JOINT OPTIMIZATION OF MICROPHONE ARRAY GEOMETRY AND REGION-OF-INTEREST BEAMFORMING WITH SPARSE CIRCULAR SECTOR ARRAYS

Gal Itzhak[1] *Simon Doclo*[2] *Israel Cohen*[1]

[1]Faculty of Electrical & Computer Engineering, Technion–Israel Institute of Technology
[2]Department of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg

ABSTRACT

This paper focuses on two key aspects: region-of-interest beamforming and optimal sparse circular sector array design. The aim is to address the problem of the unknown direction of arrival within a given region of interest while optimizing the array geometry layout and beamformer taps. This is done while meeting the maximum broadband array directivity criterion. To ensure that the desired signal is not distorted, we apply appropriate optimization constraints while maintaining a sufficiently high white noise gain. Our proposed approach outperforms a recently suggested approach in terms of the directivity factor, especially when the direction of arrival of the desired source significantly deviates from its nominal value.

Index Terms— Microphone arrays, optimal beamforming, region-of-interest beamforming, joint optimization, sparse arrays.

1. INTRODUCTION

Beamforming has been a widespread technique over the last few decades to estimate desired signals from noisy sensor observations [1–5]. Beamformers use signals from different locations in space and can be used, e.g., to attenuate undesired background noise, interferences, and reverberations, while preserving the desired source free of distortion [6–9]. When designing a beamformer, two principal matters should be considered: its taps and the underlying array geometry. While the former has been given much attention, the latter has been investigated to a lesser extent, even though its impact on the performance of the beamformer may be significant [10, 11].

Uniform linear arrays (ULAs) are a commonly used array layout that can achieve high array directivity or white noise robustness, but not both [12–14]. In addition, ULAs are heavily influenced by the direction of arrival (DOA) of the desired source [15, 16] and are sensitive to microphone imperfections [17]. Other array configurations have been considered to reduce susceptibility to the DOA of the desired source. For example, rectangular arrays (RAs) and cube arrays have been shown to perform well when the DOA is parallel to one of the array axes, but their performance tends to decline when the DOA is not parallel [18–20]. Circular (and concentric circular) arrays have the potential to exhibit DOA-independent performance (considering only the azimuth angle) [21–24]. However, this implies limited array directivity or requires accurate knowledge of the true DOA, which is impractical in most real-world scenarios.

Recent studies suggest using the concept of a region of interest (ROI) to handle DOA mismatches and adapt to changing settings. An ROI is a continuous space region from which the desired source is assumed to originate. The arrangement of the microphone array is then optimized for a certain ROI. In a recent study by Konforti et al. [25], a linear layout was proposed to maximize the array's broadband directivity while keeping the white noise gain above a specified level. This approach outperformed traditional methods, but was limited to narrow ROIs and required knowledge of the true DOA. Itzhak and Cohen [26] suggested optimizing a rectangular array with a uniform structure along one axis and a non-uniform structure along the other. This method was more flexible and ensured a constant mainlobe beamwidth above a particular threshold frequency. However, it was only suitable for relatively small ROIs and required many array microphones and a large aperture along one of its axes. Another study by Itzhak and Cohen [27] introduced a beamforming technique based on optimizing a sparse concentric circular array. This approach improved performance for wider ROIs, even with a limited number of microphones. However, the sequential optimization procedure of the array geometry followed by the derivation of the taps of the underlying beamformer yielded a sub-optimal solution. Additionally, the derivation method of the taps did not directly consider the desired ROI.

This paper introduces an ROI beamforming method that uses a sparse circular sector array (SCSA) layout to address significant deviations in the DOA of the desired source. We assume that the true DOA is unknown and formulate the problem accordingly. We then use this formulation to maximize the broadband directivity while directly accounting for the desired ROI. The optimization is performed simultaneously over the array geometry and the beamformer taps. Additionally, design constraints control the white noise gain and signal distortion. Compared to a recently proposed approach [27], our approach is preferred in terms of the directivity factor, especially when substantial DOA deviations are present.

The remainder of the paper is organized as follows. In Section 2, we present the signal model. In Section 3, we address the problem of DOA uncertainty and define the appropriate measures for ROI beamforming. In Section 4, we present our approach to jointly optimize the array geometry and the beamformer taps. In Section 5, we analyze and compare the performance of the proposed approach through simulations.

2. SIGNAL MODEL

Consider a far-field desired source propagating from an azimuth angle ϕ and a polar angle θ in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s. The plane wave impinges on a two-dimensional (2-D) uniform circular sector array (UCSA) located on the x-y plane. The UCSA is composed of M uniformly-spaced omnidirectional microphones along the radial direction with an interelement spacing δ_r and P uniformly-spaced omnidirectional microphones along the angular direction. The locations of the latter

This research was supported by the Israel Science Foundation (grant no. 1449/23) and the Pazy Research Foundation.

are lower and upper bounded by ψ_L and ψ_H, respectively, and are given by

$$\psi_p = p \times \frac{\psi_H - \psi_L}{P-1} + \psi_L, \tag{1}$$

with $p = 0, \ldots, P-1$. An illustration of the UCSA layout is depicted in Fig. 1 by considering all empty and solid circles therein. In addition, the microphone located at the center of the circle (the origin of the Cartesian coordinate system) is considered the reference microphone. The array steering vector associated with ψ_p of length $M-1$ (excluding the common reference microphone) is given by [28]

$$\mathbf{a}_{\theta,\phi,\psi_p}(f) = \left[\; e^{\jmath 2\pi f \delta_r \sin\theta\cos(\phi-\psi_p)/c} \quad e^{\jmath 4\pi f \delta_r \sin\theta\cos(\phi-\psi_p)/c} \right. \tag{2}$$
$$\left. \cdots \quad e^{\jmath 2\pi f \delta_r (M-1)\sin\theta\cos(\phi-\psi_p)/c} \; \right]^T,$$

where the superscript ^{T} denotes the transpose operator, $\jmath = \sqrt{-1}$ is the imaginary unit, and $f > 0$ is the temporal frequency. Stacking together all steering vectors $\{\mathbf{a}_{\theta,\phi,\psi_p}(f)\}_{p=0}^{P-1}$ we obtain the array steering vector $\mathbf{d}_{\theta,\phi}(f)$ of length $P(M-1)+1$ by:

$$\mathbf{d}_{\theta,\phi}(f) = \left[\; 1 \quad \mathbf{a}^T_{\theta,\phi,\psi_0}(f) \quad \cdots \quad \mathbf{a}^T_{\theta,\phi,\psi_{P-1}}(f) \; \right]^T, \tag{3}$$

in which the first element corresponds to the reference microphone.

Denoting the desired source incident angle by (θ_0, ϕ_0), the observed noisy signal vector of length $P(M-1)+1$ can be expressed in the frequency domain as [8]:

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}_{\theta_0,\phi_0}(f) X(f) + \mathbf{v}(f), \end{aligned} \tag{4}$$

where $X(f)$ and $\mathbf{v}(f)$ are the zero-mean desired source and additive noise signal vectors, respectively, as received by the reference microphone. Dropping the dependence on f, the correlation matrix of $\mathbf{y}$ is given by

$$\mathbf{\Phi}_{\mathbf{y}} = E\left(\mathbf{y}\mathbf{y}^H\right) = p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0} + \mathbf{\Phi}_{\mathbf{v}}, \tag{5}$$

where $E(\cdot)$ denotes mathematical expectation, the superscript ^{H} is the conjugate-transpose operator, $p_X = E\left(|X|^2\right)$ is the power spectral density of the desired source at the reference microphone, and $\mathbf{\Phi}_{\mathbf{v}} = E\left(\mathbf{v}\mathbf{v}^H\right)$ is the correlation matrix of $\mathbf{v}$. Note that equation (5) assumes $\mathbf{x}$ and $\mathbf{v}$ to be uncorrelated. Assuming the noise variance is approximately uniform across all sensors, equation (5) may be expressed as

$$\mathbf{\Phi}_{\mathbf{y}} = p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0} + p_V \mathbf{\Gamma}_{\mathbf{v}}, \tag{6}$$

where p_V is the power spectral density of the noise at the reference microphone and $\mathbf{\Gamma}_{\mathbf{v}} = \mathbf{\Phi}_{\mathbf{v}}/p_V$ is the pseudo-coherence matrix of the noise. From (6), we deduce that the input signal-to-noise ratio (SNR) is

$$\mathrm{iSNR} = \frac{\mathrm{tr}\left(p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0}\right)}{\mathrm{tr}\left(p_V \mathbf{\Gamma}_{\mathbf{v}}\right)} = \frac{p_X}{p_V}, \tag{7}$$

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

3. REGION-OF-INTEREST BEAMFORMING

To generate an estimate of the desired source X, a linear beamformer $\mathbf{f} = \left[\; F_1 \quad \cdots \quad F_{P(M-1)+1} \; \right]^T$ is applied to the observed signal vector $\mathbf{y}$, yielding the output signal [29]

$$\hat{X} = \mathbf{f}^H\mathbf{y} = X\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0} + \mathbf{f}^H\mathbf{v}. \tag{8}$$

If $\mathbf{d}_{\theta_0,\phi_0}$ is known, a distortionless response constraint for the desired source is given by $\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0} = 1$, which may directly be used in the derivation of the beamformer $\mathbf{f}$.

The most common performance measures for beamformers are the SNR gain, the white noise gain (WNG), and the directivity factor (DF). From (8), the output SNR can be defined as

$$\mathrm{oSNR}(\mathbf{f}) = \frac{p_X}{p_V} \times \frac{|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}, \tag{9}$$

which implies that the SNR gain is given by

$$\mathcal{G}(\mathbf{f}) = \frac{\mathrm{oSNR}(\mathbf{f})}{\mathrm{iSNR}} = \frac{|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}. \tag{10}$$

Consequently, the WNG and DF are given by

$$\mathcal{W}(\mathbf{f}) = \frac{|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}|^2}{\mathbf{f}^H\mathbf{f}}, \tag{11}$$

$$\mathcal{D}(\mathbf{f}) = \frac{|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}}, \tag{12}$$

where $\mathbf{\Gamma}_{\mathrm{d}}$ is the pseudo-coherence matrix of a spherically isotropic (diffuse) noise field [29], defined by $[\mathbf{\Gamma}_{\mathrm{d}}]_{i_1,i_2} = \mathrm{sinc}(2\pi f \Delta_{i_1,i_2}/c)$, where i_1 and i_2 denote microphone indices, Δ_{i_1,i_2} is the Euclidean distance between microphone i_1 and i_2, and $\mathrm{sinc}(x) = \sin(x)/x$.

In many real-world scenarios, involving multiple and possibly moving speakers of interest, the DOA of the desired source is unknown. Hence, the performance measures in (9)-(12) are inappropriate to properly derive $\mathbf{f}$, implying that other measures should be employed instead. Let us consider the steering vector associated with the unknown DOA, $\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}$, to be a random variable characterized by the probability density function $p_{\mathbf{d}}(\theta,\phi)$. Then, we may define

$$\begin{aligned} \bar{\mathbf{x}} &= XE[\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}] \\ &= X\int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}} p_{\mathbf{d}}(\theta,\phi)\mathbf{d}_{\theta,\phi}\sin\theta \mathrm{d}\phi\mathrm{d}\theta, \end{aligned} \tag{13}$$

which constitutes a weighted sum over all possible DOAs within the ROI: $\theta \in \Theta_{\mathrm{ROI}}, \phi \in \Phi_{\mathrm{ROI}}$.

Substituting (13) into (9), we may define the output SNR over the entire ROI as

$$\begin{aligned} \mathrm{oSNR}_{\mathrm{ROI}}(\mathbf{f}) &= \frac{1}{p_V} \times \frac{E\left[|\mathbf{f}^H\bar{\mathbf{x}}|^2\right]}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}} \\ &= \frac{p_X}{p_V} \times \frac{1}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}} \\ &\quad \times \frac{|\mathbf{f}^H\int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}}\mathbf{d}_{\theta,\phi}\sin\theta\mathrm{d}\phi\mathrm{d}\theta|^2}{|\int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}}\sin\theta\mathrm{d}\phi\mathrm{d}\theta|^2} \\ &= \frac{p_X}{p_V} \times \frac{1}{\Omega^2_{\mathrm{ROI}}} \times \frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}, \end{aligned} \tag{14}$$

where $\mathbf{b}_{\mathrm{ROI}} = \int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}}\mathbf{d}_{\theta,\phi}\sin\theta\mathrm{d}\phi\mathrm{d}\theta$ may be regarded as the ROI steering vector whose integration bounds are set by the

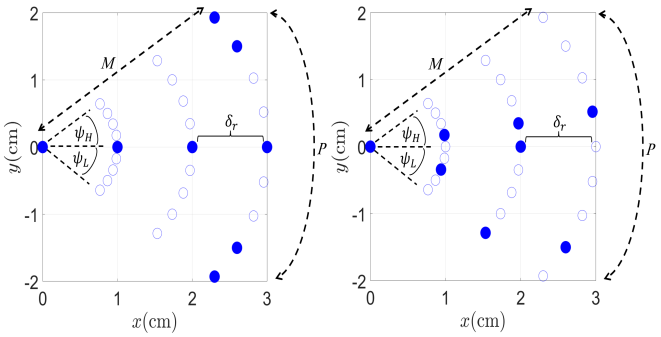

Fig. 1: Optimal layout for different ROIs. (a) $\Phi_{\mathrm{ROI}} = [-10^o, 10^o]$ and (b) $\Phi_{\mathrm{ROI}} = [-40^o, 40^o]$. Empty and solid circles indicate unoccupied and occupied microphone locations, respectively.

desired ROI, and $\Omega_{\mathrm{ROI}} = \int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}} \sin\theta \mathrm{d}\phi\mathrm{d}\theta$ is the spatial angle associated with that region.

It is evident that if the vector $\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}$ is known, (14) simplifies to the form presented in (9). It is important to note that the assumption of a uniform distribution $p_{\mathbf{d}}(\theta,\phi)$ for $\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}$ across the ROI is made in the second equality of (14). This assumption is useful in scenarios where obtaining a reliable estimation of $p_{\mathbf{d}}(\theta,\phi)$ is practically unattainable. Nevertheless, generalization to other distributions is straightforward.

In the proposed approach, as the ROI may potentially be wide, we do not require a distortionless response. Such a response could result in degradation inflicted by undesirable interferences and reverberations, particularly in silent periods. Instead, we employ the following distortion-controlled constraint

$$\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}} = 1. \tag{15}$$

Consequently, the SNR gain over the entire ROI is given by

$$\begin{aligned}\mathcal{G}_{\mathrm{ROI}}(\mathbf{f}) &= \frac{\mathrm{oSNR}_{\mathrm{ROI}}(\mathbf{f})}{\mathrm{iSNR}} \\ &= \frac{1}{\Omega^2_{\mathrm{ROI}}}\times\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{\Gamma_v}\mathbf{f}} = \frac{1}{\Omega^2_{\mathrm{ROI}}}\times\frac{1}{\mathbf{f}^H\mathbf{\Gamma_v}\mathbf{f}},\end{aligned} \tag{16}$$

which implies that the WNG over the entire ROI is obtained by

$$\mathcal{W}_{\mathrm{ROI}}(\mathbf{f}) = \frac{1}{\Omega^2_{\mathrm{ROI}}}\times\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{f}}, \tag{17}$$

and the DF over the entire ROI is

$$\mathcal{D}_{\mathrm{ROI}}(\mathbf{f}) = \frac{1}{\Omega^2_{\mathrm{ROI}}}\times\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}}. \tag{18}$$

4. JOINT OPTIMIZATION OF ARRAY GEOMETRY AND BEAMFORMING

Our study aims to propose a practical approach to beamforming that utilizes physically small arrays with a small number of microphones. To achieve this, we aim to reduce the number of microphones discussed in the previous parts associated with the UCSA structure. Instead, we will settle for a sparse circular sector array (SCSA) structure comprising a subset of the UCSA microphones. In particular, we aim to leverage the ROI formulation and present a method to jointly optimize the sparse array geometry and derive the taps of a high-directivity beamformer.

Let us dive into the technical details of the optimization function and the design constraints. To begin with, we define the ROI-oriented broadband directivity index as

$$\begin{aligned}\mathcal{DI}_{[f_L,f_H]}[\mathbf{f}] &= \frac{\int_{f_L}^{f_H}|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2 df}{\Omega^2_{\mathrm{ROI}}\int_{f_L}^{f_H}\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}df} \\ &= \frac{f_H-f_L}{\Omega^2_{\mathrm{ROI}}\int_{f_L}^{f_H}\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}df},\end{aligned} \tag{19}$$

where f_L and f_H are the minimal and maximal frequencies of interest, respectively, and the distortion-controlled constraint of (15) is assumed to hold. We aim to find the optimal sparse set of occupied microphone locations and their corresponding beamformer's taps to maximize (19). This is equivalent to solving

$$\begin{aligned}\mathbf{f}^* &= \arg\max_{\mathbf{f}}\mathcal{DI}_{[f_L,f_H]}[\mathbf{f}] \\ &= \arg\min_{\mathbf{f}}\int_{f_L}^{f_H}\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}df,\end{aligned} \tag{20}$$

where the optimal solution $\mathbf{f}^*$ is a K-sparse beamformer.

Next, we formulate the design constraints. First and foremost, we should satisfy the distortion-controlled constraint for every frequency of interest, that is,

$$\mathcal{C}_1[\mathbf{f}] : \mathbf{f}^H\mathbf{b}_{\mathrm{ROI}} = 1,\ \forall f\in[f_L,f_H]. \tag{21}$$

Then, to guarantee robustness to practical microphone imperfections, we would like to set a minimal accepted value of $\mathcal{W}_{\mathrm{ROI}}(\mathbf{f})$, denoted by ϵ, and being an application-tailored parameter. Leveraging equation (17), this implies that

$$\mathcal{C}_2[\mathbf{f}] : \mathbf{f}^H\mathbf{f} \le \frac{1}{\Omega^2_{\mathrm{ROI}}\,\epsilon},\ \forall f\in[f_L,f_H]. \tag{22}$$

Finally, we would like to ensure the K sparsity of the obtained solution. To attain that, we employ the approach taken in [25]. First, we define a binary vector $\mathbf{s} = \begin{bmatrix} S_1 & \cdots & S_{P(M-1)+1}\end{bmatrix}^T$ in which every element corresponds to a microphone location in the complete UCSA. Then, we define the following two constraints:

$$\mathcal{C}_3[\mathbf{s}] : \sum_{i=1}^{P(M-1)+1} S_i = K \tag{23}$$

which guarantees the K-sparsity attribute of $\mathbf{s}$, and

$$\begin{aligned}\mathcal{C}_4[\mathbf{f},\mathbf{s}] : |F_i|^2 \le \frac{S_i}{\Omega^2_{\mathrm{ROI}}\,\epsilon}, \\ \forall f\in[f_L,f_H],\ \forall i = 1,\ldots,P(M-1)+1,\end{aligned} \tag{24}$$

which guarantees the K-sparsity attribute of $\mathbf{f}$ without imposing further constraints on its non-zero values on top of $\mathcal{C}_2[\mathbf{f}]$. Grouping (20)-(24) together, our mixed-integer convex optimization problem is obtained as

$$\begin{aligned}\mathbf{f}_{\mathrm{SCSA}} = \arg\min_{\mathbf{f}}&\int_{f_L}^{f_H}\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}df \\ \text{s.t.}\quad &\mathcal{C}_1[\mathbf{f}],\mathcal{C}_2[\mathbf{f}],\mathcal{C}_3[\mathbf{s}],\mathcal{C}_4[\mathbf{f},\mathbf{s}],\end{aligned} \tag{25}$$

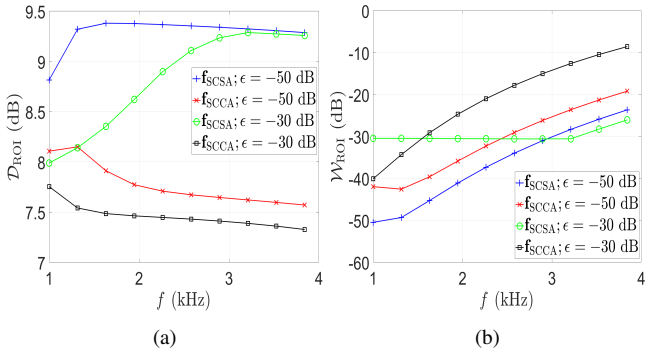

Fig. 2: DF and WNG over the entire ROI with the proposed approach $\mathbf{f}_{\mathrm{SCSA}}$ and $\mathbf{f}_{\mathrm{SCCA}}$ [27] for two values of ϵ. (a) $\mathcal{D}_{\mathrm{ROI}}$ and (b) $\mathcal{W}_{\mathrm{ROI}}$.

which may be solved using an off-the-shelf convex optimization solver as MOSEK [30]. The latter can handle binary variables using the branch-and-bound method and convex relaxation. It is important to note that this formulation exhibits fundamental advantages over previous approaches as [25–27]. Previous approaches comprised two independent procedures: optimizing the occupied microphone locations by the desired ROI and deriving a point-source-oriented beamformer accordingly. In contrast, in the proposed approach, we globally optimize over both and obtain the occupied microphone locations and their corresponding beamformer's taps simultaneously while directly accounting for the desired ROI. Additionally, due to the ROI formulation, the optimization problem we consider here is significantly reduced compared to the previous approaches, as the optimization avoids iterating over all DOAs within the ROI. This entails considerably lower computational complexity and runtime.

5. SIMULATIONS

Let us now demonstrate the performance of our proposed approach. Specifically, we set $M=4, P=9, \{\psi_0,\psi_1,\ldots,\psi_8\}=\{-40^o,-30^o,\ldots,40^o\}, \delta_r=1$ cm, focus on desired sources originating from the x-y plane, that is, $\theta_0=\pi/2$, and assume the nominal DOA of the desired source concerning the azimuth angle is 0^o. As elaborated above, this setting implies a total of 28 possible microphone locations, out of which we would like to utilize only $K=8$. To begin with, it is valuable to investigate the influence of the desired ROI on the optimal array geometry. We set $\epsilon=-50$ dB and design the proposed SCSA with $\Phi_{\mathrm{ROI}}=[-10^o,10^o]$ and $\Phi_{\mathrm{ROI}}=[-40^o,40^o]$. The results are depicted in Fig. 1 and clearly distinguish between the two ROIs. We observe that while with the narrower ROI, a symmetric geometry is exhibited, having a superdirective-like structure formed on the x-axis and pointing towards the endfire direction, with the wider ROI, the geometry is shaped differently. That is, as it is designed to accommodate a wide range of DOAs, the array geometry turns asymmetric and composed of unintuitive occupied microphone locations. This asymmetry does not necessarily imply asymmetric performance measures, as the jointly optimized beamformer may compensate for it.

Next, we focus on the more challenging case of $\Phi_{\mathrm{ROI}}=[-40^o,40^o]$ and compare the proposed approach to the sparse concentric circular array (SCCA) suggested in [27] and denoted by $\mathbf{f}_{\mathrm{SCCA}}$. Specifically, the latter is designed with a two-ring 8-sparse circular array (out of 16 uniformly-spaced possible microphone locations) and an inner-circle radius of 1 cm. We begin by evaluating

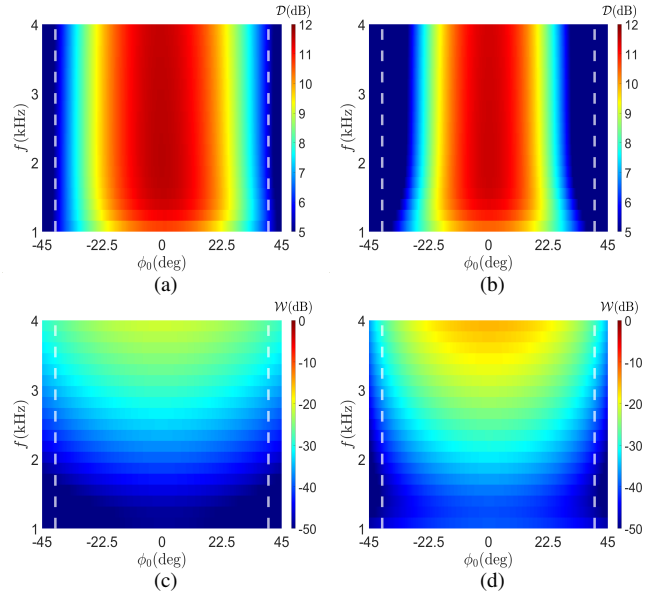

Fig. 3: DF and WNG measures with the proposed approach $\mathbf{f}_{\mathrm{SCSA}}$ and $\mathbf{f}_{\mathrm{SCCA}}$ [27]. (a) $\mathcal{D}\left(\mathbf{f}_{\mathrm{SCSA}}\right)$, (b) $\mathcal{D}\left(\mathbf{f}_{\mathrm{SCCA}}\right)$, (c) $\mathcal{W}\left(\mathbf{f}_{\mathrm{SCSA}}\right)$, and (d) $\mathcal{W}\left(\mathbf{f}_{\mathrm{SCCA}}\right)$. White vertical dashed lines indicate the boundaries of the desired ROI.

the DF and WNG over the entire ROI, which are depicted in Fig. 2 for $\epsilon=-50$ dB and $\epsilon=-30$ dB. Our proposed approach is superior in terms of $\mathcal{D}_{\mathrm{ROI}}$ for the entire frequency range considering both values of ϵ. In addition, we observe that although the proposed approach exhibits a lower value of $\mathcal{W}_{\mathrm{ROI}}$, it strictly adheres to the desired minimal value set by ϵ while globally optimizing the broadband array directivity.

Finally, in Fig. 3, we focus on the traditional DF and WNG measures, which are a function of both the frequency and the DOA. Compared to the approach taken in [27], it is clear that our approach enables tolerance to more significant DOA deviations. That is, the DF performance gap for $|\phi_0|\in[20^o,40^o]$ is substantial and is highly related to the performance gap in $\mathcal{D}_{\mathrm{ROI}}$ discussed above. In contrast, our approach exhibits a slightly inferior WNG level, particularly in high frequencies, provided that the DOA deviation is mild. We infer the proposed approach is preferable when the DOA deviation is significant, exhibiting a comparable WNG but a higher DF.

6. CONCLUSIONS

We have developed a new ROI beamforming approach, which optimizes the microphone array layout and the corresponding beamformer. Our approach considers a range of potential desired source directions rather than just one specific direction. We have solved a convex optimization problem to maximize the broadband array directivity across the entire ROI while ensuring a controlled level of desired signal distortion and sufficient WNG. Our approach is computationally efficient compared to previous methods and provides a globally optimal solution. The optimal geometry of the array layout follows the superdirective beamformer for a narrow ROI, while an asymmetric geometry is shown optimal for a wide ROI. We have compared our approach to the recent sparse concentric circular array method. We showed that our approach is superior in terms of $\mathcal{D}_{\mathrm{ROI}}$ and has notably higher directivity for significant DOA deviations.

7. REFERENCES

[1] G. W. Elko and J. Meyer, *Microphone arrays*, pp. 1021–1041, Springer Berlin Heidelberg, 2008.

[2] O. Schwartz, S. Gannot, and E. A. P. Habets, "Multi-microphone speech dereverberation and noise reduction using relative early transfer functions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 2, pp. 240–251, 2015.

[3] Z. Wang, G. Wichern, S. Watanabe, and L. R. Jonathan, "Stft-domain neural speech enhancement with very low algorithmic latency," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 397–410, 2023.

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

[5] G. Richard, P. Smaragdis, S. Gannot, P. A. Naylor, S. Makino, W. Kellermann, and A. Sugiyama, "Audio signal processing in the 21st century: The important outcomes of the past 25 years," *IEEE Signal Processing Magazine*, vol. 40, no. 5, pp. 12–26, 2023.

[6] C. Marro, Y. Mahieux, and K.U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, 1998.

[7] I. Kodrasi and S. Doclo, "Joint dereverberation and noise reduction based on acoustic multi-channel equalization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 680–693, 2016.

[8] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*, Wiley-IEEE Press, New York, 2018.

[9] W. Xiong, C. Bao, J. Zhou, M. Jia, and J. Picheral, "Joint doa estimation and dereverberation based on multi-channel linear prediction filtering and azimuth sparsity," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1481–1493, 2024.

[10] I. Kodrasi, T. Rohdenburg, and S. Doclo, "Microphone position optimization for planar superdirective beamforming," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 109–112.

[11] M. Crocco and A. Trucco, "Design of superdirective planar arrays with sparse aperiodic layouts for processing broadband signals via 3-d beamforming," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 800–815, 2014.

[12] F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "Uniform linear arrays of first-order steerable differential microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 1906–1918, 2019.

[13] G. Itzhak, J. Benesty, and I. Cohen, "On the design of differential Kronecker product beamformers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1397–1410, 2021.

[14] J. Jin, J. Benesty, G. Huang, and J. Chen, "On differential beamforming with nonuniform linear microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1840–1852, 2022.

[15] C. Pan, J. Chen, and J. Benesty, "Performance study of the mvdr beamformer as a function of the source incidence angle," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 67–79, 2014.

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

[17] S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 617–631, 2007.

[18] G. Itzhak and I. Cohen, "Differential and constant-beamwidth beamforming with uniform rectangular arrays," in *Proc. 17th International Workshop on Acoustic Signal Enhancement, IWAENC-2022*, Sep 2022.

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

[20] G. Itzhak and I. Cohen, "Differential constant-beamwidth beamforming with cube arrays," *Speech Communication*, vol. 149, pp. 98–107, 2023.

[21] G. Huang, J. Benesty, and J. Chen, "Design of robust concentric circular differential microphone arrays," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3236–3249, 2017.

[22] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 1140–1153, 2017.

[23] G. Huang, J. Chen, and J. Benesty, "Insights into frequency-invariant beamforming with concentric circular microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2305–2318, 2018.

[24] Y. Wakabayashi, K. Yamaoka, and N. Ono, "Rotation-robust beamforming based on sound field interpolation with regularly circular microphone array," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 771–775.

[25] Y. Konforti, I. Cohen, and B. Berdugo, "Array geometry optimization for region-of-interest broadband beamforming," in *Proc. 17th International Workshop on Acoustic Signal Enhancement, IWAENC-2022*, Sep 2022.

[26] G. Itzhak and I. Cohen, "Region-of-interest oriented constant-beamwidth beamforming with rectangular arrays," in *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023.

[27] G. Itzhak and I. Cohen, "Kronecker-product beamforming with sparse concentric circular arrays," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 64–72, 2024.

[28] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, Detection, Estimation, and Modulation Theory. Wiley, New York, 2004.

[29] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Simon and Schuster, Inc., USA, 1992.

[30] MOSEK ApS, "The mosek optimization toolbox for MATLAB, version 9.1," 2019.