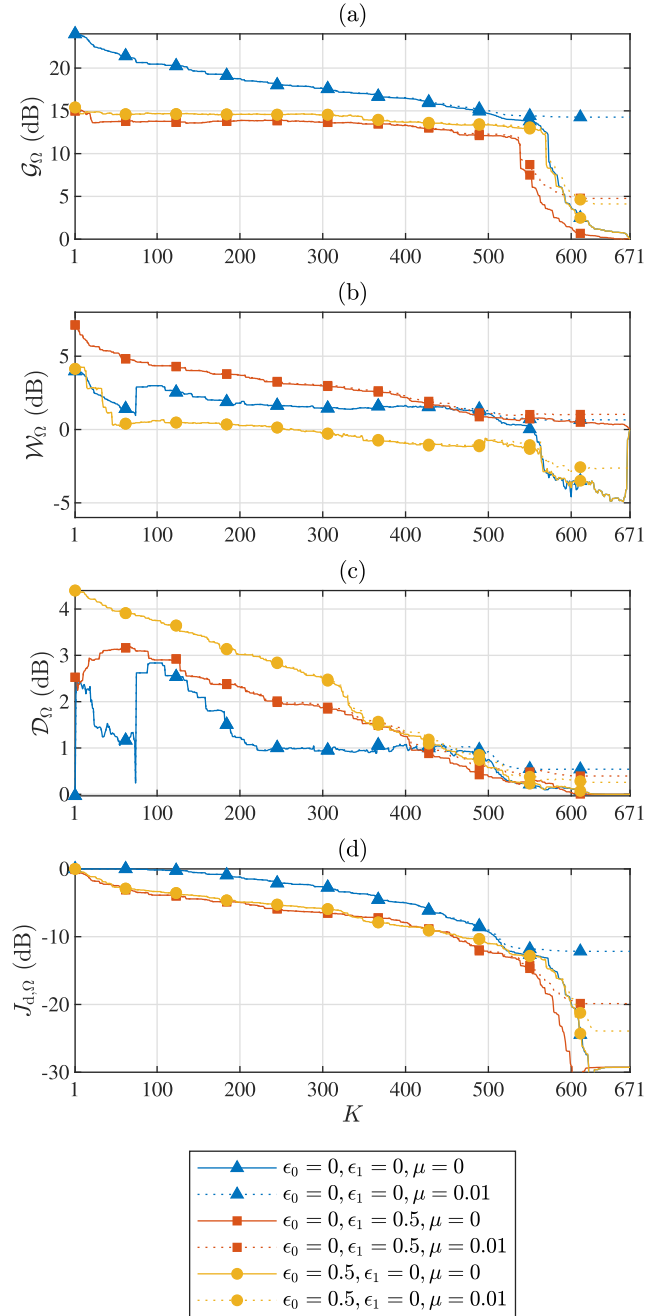


Fig. 4. Performance of the robust LDMG beamformer as a function of K for several values of ϵ_0 , ϵ_1 , and μ : (a) average array gain, (b) WNG, (c) DF, and (d) distortion.

each performance measure in the optimization problem formulated in (56).

Beam patterns are commonly used to illustrate the response of a beamformer to a plane wave arriving from the far-field as a function of its incident direction. However, the response is also a function of the source position for a nearby source in a reverberant environment. As detailed in the previous subsection, we estimated the matrices for a source located $r = 1$ m from the reference microphone and at a polar angle of 90° . Therefore, we illustrate the beamformer's response to a source located at $(r, \theta, \phi) = (1, 90^\circ, \phi)$ for $\phi \in [-90^\circ, 90^\circ]$. The broadband

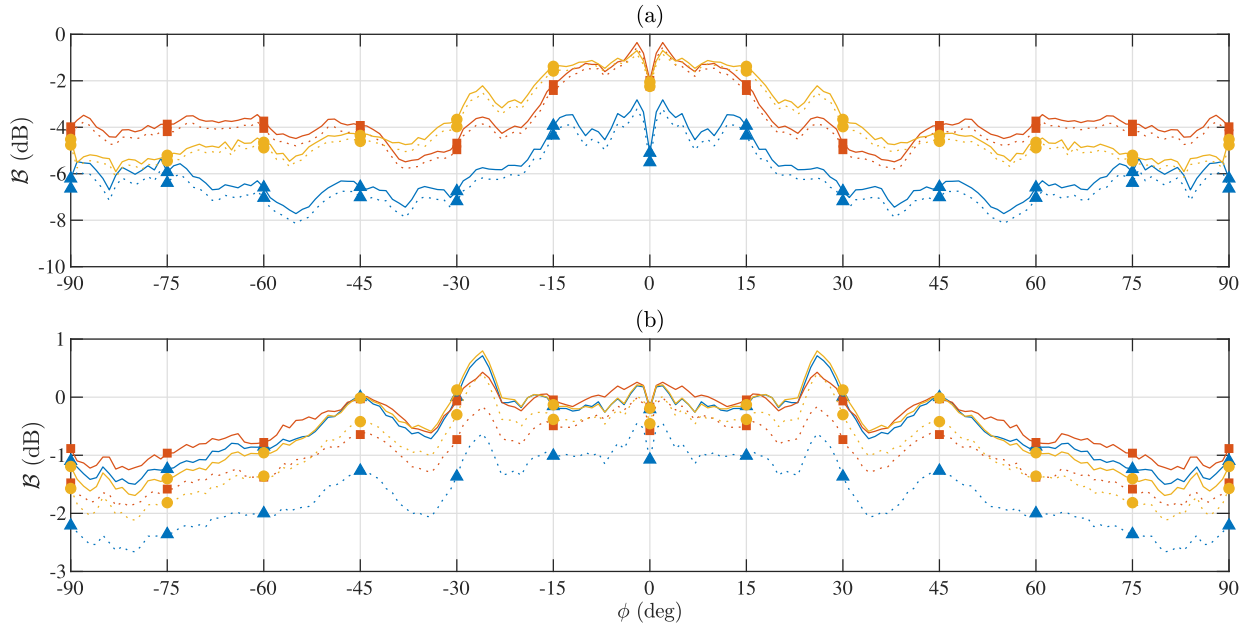


Fig. 5. Broadband beampatterns of the robust LDMG beamformer for several values of K , ϵ_0 , ϵ_1 , and μ : (a) $K = 350$, and (b) $K = 550$.

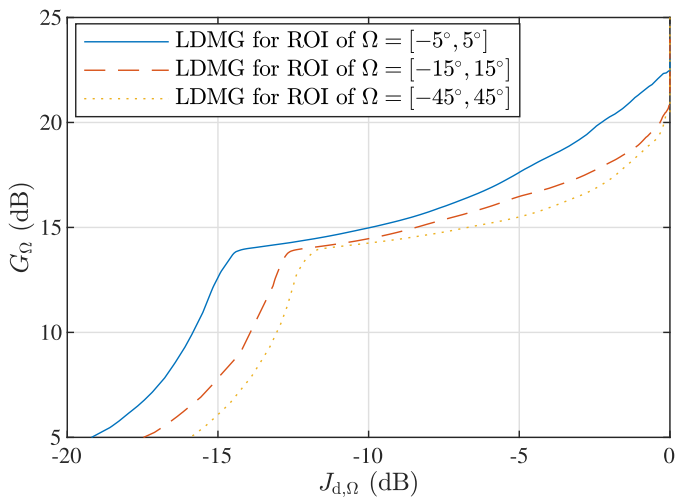


Fig. 6. Average array gain versus average distortion of the robust LDMG beamformer for different regions of interest (ROI). The three ROIs compared are: $\Omega = [-5^\circ, 5^\circ]$ (solid line), $\Omega = [-15^\circ, 15^\circ]$ (dashed line), and $\Omega = [-45^\circ, 45^\circ]$ (dotted line). Additional configurations for the beamformers: $\epsilon_0 = \epsilon_1 = \mu = 0$.

beampatterns of the robust LDMG are illustrated in Fig. 5 for $K \in \{350, 550\}$ for several values of ϵ_0 , ϵ_1 , and μ . For the larger value of K , the beampatterns are closer to 0 dB for $\phi \in \Omega$, which implies less distortion.

In Fig. 6, we examine the effect of the ROI width on the performance of the LDMG beamformer for $\epsilon_0 = \epsilon_1 = \mu = 0$. We consider three options: $\Omega = [-5^\circ, 5^\circ]$, $\Omega = [-15^\circ, 15^\circ]$, and $\Omega = [-45^\circ, 45^\circ]$. For each option, the average array gain is plotted against the corresponding average distortion, which is a function of K . We observe that for the same average distortion

level, the narrowest ROI achieves the highest average gain, while the broadest ROI yields the lowest average gain. This occurs because the narrower the ROI, the fewer directions the beamformer is constrained to limit distortion, allowing more degrees of freedom to maximize the average array gain. Therefore, the ROI should be configured as narrowly as possible while ensuring it contains the source position.

C. Comparison With Baseline Beamformers

We now compare the robust LDMG beamformer with the robust MVDR beamformer given in (80) and two additional baseline beamformers. For the LDMG beamformer, we set $K = 500$, $\mu = 0$, and $\Omega = [-15^\circ, 15^\circ]$. The robust MVDR beamformer was designed assuming $\phi_d = 0^\circ$. To obtain $\underline{\mathbf{D}}(\theta_d, \phi_d)$ for the robust MVDR beamformer, we simulated the white Gaussian noise signal to emit from the position $(r, \theta, \phi) = (1, 90^\circ, 0^\circ)$ and then computed its sample mean according to (49). The attained matrix $\underline{\mathbf{D}}^T(\theta_d, \phi_d) \underline{\mathbf{\Gamma}}_{\mathbf{v}, \epsilon}^{-1} \underline{\mathbf{D}}(\theta_d, \phi_d)$ was ill-conditioned, therefore we added a regularization term $\frac{\lambda_1}{67} \mathbf{I}_L$, where λ_1 was the largest generalized eigenvalue of $(\underline{\mathbf{\Gamma}}_{\mathbf{D}, \Omega}, \underline{\mathbf{\Gamma}}_{\mathbf{v}, \epsilon})$. The magnitude of the regularization term was tuned so that the LDMG and robust MVDR would have the same average array gain. This same regularization magnitude is also used in the other two baseline beamformers. We set $\epsilon_0 = 0$ and $\epsilon_1 = 0.1$ for all four beamformers.

The next baseline is an MVDR beamformer whose weighted average distortion over the ROI is constrained to equal 1, as proposed in [42]. While the method in [42] was formulated in the frequency domain, its structure allows a straightforward adaptation to the time domain. Formally, the beamformer we

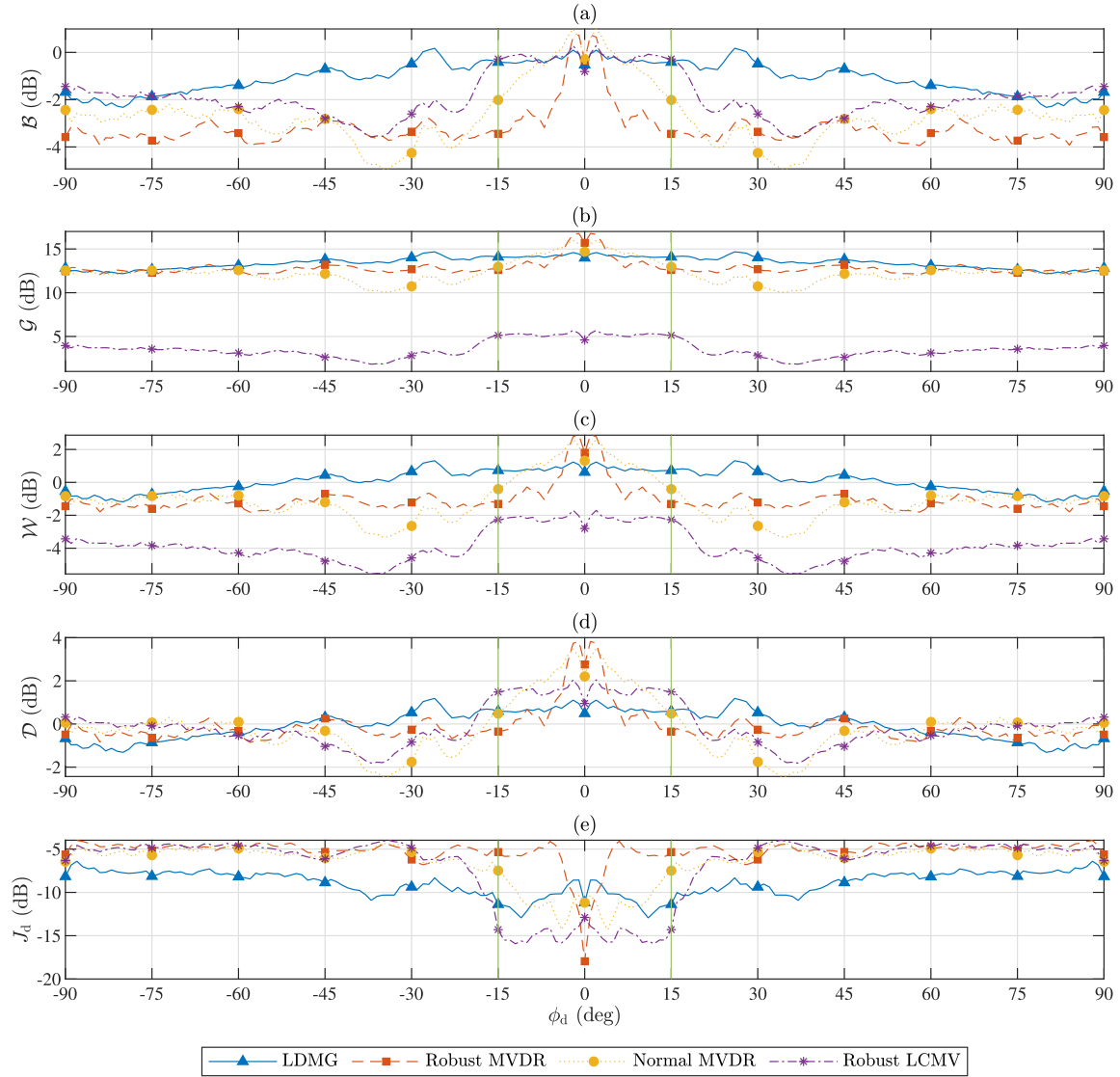


Fig. 7. Performances of the robust LDMG, robust MVDR, Normal MVDR, and robust LCMV beamformers as a function of the DOA: (a) broadband beampattern, (b) array gain, (c) WNG, (d) DF, and (e) distortion. Configurations for the LDMG beamformer: $\Omega = [-15^\circ, 15^\circ]$, $K = 500$, and $\mu = 0$. The robust MVDR beamformer was designed assuming $\phi_d = 0^\circ$. The Normal MVDR beamformer was designed with a standard deviation of $\sigma_N = 7.5^\circ$. The robust LCMV beamformer was designed with distortionless constraints over the entire ROI $\Omega = [-15^\circ, 15^\circ]$. For all beamformers: $\epsilon_0 = 0$ and $\epsilon_1 = 0.1$.

compare with is the solution to

$$\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}, \epsilon} \underline{\mathbf{h}} \quad \text{subject to} \quad \underline{\mathbf{D}}_p^T \underline{\mathbf{h}} = \mathbf{i}_l, \quad (81)$$

where

$$\underline{\mathbf{D}}_p = \int_0^\pi \int_0^{2\pi} p(\theta, \phi) \underline{\mathbf{D}}(\theta, \phi) \sin \theta d\phi d\theta \quad (82)$$

is a matrix of size $ML_y \times L$,

$$p(\theta, \phi) = \frac{p_N(\theta, \phi)}{\int_0^\pi \int_0^{2\pi} p_N(\theta, \phi) \sin \theta d\phi d\theta} \quad (83)$$

is a spatial probability function, where we choose

$$p_N(\theta, \phi) = e^{-\frac{(\phi - \phi_d)^2}{2\sigma_N^2}} \delta_D(\theta - \theta_d) \quad (84)$$

and $\sigma_N = 7.5^\circ$, representing a Normal spatial distribution centered at $(\theta_d, \phi_d) = (90^\circ, 0^\circ)$. In (84), $\delta_D(\theta)$ is the Dirac delta. The solution to (81) results in the following beamformer:

$$\underline{\mathbf{h}}_{\text{NMVDR}} = \underline{\Gamma}_{\mathbf{v}, \epsilon}^{-1} \underline{\mathbf{D}}_p \left(\underline{\mathbf{D}}_p^T \underline{\Gamma}_{\mathbf{v}, \epsilon}^{-1} \underline{\mathbf{D}}_p \right)^{-1} \mathbf{i}_l, \quad (85)$$

which we refer to as the Normal MVDR beamformer.

Last, we compare with the linearly constrained minimum variance (LCMV) beamformer [64] with distortionless constraints for each position in the ROI $\Omega = [-15^\circ, 15^\circ]$ with an azimuthal resolution of 1° . The idea of imposing multiple distortionless constraints has been proposed in [22]. Formally, the LCMV beamformer we compare with is the solution to

$$\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^T \underline{\Gamma}_{\mathbf{v}, \epsilon} \underline{\mathbf{h}} \quad \text{subject to} \quad \underline{\mathbf{D}}^T \underline{\mathbf{h}} = \bar{\mathbf{i}}_l, \quad (86)$$

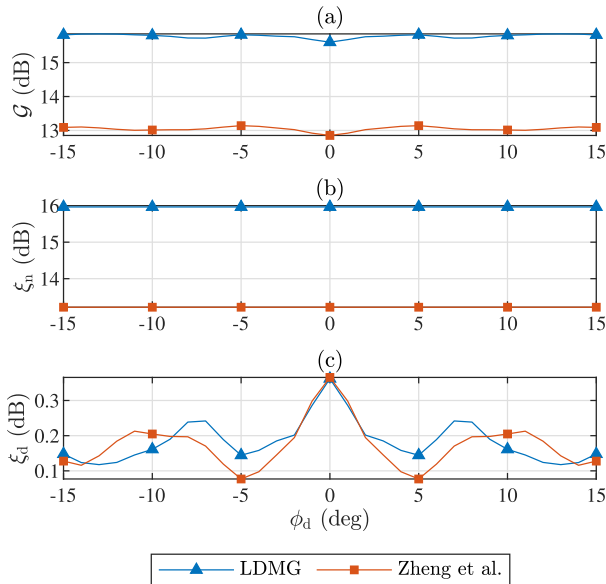


Fig. 8. Performances of the proposed LDMG beamformer and the ROI beamformer by Zheng et al. [29] as a function of the DOA for speech signals: (a) array gain, (b) noise reduction factor, and (c) desired signal reduction factor. Configurations for the LDMG beamformer: $\Omega = [-15^\circ, 15^\circ]$, $K = 500$, and $\epsilon_0 = \epsilon_1 = \mu = 0$.

where

$$\bar{\mathbf{D}} = [\mathbf{D}(\theta_d, -15^\circ) \quad \mathbf{D}(\theta_d, -14^\circ) \quad \cdots \quad \mathbf{D}(\theta_d, 15^\circ)] \quad (87)$$

is a matrix of size $ML_y \times 31L$, and

$$\bar{\mathbf{i}}_l = [\mathbf{i}_l^T \quad \mathbf{i}_l^T \quad \cdots \quad \mathbf{i}_l^T]^T \quad (88)$$

is a vector of length $31L$. The solution to (86) results in the following robust LCMV beamformer:

$$\mathbf{h}_{\text{LCMV}} = \mathbf{\Gamma}_{\mathbf{v}, \epsilon}^{-1} \bar{\mathbf{D}} \left(\bar{\mathbf{D}}^T \mathbf{\Gamma}_{\mathbf{v}, \epsilon}^{-1} \bar{\mathbf{D}} \right)^{-1} \bar{\mathbf{i}}_l. \quad (89)$$

To highlight the attributes of each beamformer, we measured the beamformers' performances as a function of the desired source DOA. We expect the robust MVDR beamformer to outperform the LDMG beamformer when the DOA is $\phi_d = 0^\circ$, which matches the design of the robust MVDR beamformer. In contrast, the LDMG should yield better results for other DOAs because it was optimized over a range of directions. We expect the robust LCMV to attain better distortion over the ROI because it was constrained to have a distortionless response over the entire ROI, potentially at the expense of the array gain. The performances of the LDMG, robust MVDR, Normal MVDR, and robust LCMV beamformers as a function of the DOA are plotted in Fig. 7.

As expected, the robust MVDR beamformer outperforms the LDMG beamformer when the desired source DOA is close to $\phi_d = 0^\circ$, due to minimal mismatch between the actual and assumed directions. The robust MVDR beamformer provides higher array gain, WNG, and DF, as well as lower distortion than the LDMG beamformer. However, when the source direction moves away from $\phi_d = 0^\circ$, the robust MVDR beamformer produces a significant increase in distortion. In contrast, the

LDMG beamformer maintains low distortion for all $\phi \in \Omega$. As a result, the LDMG beamformer has a better average distortion by 4.6 dB than the robust MVDR beamformer. Moreover, the LDMG beamformer achieves a better average WNG but a lower average DF. This is because of the ROI-distortion constraint applied to the entire ROI using the LDMG beamformer.

Also, as expected, the robust LCMV beamformer achieves lower distortion across the ROI than the LDMG beamformer. However, constraining a distortionless response across the entire ROI limits the available degrees of freedom for minimizing the noise variance, resulting in poor array gain.

The Normal MVDR beamformer attained a lower distortion and a higher array gain for a wider range of angles in the ROI compared with the robust MVDR beamformer. However, once the desired source DOA deviates even further, the LDMG beamformer achieves superior distortion and array gain. Overall, the LDMG beamformer's performances across the ROI are relatively constant due to the optimization problem from which it was derived, which maximized the average array gain over the ROI while maintaining the ROI-distortion constraint.

D. Speech Simulations

We now evaluate the performance of our proposed LDMG beamformer and the ROI beamformer by Zheng et al. [29] on a speech signal. We compare with Zheng et al. because they proposed a time-domain ROI beamformer capable of a tradeoff between noise reduction and signal distortion. The configurations for the LDMG beamformer are $\Omega = [-15^\circ, 15^\circ]$, $K = 500$, and $\epsilon_0 = \epsilon_1 = \mu = 0$. This configuration was selected to clearly illustrate the intrinsic performance differences between the methods without the influence of additional regularization or robustness terms.

Zheng et al. [29] assume near-field propagation of the signals to the microphones. The beamformer is then constrained to have a distortionless response for each position in the ROI $\Omega = [-15^\circ, 15^\circ]$ with an azimuthal resolution of 1° , and for each frequency in the band $[100, f_s/2]$ Hz with a resolution of 100 Hz. Formally, these constraints are expressed as

$$\mathbf{A}^T \mathbf{h} = \mathbf{g}_l, \quad (90)$$

where \mathbf{A} is an $ML_y \times 2A$ constraint matrix whose columns contain the real and imaginary parts of the A near-field steering vectors. The vector \mathbf{g}_l of length $2A$ holds the real and imaginary parts of the desired unit gain responses with latency of $(l-1)/f_s$ seconds. Here, $A = 2480$ is the total number of combinations of azimuthal directions and frequency bins used for the constraints. Zheng et al. relax these constraints through a singular value decomposition of the constraint matrix \mathbf{A} :

$$\mathbf{A} = \mathbf{V}\mathbf{S}\mathbf{U}^T, \quad (91)$$

where \mathbf{S} is an $ML_y \times 2A$ diagonal matrix whose main diagonal contains the singular values of \mathbf{A} in decreasing order, and the columns of matrices \mathbf{V} and \mathbf{U} are the corresponding singular vectors. By selecting the first 605 singular values and vectors, (90) is approximated by reduced-dimensional matrices:

$$\tilde{\mathbf{D}}^T \mathbf{h} = \tilde{\mathbf{i}}_l, \quad (92)$$

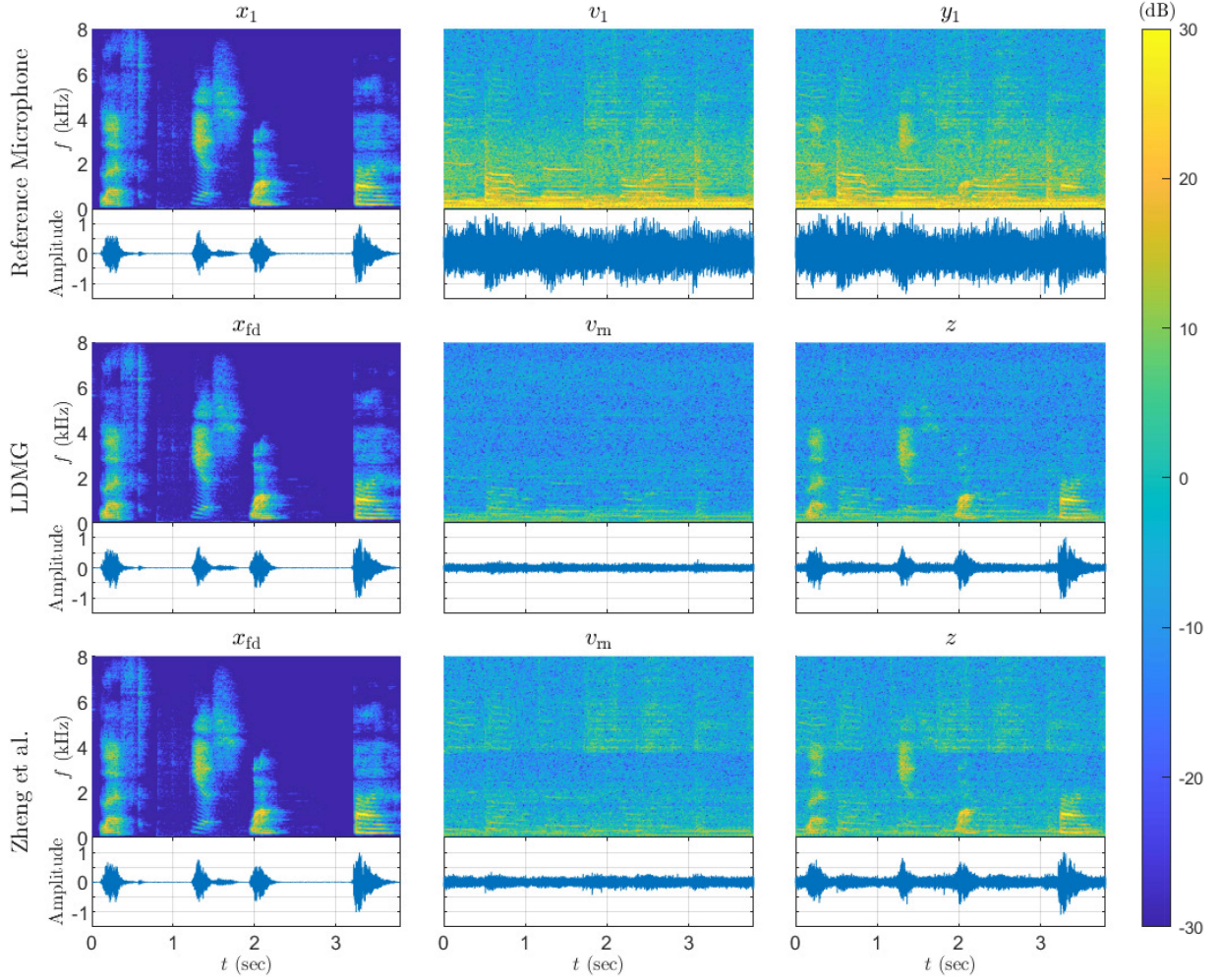


Fig. 9. Spectrograms and waveforms of the speech and noise signals received at the reference microphone and their sum (iSNR = -10 dB) and at the proposed LDMG and Zheng et al. [29] beamformers' output and their compositions [oSNR($\underline{\mathbf{h}}_{\text{LDMG},500,0,0}$) = 5.8 dB and oSNR($\underline{\mathbf{h}}_{\text{Zheng}}$) = 3.1 dB]. The desired speech originated from the direction $\phi_d = 15^\circ$. The beamformers were designed assuming $\Omega = [-15^\circ, 15^\circ]$.

where $\tilde{\mathbf{D}} = \mathbf{V}_{605}$ is a matrix of size $ML_y \times 605$, and $\tilde{\mathbf{i}}_l = \mathbf{S}_{605}^{-1} \mathbf{U}_{605}^T \mathbf{g}_l$ is a vector of length 605, where \mathbf{V}_{605} and \mathbf{U}_{605} consist of the first 605 columns of \mathbf{V} and \mathbf{U} , respectively, and \mathbf{S}_{605} is the upper-left 605×605 diagonal block of \mathbf{S} . We set the rank, which controls the tradeoff between noise reduction and signal distortion, to 605 so that the beamformer will attain the same signal distortion as our proposed LDMG beamformer. The final beamformer is obtained by minimizing the residual noise variance under the relaxed constraints. Finally, the beamformer we compare with is given as the solution to

$$\min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^T \Gamma_{\underline{\mathbf{v}}} \underline{\mathbf{h}} \quad \text{subject to} \quad \tilde{\mathbf{D}}^T \underline{\mathbf{h}} = \tilde{\mathbf{i}}_l, \quad (93)$$

which yields

$$\underline{\mathbf{h}}_{\text{Zheng}} = \Gamma_{\underline{\mathbf{v}}}^{-1} \tilde{\mathbf{D}} \left(\tilde{\mathbf{D}}^T \Gamma_{\underline{\mathbf{v}}}^{-1} \tilde{\mathbf{D}} \right)^{-1} \tilde{\mathbf{i}}_l. \quad (94)$$

For each $\phi \in \Omega$, we simulate the desired speech signal to originate from the position $(r, \theta, \phi) = (1, 90^\circ, \phi)$. The desired speech comprises four words spoken in English, each by a different person (two males and two females), with a total

duration of 3.8 s. The interference is a simultaneous playback of music from both loudspeakers. The level of the spatially white noise was set at 1/100th of the noise variance at the reference sensor. Altogether, the covariance matrix of the additive noise was estimated as $\Gamma_{\underline{\mathbf{v}}} = 0.99\Gamma_{\underline{\mathbf{i}}} + 0.01\mathbf{I}_{ML_y}$. The iSNR was set to -10 dB. For each simulation, the signals captured by the microphones were filtered by both beamformers for comparison.

The performance measures for a simulated signal are given by

$$\text{iSNR} = \frac{\text{var}[x_1(t)]}{\text{var}[v_1(t)]}, \quad (95)$$

$$\text{oSNR}(\underline{\mathbf{h}}, \theta_d, \phi_d) = \frac{\text{var}[x_{\text{id}}(t)]}{\text{var}[v_m(t)]}, \quad (96)$$

$$\mathcal{G}(\underline{\mathbf{h}}, \theta_d, \phi_d) = \frac{\text{oSNR}(\underline{\mathbf{h}}, \theta_d, \phi_d)}{\text{iSNR}}, \quad (97)$$

$$\xi_n(\underline{\mathbf{h}}, \theta_d, \phi_d) = \frac{\text{var}[v_1(t)]}{\text{var}[v_m(t)]}, \quad (98)$$

$$\xi_d(\mathbf{h}, \theta_d, \phi_d) = \frac{\text{var}[x_1(t)]}{\text{var}[x_{fd}(t)]}, \quad (99)$$

where $\text{var}(\cdot)$ is the empirical variance.

After each simulation, the performance measures given in (97)–(99) were calculated and are plotted in Fig. 8. Both beamformers demonstrated a low desired signal reduction factor for each source position within the ROI, with an average of only 0.2 dB. The proposed LDMG beamformer achieved a higher noise reduction factor, resulting in an average array gain 2.7 dB greater than the beamformer suggested by Zheng et al. [29]. This improvement highlights the advantages of our method, which leverages relative impulse responses and joint diagonalization.

Fig. 9 illustrates the waveforms and spectrograms for the experiment when the desired source direction is $\phi_d = 15^\circ$. The top subplots display the speech and noise received at the reference microphone and their sum: $x_1(t)$, $v_1(t)$, and $y_1(t)$. The middle and bottom subplots respectively display the LDMG and Zheng et al. [29] beamformers' output and their compositions: $x_{fd}(t)$, $v_m(t)$, and $z(t)$. While both beamformers introduce minimal distortion, it is evident that at higher frequencies, the beamformer by Zheng et al. [29] offers less interference suppression than the proposed LDMG beamformer. This can be attributed to our formulation's ability to capture the reverberation characteristics through the use of relative impulse responses, thereby enabling effective suppression of high-frequency interference.

VI. CONCLUSION

We presented a novel time-domain beamformer that maximizes the gain in signal-to-noise ratio within a designated spatial ROI under a constraint of minimum average distortion. This beamformer is especially valuable in scenarios where precise information about the source position is unavailable, such as in cases of source movement or estimation errors caused by background noise, interfering sources, and reverberation. Our work presents several key contributions: we established performance measures tailored for ROI beamforming and introduced a novel distortion constraint, which enabled the formulation of an optimization problem aimed at maximizing the average array gain under this constraint. By leveraging joint diagonalization, we derived an analytical solution and demonstrated through simulations that our beamformer achieves low distortion for signals from any direction within the ROI while effectively attenuating noise and interference. We showed that limiting the number of generalized eigenvectors allows a tradeoff between array gain and distortion, and that additional tradeoff parameters enhance robustness to spatially white and diffuse noise fields, albeit at the cost of reduced array gain.

Comparative analysis showcased that the proposed ROI-centric beamformer outperformed the robust MVDR beamformer in terms of average distortion across the ROI, whereas the robust MVDR beamformer excelled when the DOA was precisely known. We also outlined a practical approach for computing the beamformer that incorporates the acoustic characteristics of the microphone array. However, to calibrate the system, the approach required an external loudspeaker to play

a broadband white stationary sound as the source signal from multiple room positions. Additional limitations and challenges include manipulating large covariance matrices due to being a time-domain method. Moreover, while our method offers interpretability and control over performance tradeoffs, it lacks the adaptability of data-driven approaches and requires manual tuning of hyperparameters.

Future work could address these limitations by enabling adaptation to dynamic environments and optimizing sensor placement to improve performance. Overall, our beamformer has the potential to contribute to cutting-edge technologies, especially in the field of microphone arrays for professional conference rooms, smart homes, and wearable devices, where real-time, low-latency, computationally efficient, and high-quality audio processing is crucial.

REFERENCES

- [1] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [2] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2002.
- [3] I. Cohen, J. Benesty, and S. Gannot, Eds., *Speech Processing in Modern Communication: Challenges and Perspectives*. Berlin, Germany: Springer, 2010.
- [4] L. C. De Silva, C. Morikawa, and I. M. Petra, "State of the art of smart homes," *Eng. Appl. Artif. Intell.*, vol. 25, no. 7, pp. 1313–1321, 2012.
- [5] K. Platz and B. Shumard, "Voices carry: Beamforming technology helps everyone be heard," *Syst. Contractor News*, vol. 30, no. 10, pp. 32–33, 2023.
- [6] D. Watanabe, Y. Takeuchi, T. Matsumoto, H. Kudo, and N. Ohnishi, "Communication support system of smart glasses for the hearing impaired," in *Proc. 16th Int. Conf. Comput. Helping People Special Needs*, 2018, pp. 225–232.
- [7] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. PROC-57, no. 8, pp. 1408–1418, Aug. 1969.
- [8] J. Benesty, I. Cohen, and J. Chen, *Beamforming in the Time Domain*. Hoboken, NJ, USA: Wiley, 2018, pp. 397–443.
- [9] D. B. Ward, R. A. Kennedy, and R. C. Williamson, *Constant Directivity Beamforming*. Berlin, Germany: Springer, 2001, pp. 3–17.
- [10] Y. Zhao, W. Liu, and R. J. Langley, "Efficient design of frequency invariant beamformers with sensor delay-lines," in *Proc. 5th IEEE Sensor Array Multichannel Signal Process. Workshop*, 2008, pp. 335–339.
- [11] S. Yan, H. Sun, X. Ma, U. P. Svensson, and C. Hou, "Time-domain implementation of broadband beamformer in spherical harmonics domain," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 5, pp. 1221–1230, Jul. 2011.
- [12] W. Zhu and W. Wu, "Design of wide-band array with frequency invariant beam pattern by using adaptive synthesis method," in *Proc. Int. Conf. Image Anal. Signal Process.*, 2011, pp. 688–693.
- [13] Z. Wang, J. Li, P. Stoica, T. Nishida, and M. Sheplak, "Constant-beamwidth and constant-powerwidth wideband robust capon beamformers for acoustic imaging," *J. Acoust. Soc. America*, vol. 116, no. 3, pp. 1621–1631, 2004.
- [14] A. Frank and I. Cohen, "Constant-beamwidth Kronecker product beamforming with nonuniform planar arrays," *Front. Signal Process.*, vol. 2, 2022, Art. no. 829463.
- [15] O. Peretz and I. Cohen, "Constant elevation-beamwidth beamforming with concentric ring arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 1662–1672, 2024.
- [16] 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, Apr. 2019.

- [17] X. Wenmeng, B. Changchun, J. Maoshen, and J. Picheral, "Speech enhancement with robust beamforming for spatially overlapped and distributed sources," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 2778–2790, 2022.
- [18] F. Zhang, C. Pan, J. Benesty, and J. Chen, "On the design and implementation of maximum SNR beamformers for scattered speech sources," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, 2023, pp. 1–6.
- [19] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.
- [20] Y. Avargel and I. Cohen, "System identification in the short-time fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.
- [21] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 546–555, May 2009.
- [22] K. Takao, M. Fujita, and T. Nishi, "An adaptive antenna array under directional constraint," *IEEE Trans. Antennas Propag.*, vol. TAP-24, no. 5, pp. 662–669, Sep. 1976.
- [23] Y. Grenier, "A microphone array for car environments," *Speech Commun.*, vol. 12, no. 1, pp. 25–39, 1993.
- [24] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
- [25] M. Kajala and M. Hamalainen, "Filter-and-sum beamformer with adjustable filter characteristics," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2001, vol. 5, pp. 2917–2920.
- [26] S. Zhang and I. L. J. Thng, "Robust presteering derivative constraints for broadband antenna arrays," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 1–10, Jan. 2002.
- [27] A. Tkacenko, P. Vaidyanathan, and T. Q. Nguyen, "On the eigenfilter design method and its applications: A tutorial," *IEEE Trans. Circuits Syst. II. Analog Digit. Signal Process.*, vol. 50, no. 9, pp. 497–517, Sep. 2003.
- [28] S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using Eigenfilters," *Signal Process.*, vol. 83, no. 12, pp. 2641–2673, 2003.
- [29] Y. R. Zheng, R. A. Goubran, and M. El-Tanany, "Robust near-field adaptive beamforming with distance discrimination," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 478–488, Sep. 2004.
- [30] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1684–1696, May 2005.
- [31] J. Chen, L. Shue, H. Sun, and K. Phua, "An adaptive microphone array with local acoustic sensitivity," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2005, pp. 1–4.
- [32] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 425–437, Sep. 1997.
- [33] N. Grbić and S. Nordholm, "Soft constrained subband beamforming for hands-free speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, pp. 885–888.
- [34] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.
- [35] E. Warsitz, A. Krueger, and R. Haeb-Umbach, "Speech enhancement with a new generalized Eigenvector blocking matrix for application in a generalized sidelobe canceller," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2008, pp. 73–76.
- [36] A. Pezeshki, B. D. Van Veen, L. L. Scharf, H. Cox, and M. L. Norderivaad, "Eigenvalue beamforming using a multitrack MVDR beamformer and subspace selection," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1954–1967, May 2008.
- [37] E. Mabande and W. Kellermann, "Design of robust polynomial beamformers as a convex optimization problem," in *Proc. IEEE Int. Workshop Acoustic Echo, Noise Control*, 2010, pp. 106–110.
- [38] J. Martinez, N. Gaubitch, and W. B. Kleijn, "A robust region-based near-field beamformer," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 2494–2498.
- [39] Z. Zhong, K. Itoyama, K. Nishida, and K. Nakadai, "Design and assessment of a scan-and-sum beamformer for surface sound source separation," in *Proc. IEEE/SICE Int. Symp. Syst. Integration*, 2020, pp. 808–813.
- [40] Z. Zhong, M. Shakeel, K. Itoyama, K. Nishida, and K. Nakadai, "Assessment of a beamforming implementation developed for surface sound source separation," in *Proc. IEEE/SICE Int. Symp. Syst. Integration*, 2021, pp. 369–374.
- [41] A. Sofer, T. Kounovsky, J. Čmejla, Z. Koldovsky, and S. Gannot, "Robust relative transfer function identification on manifolds for speech enhancement," in *Proc. IEEE 29th Eur. Signal Process. Conf.*, 2021, pp. 401–405.
- [42] G. Itzhak and I. Cohen, "Robust beamforming for multispeaker audio conferencing under DOA uncertainty," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 33, pp. 139–151, 2025.
- [43] A. Davis, S. Y. Low, S. Nordholm, and N. Grbić, "A subband space constrained beamformer incorporating voice activity detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 65–68.
- [44] M. Taseska and E. A. P. Habets, "Spotforming: Spatial filtering with distributed arrays for position-selective sound acquisition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 7, pp. 1291–1304, Jul. 2016.
- [45] J. Zhang, J. Du, and L. -R. Dai, "Sensor selection for relative acoustic transfer function steered linearly-constrained beamformers," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1220–1232, 2021.
- [46] T. Jenrungrot, V. Jayaram, S. Seitz, and I. Kemelmacher-Shlizerman, "The cone of silence: Speech separation by localization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 20925–20938.
- [47] K. Tesch and T. Gerkmann, "Multi-channel speech separation using spatially selective deep non-linear filters," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 542–553, 2024.
- [48] R. Gu and Y. Luo, "ReZero: Region-customizable sound extraction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 2576–2589, 2024.
- [49] Z. Xu et al., "FoVNet: Configurable field-of-view speech enhancement with low computation and distortion for smart glasses," in *Proc. Interspeech*, 2024, pp. 3350–3354.
- [50] T. Chen, M. Itani, S. E. Eskimez, T. Yoshioka, and S. Gollakota, "Hearable devices with sound bubbles," *Nature Electron.*, vol. 7, pp. 1047–1058, 2024.
- [51] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, vol. 1. Berlin, Germany: Springer, 2008.
- [52] D. D. Dirks and S. Gilman, "Exploring azimuth effects with an anthropometric manikin," *J. Acoust. Soc. America*, vol. 66, no. 3, pp. 696–701, 1979.
- [53] R. M. Corey, N. Tsuda, and A. C. Singer, "Acoustic impulse responses for wearable audio devices," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 216–220.
- [54] Q. Xu and Y. Huang, *Anechoic and Reverberation Chambers: Theory, Design, and Measurements*. Chichester, U.K.: Wiley, 2019.
- [55] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. 14th Int. Workshop Acoustic Signal Enhancement*, 2014, pp. 313–317.
- [56] J. Čmejla, T. Kounovsky, S. Gannot, Z. Koldovsky, and P. Tandeitnik, "Mirage: Multichannel database of room impulse responses measured on high-resolution cube-shaped grid," in *Proc. IEEE 28th Eur. Signal Process. Conf.*, 2021, pp. 56–60.
- [57] T. Dietzen, R. Ali, M. Taseska, and T. van Waterschoot, "MYriAD: A multi-array room acoustic database," *Eurasip J. Audio, Speech, Music Process.*, vol. 2023, no. 1, 2023, Art. no. 17.
- [58] J. R. Jensen, J. Benesty, and M. G. Christensen, "Noise reduction with optimal variable span linear filters," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 631–644, Apr. 2016.
- [59] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1968.
- [60] W. Gabriel, "Using spectral estimation techniques in adaptive processing antenna systems," *IEEE Trans. Antennas Propag.*, vol. TAP-34, no. 3, pp. 291–300, Mar. 1986.
- [61] J. Zhang, H. Chen, L. -R. Dai, and R. C. Hendriks, "A study on reference microphone selection for multi-microphone speech enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 671–683, 2021.
- [62] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.
- [63] E. A. P. Habets, "Room impulse response generator," 2010. [Online]. Available: <https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator>
- [64] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. PROC-60, no. 8, pp. 926–935, Aug. 1972.