ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICASSP55912.2026.11461491

SEQUENTIAL AND SIMULTANEOUS OPTIMIZATION OF MICROPHONE ARRAY GEOMETRY AND REGION-OF-INTEREST BEAMFORMING

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 presents a sequential approach to simultaneously optimize a microphone array geometry and its corresponding beamformer weights. The primary objective is to address the problem of an unknown direction of arrival of a desired source within a given region of interest while maximizing the broadband array directivity criterion. Optimization constraints are developed and applied to guarantee minimal distortion of the desired source and high white noise gain. The proposed approach outperforms recent and traditional approaches, considering the average white noise gain in the entire region of interest and frequency spectrum. It is also preferable in terms of the directivity factor 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, sequential optimization, sparse arrays.

1. INTRODUCTION

Beamforming has been extensively used to extract desired source signals from noisy and reverberant sensor observations [1]–[5]. Designing a beamformer involves two primary issues: the array geometry design and the derivation of the underlying beamformer weights. While much attention has been given to the latter [6]–[9], the former has been rarely addressed despite its known influence over the attributes and performance of the involved beamformer [10], [11].

Among the most standard array layouts, uniform linear arrays (ULAs), rectangular arrays (RAs), and circular arrays have been explored to the greatest extent. In particular, ULAs have been the focus of many studies due to their simplicity and ease of application. They were also shown capable of attaining high array directivity or white noise robustness, albeit not both [12], [13]. Unfortunately, they are susceptible to mismatches between the actual and nominal direction of arrival (DOA) of the desired source [14], as well as to microphone imperfections [15]. In contrast, RAs (and their three-dimensional generalization, cube arrays) exhibit reduced susceptibility to the DOA of the desired source when the DOA is parallel to one of the array axes. Nevertheless, their performance typically drops when the DOA deviates from the array axes [16]–[18]. Leveraging their azimuthal symmetry, circular arrays may ideally exhibit DOA-independent performance on the x-y plane. Unfortunately, this usually entails limited array directivity or requires accurate knowledge of the true DOA.

Recent studies have addressed DOA uncertainty by proposing the concept of a region of interest (ROI) in different domains and across a variety of applications [19]–[22]. The ROI describes the region in space from which the desired source is expected to originate. Previous works suggested optimizing a microphone array geometry either over a linear [23] or rectangular [24] layout while maximizing the array directivity and accounting for a sufficient level of WNG. Unfortunately, although shown as superior to traditional methods and enabled other desired attributes such as a constant main-lobe beamwidth, they were only adequate for narrow ROIs. They required a large array aperture along (at least) one of its axes. In [25], a sparse concentric circular array (SCCA) approach was introduced and exhibited improved performance for wider ROIs, even with a limited number of microphones. It also utilized the Kronecker-product framework [26] to easily extend the optimized geometry for larger arrays. Nonetheless, the KP-based extension and the derivation of the underlying beamformer weights were not optimal and did not directly consider the desired ROI. In [27], [28], a joint optimization approach of the array geometry and weights was proposed. This approach exploited sparse circular sector arrays to directly account for the desired ROI, and was shown to more accurately align with it compared to previous work. However, it could not address large arrays involving many microphones, as it was designed to accommodate them simultaneously.

This paper presents a sequential approach to simultaneously optimize the microphone array geometry and beamformer weights. We assume the desired source impinges on the array from an unknown DOA within a given ROI and attempt to optimize the broadband array directivity accordingly. We ensure an acceptable lower threshold of the white noise gain (WNG) and a distortion-controlled response of the desired source. We show that the proposed approach is preferable in terms of the directivity factor (DF) when significant DOA deviations are considered and in terms of the WNG across all frequencies and directions considered.

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 sequentially optimize the array geometry and the beamformer weights at once. In Section 5, we analyze the performance of the proposed approach and compare it to existing methods.

2. SIGNAL MODEL

Consider a desired source signal propagating from the farfield in an anechoic acoustic environment at the speed of sound, i.e., $c=340$ m/s, in an elevation angle θ and an azimuth angle ϕ. The plane wave impinges on a uniform concentric circular array (UCCA) composed of N equally spaced rings whose corresponding radii are R,$2R$,...,NR, with R being the radius of the innermost ring. The array is located on the x-y plane and exhibits P equally spaced and omnidirectional microphones on each ring. Accounting for an additional microphone located in the center of the UCCA, the total number of array microphones is $M = NP + 1$. Defining the center of the UCCA as the origin of the Polar coordinate system as well as the reference point, the array steering vector of the n-th ring is given by [29]:

$$\mathbf{d}_{n;\theta,\phi}(f) = \Big[\; e^{\jmath n \frac{2\pi f R}{c}\cos(\phi-\psi_1)\sin\theta} \quad e^{\jmath n \frac{2\pi f R}{c}\cos(\phi-\psi_2)\sin\theta} \quad \dots \quad e^{\jmath n \frac{2\pi f R}{c}\cos(\phi-\psi_P)\sin\theta} \;\Big]^T, \tag{1}$$

where $n = 1, \dots, N$ is the ring index, the superscript ^{T} denotes the transpose operator, $\jmath = \sqrt{-1}$ is the imaginary unit, $f > 0$ is the temporal frequency and

$$\psi_p = \frac{2\pi(p-1)}{P} \tag{2}$$

is the angular distance between the p-th microphone on a ring ($p=1,\dots,P$) and the positive x-axis direction. Stacking the steering vectors of all rings, we obtain the full steering vector of the UCCA:

$$\mathbf{d}_{\theta,\phi}(f) = \Big[\; 1 \quad \mathbf{d}^T_{1;\theta,\phi}(f) \quad \mathbf{d}^T_{2;\theta,\phi}(f) \quad \cdots \quad \mathbf{d}^T_{N;\theta,\phi}(f) \;\Big]^T, \tag{3}$$

in which the first element corresponds to the reference microphone located at the origin.

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

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

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

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

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

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

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

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

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 $\mathbf{f}$ is applied to the observed signal vector $\mathbf{y}$, yielding the output signal [30]

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

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 WNG, and the 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

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

where $\mathbf{\Gamma}_{\mathrm{d}}$ is the pseudo-coherence matrix of a spherically isotropic (diffuse) noise field [30], 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. As a result, the performance measures in (9)-(11) may be inadequate to derive $\mathbf{f}$. To cope with such scenarios, we shall take an alternative approach and consider the steering vector associated with the true DOA to be unknown. That is, we define $\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}$ to be a uniformly distributed random variable over the entire ROI. As a consequence, we may define

$$\bar{\mathbf{x}}=XE[\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}]=\frac{X}{\Omega_{\mathrm{ROI}}}\iint_{\substack{\theta\in\Theta_{\mathrm{ROI}}\\ \phi\in\Phi_{\mathrm{ROI}}}}\mathbf{d}_{\theta,\phi}\sin\theta\mathrm{d}\phi\mathrm{d}\theta, \tag{12}$$

which constitutes the average steering vector of all possible DOAs within the ROI: $\theta\in\Theta_{\mathrm{ROI}}, \phi\in\Phi_{\mathrm{ROI}}$, whereas Ω_{ROI} is given by $\Omega_{\mathrm{ROI}}=\int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}}\sin\theta\mathrm{d}\phi\mathrm{d}\theta$. Note that this aggregation assumes complete coherence between all directions within the ROI in a similar manner to [27] and the coherently scattered source scenario examined in [31]. This assumption is valid as we consider a single desired source whose DOA is unknown (rather than a range of desired scattered sources), and hence no practical synchronization is required.

Substituting (12) into (9), we may define the output SNR over the entire ROI as

$$\begin{aligned}\mathrm{oSNR}_{\mathrm{ROI}}(\mathbf{f})&=\frac{1}{p_V}\times\frac{E\left[|\mathbf{f}^H\bar{\mathbf{x}}|^2\right]}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}\\ &=\frac{p_X}{p_V}\times\frac{1}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}\times\frac{1}{\Omega_{\mathrm{ROI}}^2}\\ &\quad\times\left|\mathbf{f}^H\iint_{\substack{\theta\in\Theta_{\mathrm{ROI}}\\ \phi\in\Phi_{\mathrm{ROI}}}}\mathbf{d}_{\theta,\phi}\sin\theta\mathrm{d}\phi\mathrm{d}\theta\right|^2\\ &=\frac{p_X}{p_V}\times\frac{1}{\Omega_{\mathrm{ROI}}^2}\times\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}},\end{aligned} \tag{13}$$

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 desired ROI. Note that in case the vector $\mathbf{d}_{\theta_{\mathrm{ROI}},\phi_{\mathrm{ROI}}}$ is known, (13) simplifies to the form presented in (9). Notice that there may be different ways to formulate (13), potentially avoiding the inner-ROI coherence assumption stated above. Nevertheless, this would typically entail computationally inefficient derivations (e.g., involving multiple eigenvalue decompositions) which would not be adequate for the iterative optimization procedure presented in this paper.

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:

$$\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}=1. \tag{14}$$

Consequently, in case (14) holds, the SNR gain over the entire ROI is given by

$$\begin{aligned}\mathcal{G}_{\mathrm{ROI}}(\mathbf{f})&=\frac{\mathrm{oSNR}_{\mathrm{ROI}}(\mathbf{f})}{\mathrm{iSNR}}\\ &=\frac{1}{\Omega_{\mathrm{ROI}}^2}\times\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}=\frac{1}{\Omega_{\mathrm{ROI}}^2}\times\frac{1}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}},\end{aligned} \tag{15}$$

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

$$\mathcal{W}_{\mathrm{ROI}}(\mathbf{f})=\frac{1}{\Omega_{\mathrm{ROI}}^2}\times\frac{\left|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}\right|^2}{\mathbf{f}^H\mathbf{f}}, \tag{16}$$

and the DF over the entire ROI is

$$\mathcal{D}_{\mathrm{ROI}}(\mathbf{f})=\frac{1}{\Omega_{\mathrm{ROI}}^2}\times\frac{\left|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}\right|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}}. \tag{17}$$

4. SEQUENTIAL AND SIMULTANEOUS OPTIMIZATION

Our primary objective is to optimize the microphone array geometry and beamformer weights concerning a given ROI, even when large arrays with many microphones are considered. First, it is crucial to establish a proper optimization criterion and design constraints for the array's design. As we are interested in designing high directivity beamformers and as speech signals are broadband, we focus on the ROI-oriented broadband directivity index defined as [27]

$$\mathcal{DI}_{[f_L,f_H]}[\mathbf{f}]=\frac{\int_{f_L}^{f_H}\left|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}\right|^2df}{\Omega_{\mathrm{ROI}}^2\int_{f_L}^{f_H}\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}df}, \tag{18}$$

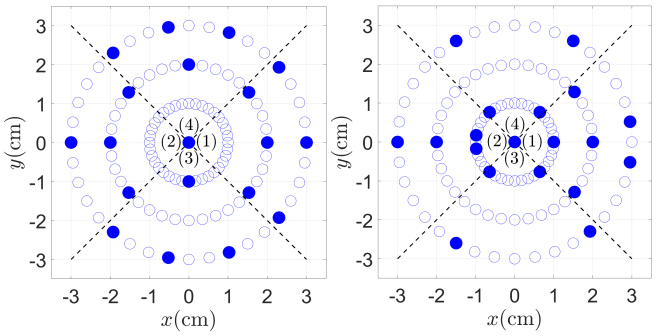

Fig. 1: Optimal layout for different ROIs. (a) $\Phi_{\text{ROI}}=[-10^o,10^o]$ and (b) $\Phi_{\text{ROI}}=[-40^o,40^o]$. Empty and solid circles indicate unoccupied and occupied microphone locations, respectively. The dashed lines indicate the division into four identical sub-arrays whose optimization sequence indices are marked in brackets.

in which the minimal and maximal frequencies of interest are denoted by f_L and f_H, respectively. Then, in case only a limited set of possible microphone locations is considered, the optimal solution may be obtained by maximizing (18):

$$\mathbf{f}^*=\underset{\mathbf{f}}{\text{argmax}}\mathcal{DI}_{[f_L,f_H]}[\mathbf{f}]. \tag{19}$$

in which the optimal solution $\mathbf{f}^*$ is a K-sparse beamformer, with K being a design parameter indicating the desired number of occupied microphone locations out of the total number of locations M. However, as shown in [27], a proper solution of (19) must be subject to performance and structural constraints that are not all convex. Specifically, the optimization problem may be classified as a mixed-integer programming (MIP) problem known as NP-hard. This implies that if either K or M are not low-valued, the associated computational complexity is immense and therefore impractical, even for offline processing.

To handle large arrays, we propose a sequential optimization approach that rigorously considers all possible microphone locations on the array but not all at once. We split the entire array layout (e.g., a complete UCCA) into smaller sub-arrays (e.g., circular sector arrays) and simultaneously optimize the array geometry and beamformer weights, considering all previously optimized occupied microphone locations. Specifically, let $t=1,2,...,T$ denote the optimization sequence index (or stage), and $M_1,M_2,...,M_T$ and $K_1,K_2,...,K_T$ denote the total number of microphone locations and desired number of occupied microphone locations for each sub-array, respectively. Then, we define:

$$\bar{K}_t=1+\sum_{l=1}^{t-1}K_l; \qquad \bar{M}_t=\bar{K}_t+M_t. \tag{20}$$

in which the additive '1' term stands for the reference microphone location in the origin that is not considered a part of any sub-array and is always marked as occupied.

Leveraging the perspective of split arrays, the single optimization problem in (19) is transformed into a series of optimization problems in the form

$$\mathbf{f}^*_{(t)}=\underset{\mathbf{f}}{\text{argmin}}\int_{f_L}^{f_H}\mathbf{f}^H_{(t)}\mathbf{\Gamma}^{(t)}_{\text{d}}\mathbf{f}_{(t)}df, \tag{21}$$

where $\mathbf{\Gamma}^{(t)}_{\text{d}}$ is a $\bar{M}_t\times\bar{M}_t$ diffuse noise pseudo-coherence matrix at stage t, taking into account all optimized microphone locations up to and including stage $t-1$, along with all potential microphone locations evaluated in stage t. Notice that this formulation assumes that the distortion-controlled constraint of (14) holds.

Each referred optimization problem must be accompanied by proper design constraints to yield a valid solution. The distortion-controlled constraint must be satisfied for every optimization stage t and frequency of interest f, that is,

$$\mathcal{C}_1\left[\mathbf{f}_{(t)}\right]:\mathbf{f}^H_{(t)}\mathbf{b}^{(t)}_{\text{ROI}}=1,\ \forall f\in[f_L,f_H], \tag{22}$$

where $\mathbf{b}^{(t)}_{\text{ROI}}=\int_{\theta\in\Theta_{\text{ROI}}}\int_{\phi\in\Phi_{\text{ROI}}}\mathbf{d}^{(t)}_{\theta,\phi}\sin\theta\text{d}\phi\text{d}\theta$, and $\mathbf{d}^{(t)}_{\theta,\phi}$ is a steering vector of length $\bar{M}_t$ in stage t constructed in a similar fashion to $\mathbf{\Gamma}^{(t)}_{\text{d}}$. In addition, to provide robustness to practical microphone imperfections, a minimal accepted value of $\mathcal{W}_{\text{ROI}}(\mathbf{f})$ is set. Denoted by ϵ, this design parameter should be selected by the application and microphones at hand. Considering equation (16), we have

$$\mathcal{C}_2\left[\mathbf{f}_{(t)}\right]:\mathbf{f}^H_{(t)}\mathbf{f}_{(t)}\le\frac{1}{\Omega^2_{\text{ROI}}\,\epsilon},\ \forall f\in[f_L,f_H]. \tag{23}$$

Next, we would like to ensure the sparsity of the obtained solution. First, We decompose the optimized beamformer into $\mathbf{f}_{(t)}=\left[\begin{array}{cc}\mathbf{f}^{(t)^T}_1 & \mathbf{f}^{(t)^T}_2\end{array}\right]^T$, in which $\mathbf{f}^{(t)}_1$ is the part of $\mathbf{f}_{(t)}$ corresponding to the previously optimized microphone locations and $\mathbf{f}^{(t)}_2$ is the part of $\mathbf{f}_{(t)}$ corresponding to the remaining M_t locations evaluated in stage t. Then, to guarantee the selection of all previously optimized microphone locations, we require

$$\begin{gathered}\mathcal{C}_3\left[\mathbf{f}^{(t)}_1\right]:\left|F^{(t)}_{1;i}\right|^2\le\frac{1}{\Omega^2_{\text{ROI}}\,\epsilon},\\ \forall f\in[f_L,f_H],\ \forall i=1,...,\bar{K}_t,\end{gathered} \tag{24}$$

with $\{F^{(t)}_{1;i}\}_{i=1}^{\bar{K}_t}$ denoting the elements of $\mathbf{f}^{(t)}_1$. Finally, we want to ensure that precisely K_t locations are optimized and selected in stage t. We define the binary vector $\mathbf{s}_{(t)}=\left[\begin{array}{ccc}S^{(t)}_1 & \cdots & S^{(t)}_{M_t}\end{array}\right]^T$ in which every element corresponds to a potential microphone location of the t-th sub-array evaluated in stage t and set the following two constraints:

$$\mathcal{C}_4\left[\mathbf{s}_{(t)}\right]:\sum_{i=1}^{M_t}S^{(t)}_i=K_t, \tag{25}$$

$$\begin{gathered}\mathcal{C}_5\left[\mathbf{f}^{(t)}_2,\mathbf{s}_{(t)}\right]:\left|F^{(t)}_{2;i}\right|^2\le\frac{S^{(t)}_i}{\Omega^2_{\text{ROI}}\,\epsilon},\\ \forall f\in[f_L,f_H],\ \forall i=1,...,M_t,\end{gathered} \tag{26}$$

with $\{F^{(t)}_{2;i}\}_{i=1}^{M_t}$ denoting the elements of $\mathbf{f}^{(t)}_2$. Note that constraints $\mathcal{C}_3\left[\mathbf{f}^{(t)}_1\right]$, $\mathcal{C}_4\left[\mathbf{s}_{(t)}\right]$, and $\mathcal{C}_5\left[\mathbf{f}^{(t)}_2,\mathbf{s}_{(t)}\right]$ guarantee the $\bar{K}_{t+1}$-sparsity of $\mathbf{f}_{(t)}$ without imposing further constraints on its non-zero values on top of $\mathcal{C}_2\left[\mathbf{f}_{(t)}\right]$.

Grouping (21)-(26) together, the sequentially optimized SCCA (SO-SCCA) beamformer of stage t is obtained by solving the following MIP optimization problem

$$\begin{aligned}\mathbf{f}^{(t)}_{\text{SO-SCCA}}=&\underset{\mathbf{f}}{\text{argmin}}\int_{f_L}^{f_H}\mathbf{f}^H_{(t)}\mathbf{\Gamma}^{(t)}_{\text{d}}\mathbf{f}_{(t)}df\\ \text{s.t.}\quad&\mathcal{C}_1\left[\mathbf{f}_{(t)}\right],\mathcal{C}_2\left[\mathbf{f}_{(t)}\right],\mathcal{C}_3\left[\mathbf{f}^{(t)}_1\right],\mathcal{C}_4\left[\mathbf{s}_{(t)}\right],\mathcal{C}_5\left[\mathbf{f}^{(t)}_2,\mathbf{s}_{(t)}\right],\end{aligned} \tag{27}$$

which we solve using the MOSEK optimization solver [32] by employing the branch-and-bound and convex relaxation methods. Ultimately, when the sequential procedure is completed, and all sub-arrays are optimized, we obtain our proposed SO-SCCA beamformer $\mathbf{f}_{\text{SO-SCCA}}=\mathbf{f}^{(T)}_{\text{SO-SCCA}}$ which is of length $K=\bar{K}_{T+1}$.

Finally, it is important to note that the sequential aspect of the optimization procedure is confined exclusively to the microphone placement, while the beamformer weights are jointly set in the final stage and hence remain globally optimized.

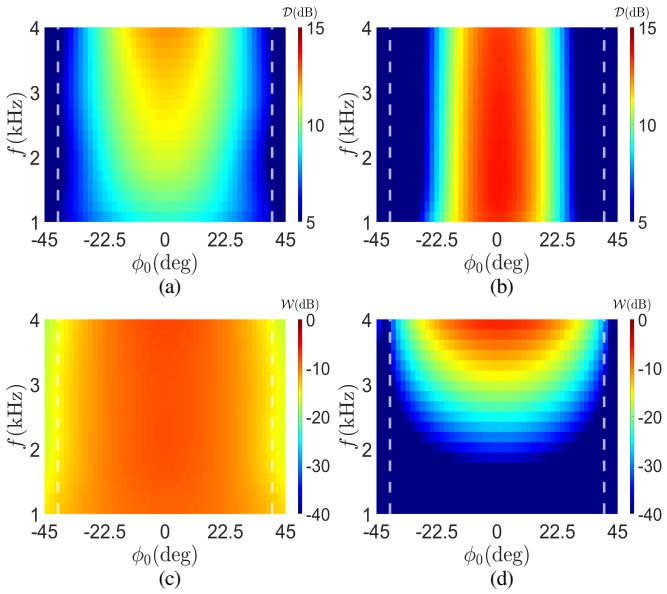

Fig. 2: DF and WNG obtained with the proposed $\mathbf{f}_{\mathrm{SO-SCCA}}$ and $\mathbf{f}_{\mathrm{SCCA}}$ of [25]. (a) $\mathcal{D}(\mathbf{f}_{\mathrm{SO-SCCA}})$, (b) $\mathcal{D}(\mathbf{f}_{\mathrm{SCCA}})$, (c) $\mathcal{W}(\mathbf{f}_{\mathrm{SO-SCCA}})$, and (d) $\mathcal{W}(\mathbf{f}_{\mathrm{SCCA}})$.

5. EXPERIMENTAL RESULTS

To demonstrate the advantages of the proposed method, we focus on a UCCA of $N=3$ rings and $P=36$ equally spaced possible microphone locations per ring with the innermost ring's radius set to $R=1$ cm. We evenly split the UCCA into $T=4$ similar-size circular-sector-shaped sub-arrays to which we sequentially apply the proposed optimization scheme, implying that $M_1 = M_2 = M_3 = M_4 = 27$ and a total of $M = 109$ possible microphone locations. The proposed SO-SCCA beamformer is designed considering desired sources originating from the x-y plane, that is, $\theta_0 = \pi/2$, and two distinct symmetric ROIs in terms of the azimuth angle around 0^o: $\Phi_{\mathrm{ROI}} = [-10^o, 10^o]$ and $\Phi_{\mathrm{ROI}} = [-40^o, 40^o]$. In addition, we set $\epsilon = -10$ dB, $K_1 = 6$ and $K_2 = K_3 = K_4 = 4$, yielding proposed beamformers of length $K = 19$.

We begin by investigating the influence of the desired ROI on the optimal array geometry as depicted in Fig. 1. We observe that for both ROIs, a symmetric geometry is obtained with a high density of occupied microphone locations along the x-axis designed to maximize the array directivity around $\phi_0 = 0$. In contrast, with the narrower ROI, the occupied locations are more widely spread to increase the array aperture and the resulting directivity, whereas with the broader ROI a more compact structure is required to enable a distortion-controlled response across the entire region.

Next, we focus on the $\Phi_{\mathrm{ROI}} = [-40^o, 40^o]$ case and compare the proposed approach to two maximum directivity factor (MDF) beamformers, each comprising of 21 microphones: the MDF/MDF beamformer with the SCCA approach [25] denoted by $\mathbf{f}_{\mathrm{SCCA}}$, and an MDF beamformer [8] applied to a UCCA geometry denoted by $\mathbf{f}_{\mathrm{UCCA}}$. The former is designed using 3 replicas of a 2-ring SCCA comprising 7 optimized microphone locations out of possible 18 equally spaced locations. The latter involves a single 3-ring complete UCCA with 7 equally spaced microphones per ring.

Figure 2 depicts the DF and WNG of the proposed approach $\mathbf{f}_{\mathrm{SO-SCCA}}$ compared to $\mathbf{f}_{\mathrm{SCCA}}$. Regarding the DF, the proposed approach is preferable for significant DOA deviations, that is, when $|\phi_0| \in [20^o, 40^o]$. Considering the WNG, it is evident that the proposed approach is superior across the entire ROI and frequency spectrum, except for small DOA deviations in the high-frequency range where the two methods perform equally well. Ultimately, Fig. 3 compares the DF and WNG over the entire ROI of the three methods. The tradeoff between the ROI's width and $\mathcal{D}_{\mathrm{ROI}}$ is demonstrated, along with a preferable performance of the proposed $\mathbf{f}_{\mathrm{SO-SCCA}}$ for frequencies over 2 kHz.

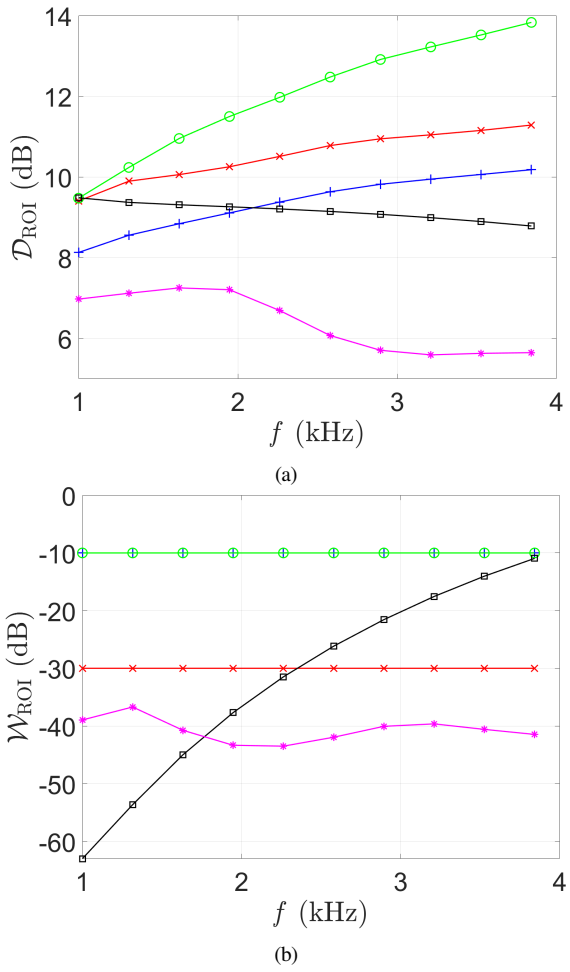

Fig. 3: (a) $\mathcal{D}_{\mathrm{ROI}}$ and (b) $\mathcal{W}_{\mathrm{ROI}}$, obtained with the proposed $\mathbf{f}_{\mathrm{SO-SCCA}}$ and existing methods for several design parameters.

Furthermore, $\mathbf{f}_{\mathrm{SO-SCCA}}$ is substantially advantageous in terms of $\mathcal{W}_{\mathrm{ROI}}$ across the entire spectrum.

6. CONCLUSIONS

We have developed an ROI beamforming approach that optimizes the microphone array geometry and the corresponding beamformer weights simultaneously. This approach involves solving a series of small MIP problems to achieve an optimal beamforming design, even with large arrays. Our focus is on maximizing the broadband array directivity criterion while considering the entire ROI and maintaining a controlled level of desired source signal distortion and sufficient WNG. We compared the performance of our approach to the recently introduced SCCA and traditional UCCA high-directivity beamformer. Our findings demonstrate that our approach is more favorable in terms of array directivity measures when significant DOA deviations are considered or in high frequencies when the entire ROI is taken into account. Additionally, our approach was found to be significantly superior in terms of WNG measures across the entire ROI and frequency spectrum.

In our future work, we will explore and refine how the microphone array is divided into sub-arrays. We will also consider the number of microphone locations being assessed and the specific number of locations we want in each sub-array.

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 *Proc. 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, 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.

[15] 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.

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

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

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

[19] A. Nemirovsky, G. Itzhak, and I. Cohen, "A robust bilinear framework for real-time speech separation and dereverberation in wearable augmented reality," *Sensors*, vol. 25, no. 17, 2025.

[20] A. Frank and I. Cohen, "Least-distortion maximum gain beamformer for time-domain region-of-interest beamforming," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 2286–2301, 2025.

[21] A. Frank and I. Cohen, "Low-latency audio front-end region-of-interest beamforming for smart glasses," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2026.

[22] J. Subhash, A. frank, I. Cohen, and N. V. George, "Generalized sidelobe canceller for time-domain region-of-interest beamforming," *to appear in IEEE Transactions on Audio, Speech and Language Processing*.

[23] 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)*, Sep 2022.

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

[25] G. Itzhak and I. Cohen, "Kronecker-product beamforming with sparse concentric circular arrays," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 64–72, 2023.

[26] J. Benesty, I. Cohen, and J. Chen, *Array Processing - Kronecker Product Beamforming*, Springer-Verlag, Switzerland, 2019.

[27] G. Itzhak, Simon Doclo, and I. Cohen, "Joint optimization of microphone array geometry and region-of-interest beamforming with sparse circular sector arrays," in *Proc. 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep 2024.

[28] G. Itzhak, S. Doclo, and I. Cohen, "Optimal region-of-interest beamforming for audio conferencing with dual perpendicular sparse circular sectors," in *Proc. 2025 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2025.

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

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

[31] 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 Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 842–852, 2019.

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