2025 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) | 979-8-3315-3745-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/WASPAA66052.2025.11231016

Optimal Region-of-Interest Beamforming for Audio Conferencing with Dual Perpendicular Sparse Circular Sectors

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

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

***Abstract*—We introduce a region-of-interest beamforming approach for audio conferencing that addresses dynamic acoustics and multiple-speaker scenarios. The approach employs a two-stage sparse optimization to select a subset of microphones from dual circular sector arrays: first on the xz plane and then on the xy plane, balancing spatial resolution and efficiency. Using the dual circular layout, we are able to reduce response variability across azimuth and elevation angles. The proposed approach maximizes broadband directivity while ensuring a controlled level of distortion and minimal white noise gain. Compared to existing methods, the mainlobe attained by the resulting beamformer is more accurately aligned with the region of interest. It also achieves a preferable sidelobe and backlobe suppression. Finally, the proposed approach is shown to be superior considering the directivity factor and white noise gain, in particular at medium and high frequencies.**

1. INTRODUCTION

Beamforming is a widely used technique for extracting desired source signals from noisy and reverberant environments, playing a crucial role in applications such as speech enhancement, teleconferencing, and spatial audio processing [1]–[5]. Extensive research has focused on optimizing beamformer weights [6]–[9], whereas the influence of the array geometry— despite its crucial role in directivity, robustness, and noise suppression— has comparatively received less attention [10], [11].

Conventional array geometries, such as uniform linear arrays (ULAs), rectangular arrays (RAs), and circular arrays, have been extensively studied. ULAs are particularly popular due to their simplicity and their ability to achieve either high array directivity or strong white noise robustness—but not both simultaneously [12], [13]. However, ULAs are highly sensitive to errors in the assumed direction of arrival (DOA) of the desired source [14] and are vulnerable to microphone imperfections [15]. RAs and their three-dimensional (3-D) counterpart, cube arrays, mitigate DOA sensitivity along their principal axes, but their performance deteriorates when the source deviates from these axes [16]–[18]. Circular arrays, due to their angular symmetry, offer the potential for DOA-independent performance considering the plane on which they are positioned, but often suffer from reduced directivity unless the true DOA is precisely known.

To address the challenges of DOA uncertainty, recent approaches have introduced the concept of a region of interest (ROI), which defines the spatial region from which the desired source is expected to originate. Optimizing microphone array geometries over an ROI can enhance spatial filtering capabilities without relying on precise DOA estimates. Previous studies have explored array optimization for linear [19] and rectangular [20] geometries, focusing on maximizing array directivity while maintaining adequate white noise gain (WNG). Although these methods outperform traditional approaches and enable desirable features such as a constant mainlobe beamwidth, they are limited to narrow ROIs and generally require a large array aperture along at least one axis. In [21], a sparse concentric circular array (SCCA) approach was proposed, demonstrating enhanced performance for wider ROIs even with a limited number of microphones.

In [22], an approach to jointly optimize the array geometry and beamformer weights was introduced. This approach utilized sparse circular sector arrays (SCSAs) to directly account for the desired ROI and demonstrated superior performance over previous methods, particularly in scenarios with significant DOA deviations. However, its planar geometry enforced a symmetric spatial response, making it unsuitable for asymmetric ROIs in the elevation direction.

In [23], a robust beamforming method for time-domain audio conferencing was proposed and demonstrated to be effective when applied to linear arrays, particularly in scenarios involving small ROIs. In [24], [25], a distinctive 3-D array configuration was presented, combining a uniform concentric circular array (UCCA) on the xy plane with a ULA aligned along the z-axis. This layout proved advantageous in dynamic scenarios and when accounting for multiple desired speakers. Nevertheless, it exhibited high sensitivity to variations in the vertical location of the ROI, indicating that the horizontal placement of the microphone array influences the array's directivity. Additionally, it did not involve explicit optimization of the array geometry, but instead relied on a potentially azimuth-angle-independent geometry.

In this work, we propose an ROI beamforming approach utilizing dual circular sectors for audio conferencing: one located on the xz plane and the other on the xy plane. Compared to the configuration in [24], [25], this layout enables lower variability for signals originating from distinct elevation directions. Optimizing a subset of array microphones from many possible locations, the proposed approach aims to maximize the broadband array directivity while jointly determining the beamformer weights, subject to design constraints that maintain controlled distortion and minimal WNG. Compared to existing methods, the resulting beamformer better aligns with the ROI, considering its mainlobe, whereas the sidelobes and backlobe are more suppressed. It also attains superior directivity across the entire spectrum and maintains a favorable WNG at medium and high frequencies.

2. SIGNAL MODEL

Consider a desired source signal that is associated with a speaker in a conference room propagating from the farfield in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, impinging on a two-dimensional (2-D) uniform circular sector array (UCSA) located on the xz plane from an elevation angle θ_0 and an azimuth angle ϕ_0. The UCSA is composed of M_{xz} uniformly-spaced omnidirectional microphones along the radial direction with an interelement spacing δ_{xz} and Q uniformly-spaced omnidirectional microphones along the elevation angle direction. The locations of the latter are lower and upper bounded by γ_L and γ_H, respectively, and are given by

$$\gamma_q = q \times \frac{\gamma_H - \gamma_L}{Q-1} + \gamma_L, \tag{1}$$

with $q = 0, \ldots, Q-1$. Considering the center of the circle as the reference point, the array steering vector associated with γ_q of length M_{xz} is given by [26]

$$\mathbf{a}^{\mathsf{xz}}_{\theta_0,\phi_0,\gamma_q}(f) = \Big[\; e^{\jmath 2\pi \delta_{\mathsf{xz}} f (\sin\gamma_q \sin\theta_0 \cos\phi_0 + \cos\gamma_q \cos\theta_0)/c} \quad \cdots \quad e^{\jmath 2\pi \delta_{\mathsf{xz}} M_{\mathsf{xz}} f (\sin\gamma_q \sin\theta_0 \cos\phi_0 + \cos\gamma_q \cos\theta_0)/c} \;\Big]^T, \tag{2}$$

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 $\{\mathbf{a}^{\mathsf{xz}}_{\theta_0,\phi_0,\gamma_q}(f)\}_{q=0}^{Q-1}$ we obtain the array steering vector of length $QM_{\mathsf{xz}}+1$ corresponding to the circular sector on the xz plane:

$$\mathbf{a}^{\mathsf{xz}}_{\theta_0,\phi_0}(f)=\left[\begin{array}{cccc} 1 & \mathbf{a}^{\mathsf{xz}}_{\theta_0,\phi_0,\gamma_0}{}^T(f) & \cdots & \mathbf{a}^{\mathsf{xz}}_{\theta_0,\phi_0,\gamma_{Q-1}}{}^T(f) \end{array}\right]^T, \quad (3)$$

where the first element refers to an additional microphone located at the reference point (which also serves as the origin of the coordinate system). Note that this resembles a truncated UCCA geometry located on the xz plane for which all microphones outside the $[\gamma_0, \gamma_{Q-1}]$ range have been eliminated.

In addition to the UCSA on the xz plane, consider a UCSA on the xy plane that is composed of M_{xy} uniformly-spaced omnidirectional microphones along the radial direction and P uniformly-spaced omnidirectional microphones along the azimuth angle direction. The array steering vector associated with the azimuth angle ψ_p of length M_{xy} is given by

$$\mathbf{a}^{\mathsf{xy}}_{\theta_0,\phi_0,\psi_p}(f)=\Big[\begin{array}{ccc} e^{\jmath 2\pi f \delta_{\mathsf{xy}} \sin\theta_0 \cos(\phi_0-\psi_p)/c} & \cdots & e^{\jmath 2\pi f \delta_{\mathsf{xy}} M_{\mathsf{xz}} \sin\theta_0 \cos(\phi_0-\psi_p)/c} \end{array}\Big]^T, \quad (4)$$

where δ_{xy} denotes the interelement spacing for this UCSA, and we have assumed its reference point to be at the center of its corresponding circle. Note that this formulation assumes the centers of both UCSAs to coincide. Similarly to (3), the array steering vector of length PM_{xy} that corresponds to the circular sector on the xz plane is given by:

$$\mathbf{a}^{\mathsf{xy}}_{\theta_0,\phi_0}(f)=\left[\begin{array}{ccc} \mathbf{a}^{\mathsf{xy}}_{\theta_0,\phi_0,\psi_0}{}^T(f) & \cdots & \mathbf{a}^{\mathsf{xy}}_{\theta_0,\phi_0,\psi_{P-1}}{}^T(f) \end{array}\right]^T. \quad (5)$$

Stacking the steering vectors of both UCSAs in (3) and (5), the complete steering vector is given by:

$$\mathbf{d}_{\theta_0,\phi_0}(f)=\left[\begin{array}{cc} \mathbf{a}^{\mathsf{xz}}_{\theta_0,\phi_0}{}^T(f) & \mathbf{a}^{\mathsf{xy}}_{\theta_0,\phi_0}{}^T(f) \end{array}\right]^T. \quad (6)$$

An illustration of the proposed array layout as well as the discussed structural parameters is depicted in Fig. 1 by considering all empty and solid circles therein.

Considering (6), the observed noisy signal vector of length $M = QM_{\mathsf{xz}}+PM_{\mathsf{xy}}+1$ can be expressed in the frequency domain as [8]:

$$\mathbf{y}(f)=\mathbf{x}(f)+\mathbf{v}(f)=\mathbf{d}_{\theta_0,\phi_0}(f)X(f)+\mathbf{v}(f), \quad (7)$$

where $X(f)$ and $\mathbf{v}(f)$ are the (zero-mean) desired source signal and the noise vector, respectively, as received by the reference microphone. Dropping the dependence on f and assuming $\mathbf{x}$ and $\mathbf{v}$ to be uncorrelated, the noisy correlation matrix is given by

$$\mathbf{\Phi}_{\mathbf{y}}=E\left(\mathbf{y}\mathbf{y}^H\right)=p_X\mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0}+\mathbf{\Phi}_{\mathbf{v}}, \quad (8)$$

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 (PSD) of the desired source at the reference microphone, and $\mathbf{\Phi}_{\mathbf{v}}=E\left(\mathbf{v}\mathbf{v}^H\right)$ is the noise correlation matrix. Assuming a compact array such that the noise variance is approximately equal for all sensors, equation (8) may be expressed as

$$\mathbf{\Phi}_{\mathbf{y}}=p_X\mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0}+p_V\mathbf{\Gamma}_{\mathbf{v}}, \quad (9)$$

where p_V is the noise PSD at the reference microphone and $\mathbf{\Gamma}_{\mathbf{v}}=\mathbf{\Phi}_{\mathbf{v}}/p_V$ is the pseudo-coherence matrix of the noise. From (9), we deduce that the input signal-to-noise ratio (SNR) is

$$\mathrm{iSNR}=\frac{\mathrm{tr}\left(p_X\mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0}\right)}{\mathrm{tr}(p_V\mathbf{\Gamma}_{\mathbf{v}})}=\frac{p_X}{p_V}, \quad (10)$$

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

3. REGION-OF-INTEREST BEAMFORMING

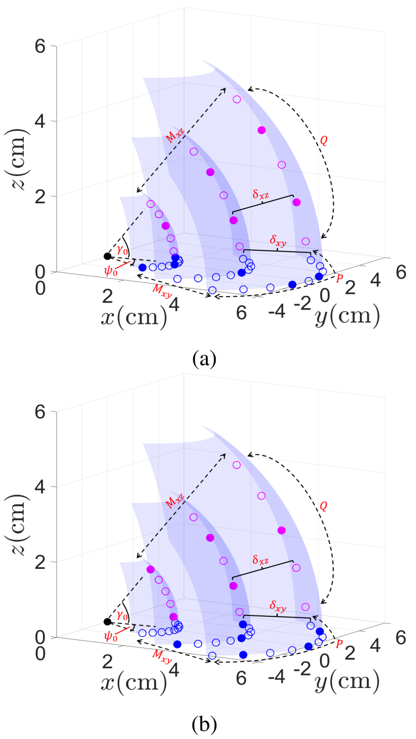
Conventionally, assuming $\mathbf{d}_{\theta_0,\phi_0}$ is known, a linear beamformer $\mathbf{f}$ is applied to the observed signal vector $\mathbf{y}$ in order to generate an estimate of the desired source X [27]:

$$\hat{X}=\mathbf{f}^H\mathbf{y}=X\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}+\mathbf{f}^H\mathbf{v}. \quad (11)$$

Hence, the output SNR is equal to

$$\mathrm{oSNR}(\mathbf{f})=\frac{p_X}{p_V}\times\frac{|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}. \quad (12)$$

While this expression may hold in relatively static (slowly varying) scenarios and assuming a single desired source, it is inapplicable for dynamic acoustics scenarios. For example, consider a conference meeting scenario, in which multiple speakers sit around a table and at least one of the speakers is physically moving while actively presenting. Very often in such scenarios, a communication device is placed on the conference table to allow sharing the conference with remote participants or recording the meeting for future use. To address scenarios of this kind, the output SNR over the entire ROI was considered [24]:



(a)

(b)

Fig. 1: Optimal beamformers' layout. (a) $\mathbf{f}_{\mathrm{DCS}}$ and (b) $\mathbf{f}_{\mathrm{FT-DCS}}$. Empty and solid circles indicate unoccupied and occupied microphone locations, respectively. Color coding refers to the microphone spatial locations: black (origin), magenta (xz plane), and blue (xy plane).

$$\mathrm{oSNR}_{\mathrm{ROI}}(\mathbf{f})=\frac{p_X}{p_V}\times\frac{1}{|\Omega_{\mathrm{ROI}}|^2}\times\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}, \quad (13)$$

where the ROI steering vector $\mathbf{b}_{\mathrm{ROI}} = \int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}}\mathbf{d}_{\theta,\phi}\sin\theta \mathrm{d}\phi\mathrm{d}\theta$ is defined as the average steering vector of all DOAs within the ROI, and the spatial angle associated with the ROI is expressed as and $|\Omega_{\mathrm{ROI}}| = \int_{\theta\in\Theta_{\mathrm{ROI}}}\int_{\phi\in\Phi_{\mathrm{ROI}}}\sin\theta\mathrm{d}\phi\mathrm{d}\theta$. Consequently, the following distortion-controlled constraint was employed:

$$\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}=1, \quad (14)$$

as well as the SNR gain over the entire ROI:

$$\mathcal{G}_{\mathrm{ROI}}(\mathbf{f})=\frac{\mathrm{oSNR}_{\mathrm{ROI}}(\mathbf{f})}{\mathrm{iSNR}}=\frac{|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}|^2}{|\Omega_{\mathrm{ROI}}|^2\mathbf{f}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{f}}, \quad (15)$$

and its corresponding WNG and DF measures:

$$\mathcal{W}_{\mathrm{ROI}}(\mathbf{f})=\frac{\left|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}\right|^2}{|\Omega_{\mathrm{ROI}}|^2\mathbf{f}^H\mathbf{f}};\ \mathcal{D}_{\mathrm{ROI}}(\mathbf{f})=\frac{\left|\mathbf{f}^H\mathbf{b}_{\mathrm{ROI}}\right|^2}{|\Omega_{\mathrm{ROI}}|^2\mathbf{f}^H\mathbf{\Gamma}_{\mathrm{d}}\mathbf{f}}, \quad (16)$$

where $\mathbf{\Gamma}_{\mathrm{d}}$ is the pseudo-coherence matrix of a spherically isotropic (diffuse) noise field [27].

The beampattern, describing the spatial response of the beamformer for all spatial angles, is defined as

$$\mathcal{B}_{\theta,\phi}(\mathbf{f})=\mathbf{f}^H\mathbf{d}_{\theta,\phi}. \quad (17)$$

Note that the beampattern embodies an essential measure of undesired signal distortion (within the ROI) and susceptibility to interfering sources (outside the ROI).

4. OPTIMAL AND FREQUENCY-TUNED BEAMFORMERS

Our main objective is to optimize both the layout of the microphone array and the weights of the beamformer simultaneously, while considering a specified ROI for the conference room scenario described above. Unlike previous studies [22], [24], this research places particular emphasis on both azimuth and elevation angles. We do not assume that the desired speaker is located on the xy-plane or around it. This corresponds to a scenario with a standing (or walking) desired speaker while a small communication device is placed on a conference table at a significantly lower position than the speaker. We will also introduce a tuning technique that enables further design flexibility by accentuating the directivity in specific frequencies at the expense of others.

First, let us establish a proper optimization criterion and design constraints for deriving the beamformer weights. Since speech signals are broadband, we are interested in designing high-directivity beamformers across a large frequency range. Therefore, we utilize the ROI-oriented broadband directivity index defined as [22]

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

where f_L and f_H denote the minimal and maximal frequencies of interest, respectively. Then, we would like to maximize the criterion in (18) subject to certain constraints. For instance, it is often practically desirable to limit the total number of microphones to a subset of K out of M microphones, thereby balancing spatial resolution and efficiency. Thus, an optimal solution would be obtained by solving:

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

with $\mathbf{f}^*$ being a K-sparse beamformer. Moreover, the latter may not all be convex and involve NP-hard mixed-integer programming (MIP) optimization, e.g., as shown in [22]. Therefore, we perform the optimization in two sequential stages: first considering merely the UCSA on the xz plane and optimizing K_{xz} out of the QM_{xz} possible microphone locations, and then optimizing the remaining K_{xy} out of the PM_{xy} possible microphone locations on the xy plane while accounting for the K_{xz} previously optimized locations. Note that in this formulation $K = K_{\mathsf{xz}} + K_{\mathsf{xy}} + 1$, since the reference microphone location in the origin is always marked as occupied.

In the first stage, we focus on the occupied microphone locations for the UCSA on the xz plane. Following the expression in (19) and assuming the distortion-controlled constraint holds, we have

$$\mathbf{f}_{\mathsf{xz}}^* = \underset{\mathbf{f}_{\mathsf{xz}}}{\operatorname{argmin}} \int_{f_L}^{f_H} \mathbf{f}_{\mathsf{xz}}^H \mathbf{\Gamma}_{\mathrm{d}}^{\mathsf{xz}} \mathbf{f}_{\mathsf{xz}} df, \tag{20}$$

where $\mathbf{\Gamma}_{\mathrm{d}}^{\mathsf{xz}}$ is the $(QM_{\mathsf{xz}}+1) \times (QM_{\mathsf{xz}}+1)$ diffuse noise pseudo-coherence matrix that takes into account all UCSA microphone locations on the xz plane as well as the reference microphone.

To ensure the K-sparsity of the solution as well as the distortion-controlled constraint and a minimal acceptable level of WNG, we set the following four design constraints. First, accounting for the distortion-controlled constraint of (14) for all considered frequencies, we define

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

where $\mathbf{b}_{\mathrm{ROI}}^{\mathsf{xz}} = \int_{\theta \in \Theta_{\mathrm{ROI}}} \mathbf{a}_{\theta,\phi_{\mathrm{mid}}}^{\mathsf{xz}} \sin\theta \mathrm{d}\theta$, and ϕ_{mid} is the center of the ROI concerning the azimuth angle. Note that this constraint merely considers the elevation-angle part of the ROI, and that the complete ROI will be addressed in the second stage upon optimizing the UCSA on the xy plane. Moreover, it is highly desirable to guarantee a minimal level of WNG in practice. Invoking (16), we require that

$$\mathcal{C}_2[\mathbf{f}_{\mathsf{xz}}] : \mathbf{f}_{\mathsf{xz}}^H \mathbf{f}_{\mathsf{xz}} \leq \frac{1}{|\Theta_{\mathrm{ROI}}|^2 \epsilon}, \ \forall f \in [f_L, f_H], \tag{22}$$

in which the design parameter ϵ constitutes the minimal accepted level of $\mathcal{W}_{\mathrm{ROI}}(\mathbf{f})$, and $|\Theta_{\mathrm{ROI}}| = \int_{\theta \in \Theta_{\mathrm{ROI}}} \sin\theta \mathrm{d}\theta$. Then, to ensure the sparsity of the obtained solution, we denote $\mathbf{f}_{\mathsf{xz}} = \begin{bmatrix} F_1^{\mathsf{xz}} & \cdots & F_{QM_{\mathsf{xz}}+1}^{\mathsf{xz}} \end{bmatrix}^T$ and require

$$\mathcal{C}_3[\mathbf{f}_{\mathsf{xz}}] : \left|F_i^{\mathsf{xz}}\right|^2 \leq \frac{S_i^{\mathsf{xz}}}{|\Theta_{\mathrm{ROI}}|^2 \epsilon}, \ \forall f \in [f_L, f_H], \ \forall i = 1, \ldots, QM_{\mathsf{xz}}+1, \tag{23}$$

and

$$\mathcal{C}_4[\mathbf{s}_{\mathsf{xz}}] : \sum_{i=1}^{QM_{\mathsf{xz}}+1} S_i^{\mathsf{xz}} = K_{\mathsf{xz}} + 1, \tag{24}$$

where $\mathbf{s}_{\mathsf{xz}} = \begin{bmatrix} 1 & S_2^{\mathsf{xz}} & \cdots & S_{QM_{\mathsf{xz}}+1}^{\mathsf{xz}} \end{bmatrix}^T$ is a binary $(K_{\mathsf{xz}}+1)$-sparse vector ensuring that the beamformer weight corresponding to the microphone at the origin is always chosen. Ultimately, to receive the optimal solution $\mathbf{f}_{\mathsf{xz}}^*$ and its corresponding $K_{\mathsf{xz}}+1$ occupied microphone locations described by $\mathbf{s}_{\mathsf{xz}}$, we solve:

$$\begin{aligned} \mathbf{f}_{\mathsf{xz}}^* = \underset{\mathbf{f}_{\mathsf{xz}}}{\operatorname{argmin}} & \int_{f_L}^{f_H} W \mathbf{f}_{\mathsf{xz}}^H \mathbf{\Gamma}_{\mathrm{d}}^{\mathsf{xz}} \mathbf{f}_{\mathsf{xz}} df \\ \text{s.t.} \quad & \mathcal{C}_1[\mathbf{f}_{\mathsf{xz}}], \mathcal{C}_2[\mathbf{f}_{\mathsf{xz}}], \mathcal{C}_3[\mathbf{f}_{\mathsf{xz}}], \mathcal{C}_4[\mathbf{s}_{\mathsf{xz}}], \end{aligned} \tag{25}$$

where W is a frequency-dependent weighting term that may be applied to accentuate certain frequencies (e.g., placing greater penalties on low array directivity at lower frequencies).

For the second optimization stage, we consider the complete ROI, which accounts for both the elevation and azimuth angles, the occupied microphone locations optimized in the first stage, and all possible locations on the xy plane. Note that we do not directly account for the actual weights of $\mathbf{f}_{\mathsf{xz}}^*$, but rather only for the positions of its non-zero elements. Denoting the corresponding steering vector of length $PM_{\mathsf{xy}} + K_{\mathsf{xz}} + 1$ and the diffuse noise pseudo-coherence matrix of size $(PM_{\mathsf{xy}} + K_{\mathsf{xz}} + 1) \times (PM_{\mathsf{xy}} + K_{\mathsf{xz}} + 1)$ by $\bar{\mathbf{d}}_{\theta,\phi}$ and $\mathbf{\Gamma}_{\mathrm{d}}^{\mathsf{xyz}}$, respectively, we define the following constraints:

$$\mathcal{C}_5[\mathbf{f}_{\mathsf{xyz}}] : \mathbf{f}_{\mathsf{xyz}}^H \mathbf{b}_{\mathrm{ROI}}^{\mathsf{xyz}} = 1, \ \forall f \in [f_L, f_H], \tag{26}$$

where $\mathbf{f}_{\mathsf{xyz}}$ is the optimized beamformer and $\mathbf{b}_{\mathrm{ROI}}^{\mathsf{xyz}} = \int_{\theta \in \Theta_{\mathrm{ROI}}} \int_{\phi \in \Phi_{\mathrm{ROI}}} \bar{\mathbf{d}}_{\theta,\phi} \sin\theta \mathrm{d}\phi \mathrm{d}\theta$. Addressing the minimal WNG constraint, we have

$$\mathcal{C}_6[\mathbf{f}_{\mathsf{xyz}}] : \mathbf{f}_{\mathsf{xyz}}^H \mathbf{f}_{\mathsf{xyz}} \leq \frac{1}{|\Omega_{\mathrm{ROI}}|^2 \epsilon}, \ \forall f \in [f_L, f_H], \tag{27}$$

whereas the two following constraints guarantee the K-sparsity property of the optimal solution:

$$\begin{aligned} \mathcal{C}_7[\mathbf{f}_{\mathsf{xyz}}] : \left|F_i^{\mathsf{xyz}}\right|^2 \leq \frac{S_i^{\mathsf{xyz}}}{|\Omega_{\mathrm{ROI}}|^2 \epsilon}, \ \forall f \in [f_L, f_H], \\ \forall i = 1, \ldots, PM_{\mathsf{xy}} + K_{\mathsf{xz}} + 1, \end{aligned} \tag{28}$$

and

$$\mathcal{C}_8[\mathbf{s}_{\mathsf{xyz}}] : \sum_{i=1}^{PM_{\mathsf{xy}}+K_{\mathsf{xz}}+1} S_i^{\mathsf{xyz}} = K, \tag{29}$$

where $\mathbf{f}_{\mathsf{xyz}}$ is defined in a similar manner to $\mathbf{f}_{\mathsf{xz}}$, and $\mathbf{s}_{\mathsf{xyz}} = \begin{bmatrix} \mathbf{1}_{K_{\mathsf{xz}}+1} & S_{K_{\mathsf{xz}}+2}^{\mathsf{xyz}} & \cdots & S_{PM_{\mathsf{xy}}+K_{\mathsf{xz}}+1}^{\mathsf{xyz}} \end{bmatrix}^T$ is a binary K-sparse vector of length $PM_{\mathsf{xy}} + K_{\mathsf{xz}} + 1$ with $\mathbf{1}_{K_{\mathsf{xz}}+1}$ being an all-ones vector of length $K_{\mathsf{xz}}+1$. Finally, the proposed frequency-tuned dual circular sector (FT-DCS) beamformer of length K is obtained by solving

$$\begin{aligned} \mathbf{f}_{\mathrm{FT-DCS}} = \underset{\mathbf{f}_{\mathsf{xyz}}}{\operatorname{argmin}} & \int_{f_L}^{f_H} W \mathbf{f}_{\mathsf{xyz}}^H \mathbf{\Gamma}_{\mathrm{d}}^{\mathsf{xyz}} \mathbf{f}_{\mathsf{xyz}} df \\ \text{s.t.} \quad & \mathcal{C}_5[\mathbf{f}_{\mathsf{xyz}}], \mathcal{C}_6[\mathbf{f}_{\mathsf{xyz}}], \mathcal{C}_7[\mathbf{f}_{\mathsf{xyz}}], \mathcal{C}_8[\mathbf{s}_{\mathsf{xyz}}], \end{aligned} \tag{30}$$

which may be solved by any off-the-shelf optimization solver (e.g., MOSEK [28]). Note that in the special (frequency-untuned) case of $W = 1$, we refer to this beamformer as $\mathbf{f}_{\mathrm{DCS}}$.

5. EXPERIMENTAL RESULTS

In this part, we evaluate the performance of the proposed approach and compare it to existing beamformers from the literature. Specifically, we set $M_{\mathsf{xz}} = M_{\mathsf{xy}} = 3$, $\delta_{\mathsf{xz}} = \delta_{\mathsf{xy}} = 2$ cm, $Q = 5$ with $\{\gamma_0, \gamma_1, \ldots, \gamma_4\} = \{40^o, 50^o, \ldots, 80^o\}$, $P = 9$ with $\{\psi_0, \psi_1, \ldots, \psi_8\} = \{-40^o, -30^o, \ldots, 40^o\}$, $K_{\mathsf{xz}} = 5$, $K_{\mathsf{xy}} = 6$, $\epsilon = -30$ dB, and $\Theta_{\mathrm{ROI}} = [40^o, 80^o]; \Phi_{\mathrm{ROI}} = [-40^o, 40^o]$. We also set $W = e^{(8-f[\mathrm{kHz}])}$ for $\mathbf{f}_{\mathrm{FT-DCS}}$. Note that this implies utilizing 12 microphones out of 43 locations. Fig. 1 depicts the optimal array layout for $\mathbf{f}_{\mathrm{DCS}}$ and $\mathbf{f}_{\mathrm{FT-DCS}}$, respectively. We observe that, for both beamformers, the layout is symmetric around the center of the ROI (considering both angles); however, the aperture is greater with $\mathbf{f}_{\mathrm{FT-DCS}}$. This is apparent in the innermost ring of the solid-magenta-colored sparse sector on the xz plane and the

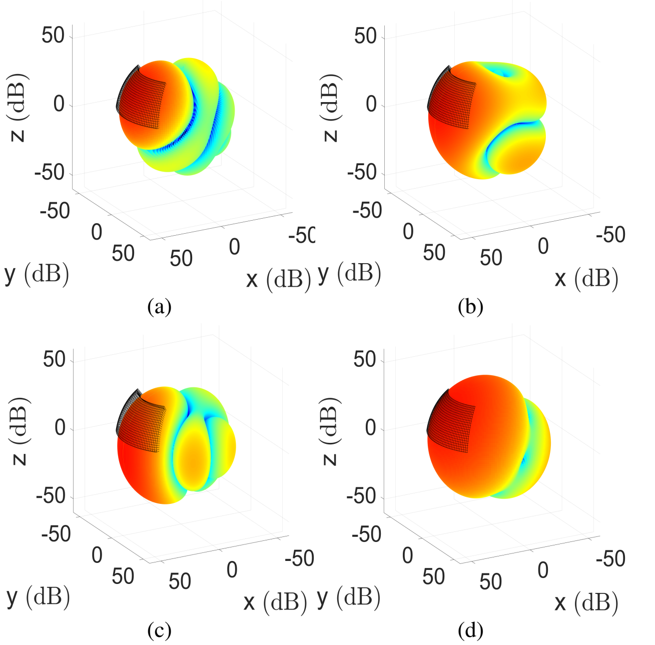


Fig. 2: 3-D beampatterns. (a) $\mathbf{f}_{\mathrm{FT-DCS}}$, (b) $\mathbf{h}_{\mathrm{NER}}$ [24], (c) $\mathbf{h}_{\mathrm{UCCA}}$, and (d) $\mathbf{h}_{\mathrm{spherical}}$. The distortionless response within the ROI is black shaded.

two larger rings of the solid-blue-colored sparse sector on the xy plane. This embodies $\mathbf{f}_{\mathrm{FT-DCS}}$'s design bias towards the lower frequencies for which the directivity benefits from a higher array aperture.

Next, we compare the 3-D beampatterns of the proposed $\mathbf{f}_{\mathrm{FT-DCS}}$ with those of three existing beamformers. First, is the near-end region (NER) of [24], $\mathbf{h}_{\mathrm{NER}}$, designed with a single ring of 7 microphones on the xy plane and 5 microphones on the z-axis. In addition, we simulate two maximum directivity beamformers of two distinct layouts: a UCCA, denoted by $\mathbf{h}_{\mathrm{UCCA}}$ and composed of 2 rings of 6 equally spaced microphones on each, and a uniform spherical array whose microphones are placed at the same angles of the ROI, with 4 equally spaced microphones along the azimuth direction and 3 equally spaced microphones along the elevation direction. It is denoted by $\mathbf{h}_{\mathrm{spherical}}$. We define the inner radius of $\mathbf{h}_{\mathrm{UCCA}}$ and $\mathbf{h}_{\mathrm{NER}}$ as well as the latter's z-axis interelement spacing to be 2 cm. The radius of $\mathbf{h}_{\mathrm{spherical}}$ is set to 3 cm. The beampatterns are shown in Fig. 2 for $f = 1200$ Hz, where the desired response within the ROI is black shaded to indicate potential distortion. We notice that $\mathbf{h}_{\mathrm{spherical}}$ exhibits a significant spatial response for a region well beyond the ROI as well as for the backlobe. Considering $\mathbf{h}_{\mathrm{NER}}$ and $\mathbf{h}_{\mathrm{UCCA}}$, we note that both exhibit a reduced mainlobe compared to $\mathbf{h}_{\mathrm{spherical}}$. However, $\mathbf{h}_{\mathrm{UCCA}}$'s planar geometry dictates an undesirably symmetric spatial response with respect to the xy plane that yields potential susceptibility to noise sources impinging from lower room positions, whereas both beamformers exhibit significant sidelobes. The latter further displays significant distortion at the top boundary of the ROI. In contrast, the mainlobe of the proposed $\mathbf{f}_{\mathrm{FT-DCS}}$ accurately matches the ROI with no signal distortion and without heavily amplifying unwanted regions outside of it. It also exhibits the lowest sidelobe and backlobe levels of all.

We end by comparing the $\mathcal{D}_{\mathrm{ROI}}$ and $\mathcal{W}_{\mathrm{ROI}}$ measures that are depicted in Fig. 3. Focusing on $\mathcal{D}_{\mathrm{ROI}}$, it is clear that the ROI beamformers, that is, $\mathbf{f}_{\mathrm{FT-DCS}}$, $\mathbf{f}_{\mathrm{DCS}}$, and $\mathbf{h}_{\mathrm{NER}}$ outperform the other two. Additionally, the latter is outperformed by the two proposed beamformers, a consequence of the linear structure of $\mathbf{h}_{\mathrm{NER}}$ along the z-axis, which mandates undesirably large mainlobe beamwidths along the elevation direction. We also note that $\mathbf{f}_{\mathrm{FT-DCS}}$ performs better than $\mathbf{f}_{\mathrm{DCS}}$ for low frequencies at the expense of reduced performance for high frequencies. Considering $\mathcal{W}_{\mathrm{ROI}}$, the proposed beamformers are shown to satisfy the minimal WNG level, with $\mathbf{f}_{\mathrm{FT-DCS}}$ performing best for all frequencies over $f = 1800$.

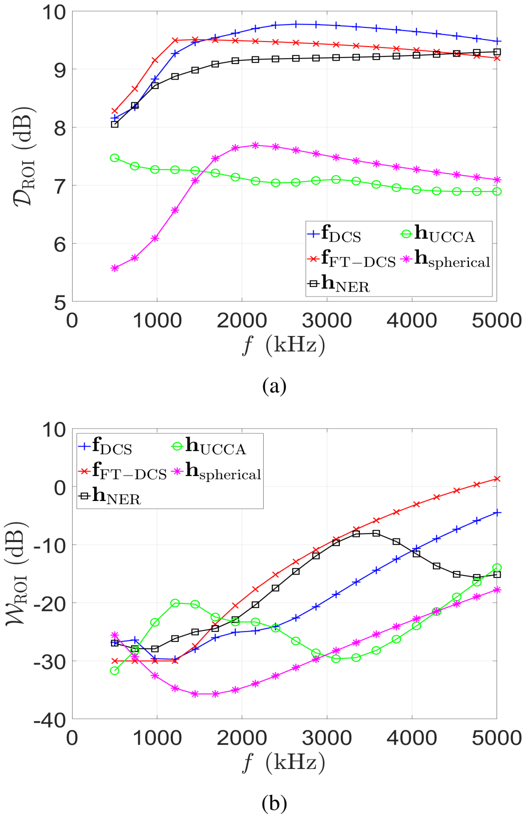


Fig. 3: (a) $\mathcal{D}_{\mathrm{ROI}}$ and (b) $\mathcal{W}_{\mathrm{ROI}}$ for the proposed and existing methods.

6. CONCLUSIONS

We have introduced an ROI beamforming approach that utilizes dual circular sectors for audio conferencing, each located either on the xz plane or the xy plane. Through a two-stage sparse optimization, we have proposed a practical approach that requires only a subset of array microphones whose cardinality is set by design. First, we consider the elevation-angle part of the ROI and optimize the occupied locations corresponding to the sector on the xz plane. Then, the latter are held fixed and the remaining locations are optimized on the xy plane while accounting for the complete ROI. Appropriate design constraints are maintained to ensure a controlled level of distortion over the entire ROI and a minimal level of WNG. We derive an optimal ROI beamformer that exhibits a controlled bias towards certain frequencies. We have evaluated the performance of two versions of the proposed beamformer, which differ in their frequency bias, and compared them to three beamforming methods from the literature. We have demonstrated that the proposed approach achieves a mainlobe that is better aligned with the ROI, while attenuating the sidelobes and backlobe to a greater extent. Finally, the proposed approach has been demonstrated to outperform existing methods in terms of the DF throughout the entire spectrum and at a preferred level of WNG for medium and high frequencies.

7. ACKNOWLEDGMENTS

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

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)*, Prague, 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-2022*, Bamberg, 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] 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*, Bamberg, Sep 2022.

[20] 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)*, New Paltz, New York, 2023.

[21] 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.

[22] G. Itzhak, S. 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-2024*, Aalborg, Sep 2024, pp. 135–139.

[23] 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.

[24] G. Itzhak and I. Cohen, "Robust beamforming for multispeaker audio conferencing under DOA uncertainty," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 139–151, 2025.

[25] G. Itzhak and I. Cohen, "STFT-domain least-distortion region-of-interest beamforming," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 2803–2816, 2025.

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

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

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