Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

Kronecker-Product Beamforming with Sparse Concentric Circular Arrays

Gal Itzhak[1] and Israel Cohen[1]

[1]Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 3200003, Israel

e-mail: Gal Itzhak: galitz@technion.ac.il; Israel Cohen: icohen@ee.technion.ac.il

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

ABSTRACT This paper presents a Kronecker-product (KP) beamforming approach incorporating sparse concentric circular arrays (SCCAs). The locations of the microphones on the SCCA are optimized concerning the broadband array directivity over a wide range of direction-of-arrival (DOA) deviations of a desired signal. A maximum directivity factor (MDF) sub-beamformer is derived accordingly with the optimal locations. Then, we propose two global beamformers obtained as a Kronecker product of a uniform linear array (ULA) and the SCCA sub-beamformer. The global beamformers differ by the type of the ULA, which is designed either as an MDF sub-beamformer along the x-axis or as a maximum white noise gain sub-beamformer along the y-axis. We analyze the performance of the proposed beamformers in terms of the directivity factor, the white noise gain, and their spatial beampatterns. Compared to traditional beamformers, the proposed beamformers exhibit considerably larger tolerance to DOA deviations concerning both the azimuth and elevation angles. Experimental results with speech signals in noisy and reverberant environments demonstrate that the proposed approach outperforms traditional beamformers regarding the perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) scores when the desired speech signals deviate from the nominal DOA.

INDEX TERMS Microphone arrays, Kronecker-product beamforming, concentric circular beamformers, direction-of-arrival deviations, sparse arrays.

I. Introduction

During the past few decades, beamforming has been investigated as a means to extract signals of interest from simultaneously sampled noisy observations in space [1]–[4]. Broadly, beamforming consists of two main ingredients: the spatial array geometry and the filter coefficients. While numerous methods and optimization criteria have been proposed to derive the coefficients [5]–[9], the array geometry diversity is traditionally more limited.

Efficiently designed and analyzed, uniform linear arrays (ULAs) are the most common array geometry in the literature and practice. While they may attain either a high white noise gain (WNG) or a high directivity factor (DF) (albeit not both simultaneously), it is well-known they are highly susceptible to the direction of arrival (DOA) of the desired signal [10], [11]. To cope with this problem, more complex array structures have been exploited. For example, rectangular arrays (RAs) exhibit reduced susceptibility to the desired signal's DOA. Furthermore, it may enable optimization concerning several criteria at once by taking advantage of the Kronecker-product (KP) beamforming framework [12]–[16]. RA structures are also known to be valuable in differential (closely-spaced microphones) settings [17], [18] and for DOA estimation methods [19], [20]. Nevertheless, RAs are not entirely DOA-independent as their symmetry dictates a clear preference for signals impinging on the array parallel to one of its axes.

Uniform circular arrays (UCAs) have been shown to allow a high level of control, compromising between the WNG, the DF, and a frequency-independent spatial response [21], [22]. Other studies have generalized UCAs by proposing uniform concentric circular array (UCCA) structures [23]–[25], which may also enable a constant mainlobe beamwidth concerning the azimuth angle, the elevation angle, or both [26]–[28]- a highly desirable property with broadband applications which has been extensively addressed in the

literature [29]–[33]. In addition, it is known that UCCA structures can mitigate low-WNG issues and beampattern irregularities in case adjacent microphones are closely spaced on the circumference of the UCA [34]. Recently, a linear array topology optimization approach has been proposed to maximize the broadband array directivity, considering a potential range of DOA deviations [35]. While it was shown to be valuable for relatively small deviations, it suffers from several inherent drawbacks. To begin with, although the topology was optimized considering a deviation range, the resulting beamformer required knowledge of the nominal desired signal's DOA, which is often unknown in practice. Moreover, its spatial sensing level is limited due to its linear nature. For example, it lacks the information to distinguish between signals impinging from any theoretical circle drawn around the linear array (cylindrical beampattern symmetry). Finally, this approach may not fit practical appliances with more than a few microphones due to the large physical size of such a nonuniform array, implying high computational complexity.

To overcome some of the drawbacks in [35], an optimized RA topology was introduced in [36], suggesting a uniform structure along one axis and a nonuniform structure along the other. This yielded high array directivity over a desirable deviation range and a constant mainlobe beamwidth simultaneously. Indeed, this approach did not require the nominal value of the desired signal's DOA and outperformed traditional topologies in scenarios involving DOA deviations. Unfortunately, the single-axis geometry optimization and the explicit constant beamwidth requirement dictated tolerance to relatively small DOA deviations, which entails a large physical size along the uniform axis of the array.

This paper combines a KP beamforming approach with sparse concentric circular arrays (SCCAs). Unlike previous studies, in this work, we seek tolerance for significant deviations of the desired signal DOA while mitigating the necessity of its precise *a priori* knowledge to derive the beamformers. First, we optimize the locations of the microphones on the SCCA for the broadband array directivity over a wide range of DOA deviations and derive a maximum directivity factor (MDF) sub-beamformer. Then, to extend the SCCA sub-beamformer without adding considerable computational complexity, we propose two global beamformers obtained as a KP of a ULA and the former. The global beamformers differ by the type of the ULA, designed either as an MDF beamformer along the x-axis or as a maximum white noise gain (MWNG) beamformer along the y-axis. The proposed beamformers outperform traditional and common beamformers in the DF and WNG, considering both the azimuth and elevation angles. Finally, experimental results with noisy speech signals in reverberant environments indicate that the proposed method is preferable to the compared methods regarding both the perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) scores. This is particularly emphasized when the desired speech signal significantly deviates from its nominal DOA.

The rest of the paper is organized as follows. Section II presents the signal model, including the mathematical formulation and notations. In Section III, we briefly review the KP beamforming framework. Section IV addresses the array topology optimization followed by the derivation of the two proposed linear SCCA beamformers. Next, Section V is dedicated to extensive simulations. We analyze the beampatterns and the DF and WNG of the proposed approach, considering a desirable range of deviations in both the azimuth and elevation angles, and compare its performance to traditional beamformers. Then, in the last part of this section, we perform speech signal simulations in noisy and reverberant environments. Finally, in Section VI, we summarize and conclude this study.

II. Signal Model

Consider a signal of interest propagating from the farfield in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, in an elevation angle θ and an azimuth angle ϕ. The plane wave impinges on a uniform concentric circular array (UCCA) composed of M microphones and N equally spaced rings whose corresponding radii are $R, 2R, \ldots, NR$ located on the x-y plane. The microphones are uniformly distributed and equally spaced on the array's rings (that is, the number of microphones on each ring is M/N, and the angular distance between every two adjacent microphones on the same ring is $2\pi N/M$). Defining the origin of the CCA as the reference point and using the Polar coordinate system, the array steering vector of the n-th ring is given by [2]:

$$\mathbf{a}_{n;\theta,\phi}\left(f\right) = \Big[\; e^{\jmath n \frac{2\pi f R}{c}\cos(\phi)\sin\theta} \quad e^{\jmath n \frac{2\pi f R}{c}\cos(\phi-\psi_1)\sin\theta} \quad \cdots \quad e^{\jmath n \frac{2\pi f R}{c}\cos(\phi-\psi_{M/N})\sin\theta} \;\Big]^T, \tag{1}$$

where $n = 1, \ldots, N$ 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

$$\psi_i = \frac{2\pi N(i-1)}{M} \tag{2}$$

is the angular distance between the i-th microphone on a ring ($i = 1, \ldots, M/N$) and the positive x-axis direction. Stacking the steering vectors of all rings, we obtain the full steering vector of the CCA:

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

Let us consider a uniform linear array (ULA) composed of $P \geq 2$ omnidirectional microphones with an interelement spacing equal to δ. Assuming the ULA is aligned with the x-axis, its corresponding steering vector is given by [2]

$$\mathbf{b}_{\theta,\phi;\mathsf{x}}\left(f\right) = \Big[\; 1 \quad \cdots \quad e^{\jmath 2\pi f(P-1)\delta\cos\phi\sin\theta/c} \;\Big]^T. \tag{4}$$

Alternatively, when the ULA is aligned with the y-axis, its corresponding steering vector is

$$\mathbf{b}_{\theta,\phi;\mathsf{y}}(f) = \left[\begin{array}{ccc} 1 & \cdots & e^{\jmath 2\pi f(P-1)\delta \sin\phi \sin\theta/c} \end{array}\right]^T. \quad (5)$$

With the aforementioned steering vectors at hand, we can exploit the Kronecker-product (KP) beamforming framework [14] and define $\mathbf{d}_{\theta,\phi;\mathsf{x}}(f)$ and $\mathbf{d}_{\theta,\phi;\mathsf{y}}(f)$ by

$$\mathbf{d}_{\theta,\phi;\mathsf{x}}(f) = \mathbf{b}_{\theta,\phi;\mathsf{x}}(f) \otimes \mathbf{a}_{\theta,\phi}(f) \quad (6)$$

$$\mathbf{d}_{\theta,\phi;\mathsf{y}}(f) = \mathbf{b}_{\theta,\phi;\mathsf{y}}(f) \otimes \mathbf{a}_{\theta,\phi}(f), \quad (7)$$

where $\otimes$ is the KP operator. We note that in this formulation, the KP operation yields an array constructed by replicas of the CCA - P replicas along the x-axis in case of $\mathbf{d}_{\theta,\phi;\mathsf{x}}(f)$, and P replicas along the y-axis in case of $\mathbf{d}_{\theta,\phi;\mathsf{y}}(f)$. Focusing on $\mathbf{d}_{\theta,\phi;\mathsf{x}}(f)$, without loss of generality, the observed signal vector of length MP can be expressed in the frequency domain as [10]:

$$\begin{aligned}\mathbf{y}(f) &= \left[\begin{array}{ccc} \mathbf{y}_1^T(f) & \mathbf{y}_2^T(f) & \cdots \quad \mathbf{y}_P^T(f) \end{array}\right]^T \\ &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}_{\theta,\phi;\mathsf{x}}(f) X(f) + \mathbf{v}(f), \end{aligned} \quad (8)$$

where $X(f)$ is the zero-mean desired source signal, $\mathbf{v}(f)$ is the zero-mean additive noise signal vector,

$$\mathbf{y}_p(f) = \left[\begin{array}{cccc} \mathbf{y}_{p;1}(f) & \mathbf{y}_{p;2}(f) & \cdots & \mathbf{y}_{p;N}(f) \end{array}\right]^T, \quad (9)$$

with $p = 1, 2, \ldots, P$ being the CCA replica index and

$$\mathbf{y}_{p;n}(f) = \left[\begin{array}{cccc} Y_{p;n;1}(f) & Y_{p;n;2}(f) & \cdots & Y_{p;n;M/N}(f) \end{array}\right]^T \quad (10)$$

is the observed signal vector of the N-th ring in the p-th replica, composed of the set of M/N observations $\{Y_{p;n;i}(f)\}_{i=1}^{M/N}$. Denoting the desired signal incident angle by (θ_0, ϕ_0) and dropping the explicit dependence on f, (8) becomes:

$$\mathbf{y} = (\mathbf{b}_{\theta_0,\phi_0;\mathsf{x}} \otimes \mathbf{a}_{\theta_0,\phi_0}) X + \mathbf{v}, \quad (11)$$

where $\mathbf{b}_{\theta_0,\phi_0;\mathsf{x}} \otimes \mathbf{a}_{\theta_0,\phi_0} = \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}$ is the steering matrix at θ_0, ϕ_0, and the covariance matrix of $\mathbf{y}$ is

$$\mathbf{\Phi}_{\mathbf{y}} = E\left(\mathbf{y}\mathbf{y}^H\right) = p_X \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}} \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}^H + \mathbf{\Phi}_{\mathbf{v}}, \quad (12)$$

where $E(\cdot)$ denotes mathematical expectation, the superscript ^{H} is the conjugate-transpose operator, $p_X = E\left(|X|^2\right)$ is the variance of X, and $\mathbf{\Phi}_{\mathbf{v}} = E\left(\mathbf{v}\mathbf{v}^H\right)$ is the covariance matrix of $\mathbf{v}$. Assuming that the variance of the noise is approximately the same at all sensors, we can express (12) as

$$\mathbf{\Phi}_{\mathbf{y}} = p_X \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}} \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}^H + p_V \mathbf{\Gamma}_{\mathbf{v}}, \quad (13)$$

where p_V is the variance of the noise at a reference microphone (e.g., the first microphone on the inner ring of the first CCA replica) and $\mathbf{\Gamma}_{\mathbf{v}} = \mathbf{\Phi}_{\mathbf{v}}/p_V$ is the pseudo-coherence matrix of the noise. From (13), we deduce that the input signal-to-noise ratio (SNR) is

$$\mathrm{iSNR} = \frac{\mathrm{tr}\left(p_X \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}} \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}^H\right)}{\mathrm{tr}\left(p_V \mathbf{\Gamma}_{\mathbf{v}}\right)} = \frac{p_X}{p_V}, \quad (14)$$

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

We end this part by noting that equations (8)-(14) remain valid by substituting $\mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}$ and $\mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}$ with $\mathbf{b}_{\theta_0,\phi_0;\mathsf{y}}$ and $\mathbf{d}_{\theta_0,\phi_0;\mathsf{y}}$, respectively.

III. Kronecker-Product Beamforming

Let us assume a global beamformer $\mathbf{f}$ of length MP is obtained by a KP of two sub-beamformers, $\mathbf{g}$ and $\mathbf{h}$. Hence, $\mathbf{f}$ is of the form:

$$\mathbf{f} = \mathbf{g} \otimes \mathbf{h}, \quad (15)$$

where $\mathbf{h}$ is a sub-beamformer of length M corresponding to a CCA, and $\mathbf{g}$ is a sub-beamformer of length P corresponding to a ULA. Then, the global beamformer output signal is

$$\begin{aligned} Z &= \mathbf{f}^H \mathbf{y} \\ &= \left(\mathbf{g}^H \mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}\right)\left(\mathbf{h}^H \mathbf{a}_{\theta_0,\phi_0}\right) X + (\mathbf{g} \otimes \mathbf{h})^H \mathbf{v}, \end{aligned} \quad (16)$$

where Z is an estimate of X composed of a filtered desired signal term and a residual noise term, and $\mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}$ is used without loss of generality. In addition, it is clear that a distortionless constraint is satisfied by

$$\mathbf{h}^H \mathbf{a}_{\theta_0,\phi_0} = 1, \ \mathbf{g}^H \mathbf{b}_{\theta_0,\phi_0;\mathsf{x}} = 1. \quad (17)$$

The output SNR and the gain in SNR are, respectively,

$$\mathrm{oSNR}(\mathbf{f}) = \frac{p_X}{p_V} \times \frac{\left|\mathbf{f}^H \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}\right|^2}{\mathbf{f}^H \mathbf{\Gamma}_{\mathbf{v}} \mathbf{f}}, \quad (18)$$

and

$$\mathcal{G}(\mathbf{f}) = \frac{\mathrm{oSNR}(\mathbf{f})}{\mathrm{iSNR}} = \frac{\left|\mathbf{f}^H \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}\right|^2}{\mathbf{f}^H \mathbf{\Gamma}_{\mathbf{v}} \mathbf{f}}, \quad (19)$$

from which we deduce the WNG:

$$\begin{aligned} \mathcal{W}(\mathbf{f}) &= \frac{\left|\mathbf{f}^H \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}\right|^2}{\mathbf{f}^H \mathbf{f}} = \frac{\left|\mathbf{g}^H \mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}\right|^2}{\mathbf{g}^H \mathbf{g}} \times \frac{\left|\mathbf{h}^H \mathbf{a}_{\theta_0,\phi_0}\right|^2}{\mathbf{h}^H \mathbf{h}} \\ &= \mathcal{W}(\mathbf{g}) \times \mathcal{W}(\mathbf{h}), \end{aligned} \quad (20)$$

and the DF:

$$\mathcal{D}(\mathbf{f}) = \frac{\left|\mathbf{f}^H \mathbf{d}_{\theta_0,\phi_0;\mathsf{x}}\right|^2}{\mathbf{f}^H \mathbf{\Gamma}_{\mathrm{d}} \mathbf{f}}, \quad (21)$$

where $\mathbf{\Gamma}_{\mathrm{d}}$ is the $MP \times MP$ pseudo-coherence matrix of the spherically isotropic (diffuse) noise field [1] defined by

$$[\mathbf{\Gamma}_{\mathrm{d}}]_{i_1,i_2} = \mathrm{sinc}(2\pi f r_{i_1,i_2}/c), \quad (22)$$

$i_1, i_2 = 1, 2, \ldots, MP$ are the microphone indices of the global beamformer, r_{i_1,i_2} is the Euclidean distance between the i_1-th and i_2-th microphones, and $\mathrm{sinc}(x) = \sin(x)/x$.

Finally, the beampattern is given by

$$\begin{aligned} \mathcal{B}_{\theta,\phi}(\mathbf{f}) &= \mathbf{f}^H \mathbf{d}_{\theta,\phi;\mathsf{x}} = \left(\mathbf{g}^H \mathbf{b}_{\theta,\phi;\mathsf{x}}\right)\left(\mathbf{h}^H \mathbf{a}_{\theta,\phi}\right) \\ &= \mathcal{B}_{\theta,\phi}(\mathbf{g})\, \mathcal{B}_{\theta,\phi}(\mathbf{h}), \end{aligned} \quad (23)$$

where $\mathcal{B}_{\theta,\phi}(\mathbf{g}) = \mathbf{g}^H \mathbf{b}_{\theta,\phi;\mathsf{x}}$ may be seen as the beampattern of $\mathbf{g}$ and $\mathcal{B}_{\theta,\phi}(\mathbf{h}) = \mathbf{h}^H \mathbf{a}_{\theta,\phi}$ may be seen as the beampattern of $\mathbf{h}$.

This article has been accepted for publication in IEEE Open Journal of Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2023.3339433

IV. SCCA-Based KP Beamformers

We aim to design a high-directivity beamformer tolerant of significant DOA errors concerning the azimuth angle. That is, while most studies assume the desired signal can be modeled as a point source whose DOA is either known in advance or may be precisely estimated, in this work, we take into account considerable deviations of the DOA from its nominal value. For example, focusing on the endfire direction ($\phi_0 = 0^o$ with respect to the positive x-axis direction), we assume a wide potential range of DOA deviations on the x-y plane of all angles satisfying $|\phi| \leq \phi_H$. We note that unlike previous studies, which considered a relatively narrow deviation range ($\phi_H \approx 20^o - 30^o$) [33], [35], [36], in this study, we target more severe cases which require a broader range. To achieve this, we first take advantage of the microphone array topology optimization approach suggested in [35] with ULAs and in [36] with nonuniform rectangular arrays and optimize a K-sparse CCA structure out of a UCCA of M possible microphone locations (uniformly distributed and equally spaced on the N UCCA's rings, as elaborated in Section II). The topology optimization is performed concerning the broadband directivity index:

$$\mathcal{DI}_{[f_L,f_H]}\left[\mathbf{h}\right] = \frac{\int_{f_L}^{f_H} \left|\mathbf{a}_{\theta_0,\phi_0}^H \mathbf{h}\right|^2 df}{\int_{f_L}^{f_H} \mathbf{h}^H \mathbf{\Gamma}_{\mathrm{d,UCCA}} \mathbf{h} df}, \tag{24}$$

where $\mathbf{\Gamma}_{\mathrm{d,UCCA}}$ is the $M \times M$ pseudo-coherence diffuse noise of the complete UCCA whose elements are defined as in (22), and f_L and f_H are the minimal and maximal frequencies of interest, respectively. It is important to mention that as in [35], [36], M can not be set too high, or else the computational complexity of the underlying iterative optimization algorithm would turn impractically high.

The output of the convex optimization algorithm is a subset of K microphones which correspond to a steering vector $\bar{\mathbf{a}}_{\theta_0,\phi_0}$ of length K whose elements are taken from the appropriate elements of the complete UCCA steering vector $\mathbf{a}_{\theta_0,\phi_0}$. With $\bar{\mathbf{a}}_{\theta_0,\phi_0}$ at hand, we can design a maximum directivity factor (MDF) sub-beamformer of length K that minimizes the diffuse noise considering the optimal sparse array topology and the nominal DOA $(\theta_0, \phi_0) = (90^o, 0^o)$, which is given upon solving

$$\min_{\mathbf{h}} \ \mathbf{h}^H \mathbf{\Gamma}_{\mathrm{d,SCCA}} \mathbf{h} \quad \text{s. t.} \quad \mathbf{h}^H \bar{\mathbf{a}}_{\theta_0,\phi_0} = 1. \tag{25}$$

The solution is obtained by

$$\mathbf{h}_{\mathrm{MDF}} = \frac{\mathbf{\Gamma}_{\mathrm{d,SCCA}}^{-1} \bar{\mathbf{a}}_{\theta_0,\phi_0}}{\bar{\mathbf{a}}_{\theta_0,\phi_0}^H \mathbf{\Gamma}_{\mathrm{d,SCCA}}^{-1} \bar{\mathbf{a}}_{\theta_0,\phi_0}}, \tag{26}$$

where $\mathbf{\Gamma}_{\mathrm{d,SCCA}}$ is the $K \times K$ diffuse noise pseudo-coherence matrix of the SCCA whose elements are defined as in (22).

Next, to extend the SCCA sub-beamformer without adding significant computational complexity, we design two versions of the ULA sub-beamformer of length P and an interelement spacing that equals δ. The first ULA, designed as an MDF sub-beamformer, is aligned with the x-axis, corresponds to $\mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}$ in Section II and is denoted by $\mathbf{g}_{\mathrm{MDF}}$. It is given by

$$\mathbf{g}_{\mathrm{MDF}} = \frac{\mathbf{\Gamma}_{\mathrm{d,ULA}}^{-1} \mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}}{\mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}^H \mathbf{\Gamma}_{\mathrm{d,ULA}}^{-1} \mathbf{b}_{\theta_0,\phi_0;\mathsf{x}}}, \tag{27}$$

where $\mathbf{\Gamma}_{\mathrm{d,ULA}}$ is the $P \times P$ pseudo-coherence diffuse noise of the ULA whose elements are defined as in (22). Then, by invoking the KP beamforming approach discussed in Section III, we can define $\mathbf{f}_{\mathrm{MDF/MDF}}$ by

$$\mathbf{f}_{\mathrm{MDF/MDF}} = \mathbf{g}_{\mathrm{MDF}} \otimes \mathbf{h}_{\mathrm{MDF}}, \tag{28}$$

which corresponds to an x-linear SCCA constructed by replicas of an SCCA along the x-axis. We note that the steering vector of a farfield signal impinging on this array is given by

$$\bar{\mathbf{d}}_{\theta,\phi;\mathsf{x}} = \mathbf{b}_{\theta,\phi;\mathsf{x}} \otimes \bar{\mathbf{a}}_{\theta,\phi}. \tag{29}$$

Alternatively, we may design the ULA as an MWNG sub-beamformer aligned with the y-axis. It corresponds to $\mathbf{b}_{\theta_0,\phi_0;\mathsf{y}}$ and is given by

$$\mathbf{g}_{\mathrm{MWNG}} = \frac{\mathbf{b}_{\theta_0,\phi_0;\mathsf{y}}}{\mathbf{b}_{\theta_0,\phi_0;\mathsf{y}}^H \mathbf{b}_{\theta_0,\phi_0;\mathsf{y}}} = \frac{\mathbf{b}_{\theta_0,\phi_0;\mathsf{y}}}{P} = \frac{\mathbf{1}}{P}, \tag{30}$$

where $\mathbf{1}$ is an all-ones vector of length P. Then, we have

$$\mathbf{f}_{\mathrm{MWNG/MDF}} = \mathbf{g}_{\mathrm{MWNG}} \otimes \mathbf{h}_{\mathrm{MDF}}, \tag{31}$$

which corresponds to a y-linear SCCA constructed by replicas of an SCCA along the y-axis. The appropriate steering vector of a farfield signal impinging on this array is given by

$$\bar{\mathbf{d}}_{\theta,\phi;\mathsf{y}} = \mathbf{b}_{\theta,\phi;\mathsf{y}} \otimes \bar{\mathbf{a}}_{\theta,\phi}. \tag{32}$$

We end this part by noting that while $\mathbf{h}_{\mathrm{MDF}}$ is designed by considering a desirable deviation range concerning merely the azimuth angle, it will be shown next that both $\mathbf{f}_{\mathrm{MDF/MDF}}$ and $\mathbf{f}_{\mathrm{MWNG/MDF}}$ exhibit a significant tolerance to DOA deviations concerning the elevation angle as well. While this property is not directly optimized, it is received as an artifact of the planar array structure, which exhibits lower sensitivity to signals impinging outside the x-y plane.

V. Experimental Results

A. Empirical Optimization of K

The convex optimization algorithm discussed in the previous section requires the configuration of several parameters, e.g., M, N, R, ϕ_H and K. Most parameters are usually confined by practical considerations involving the physical size of the array and possible microphone placements. These include M, N, and R. Additionally, the parameter ϕ_H embodies the desirable deviation range: the tolerance to the desired signal's DOA deviation concerning the azimuth angle, which is typically a requirement derived from a scenario of interest. In contrast, it is not *a priori* clear what value should be assigned to the parameter K, and what its effect on the performance of the beamformer.

To roughly assess the appropriate or "optimal" value of K, we iterate over all possible values from 2-to-18 with

$M = 20$, $N = 2$, $R = 1$ cm and $\phi_H = 45^o$, and investigate the DF and WNG measures of the corresponding version of the sub-beamformer $\mathbf{h}_{\mathrm{MDF}}$ as a function of the azimuth angle ϕ and frequency f. The results are depicted in Figure 1. When $K = 4$, we observe that the WNG is roughly independent of ϕ within the analyzed deviation range and strictly positive for high frequencies, whereas the DF is poor. While this implies high tolerance to DOA deviations, the mainlobe width is inevitably large, and the sensitivity to diffuse noise (and reverberations) is considerable. As the number of array microphones K increases, the DF improves at the expense of the WNG. For example, when $K = 16$, the directivity of $\mathbf{h}_{\mathrm{MDF}}$ is high. However, the WNG is low, particularly in low frequencies. Consequently, we infer that a practical value of K should be selected around 8 as a compromise between the two measures.

B. Performance Analysis of the SCCA Beamformers

Based on the former part, let us now set the following parameters to design $\mathbf{f}_{\mathrm{MDF/MDF}}$ and $\mathbf{f}_{\mathrm{MWNG/MDF}}$. We set $M = 24$ equally spaced and uniformly distributed possible microphone locations on a CCA with $N = 2$ rings. The number of microphones per SCCA is set to $K = 8$, whereas the number of SCCA replicas is set to $P = 3$. We note that P should be chosen small from practical considerations as it multiplies the number of array microphones. Consequently, in this setting, we employ 24 microphones. In addition, the inner-ring radius is set to $R = 1$ cm, the interelement spacing of the ULA is $\delta = 2.5$ cm, the one-sided threshold azimuth angle is $\phi_H = 45^o$, the minimal value of the WNG for the convex optimization is set to -50dB as a compromise to allow high directivity while avoiding extreme white noise amplification, and the frequency range of interest is $[f_L, f_H] = [1, 6]$ kHz.

Figure 2 shows the two optimal x-linear SCCAs and y-linear SCCAs. Empty circles indicate potential microphone locations, while filled circles indicate microphone placements. We observe that the optimal SCCA topology comprises more microphones placed on the inner ring (5) than on the outer ring (3). This makes much sense as it is well known that MDF (or super-directive) beamformers improve their directivity as the Euclidean distance between adjacent microphones reduces. In addition, as the deviation range is centered around $\theta_0, \phi_0 = 0^o$ and wide, it is clear why the microphones located on the outer ring are on its left half - they form differential ULA-like structures when combined with the microphones located on the inner ring, for a wide DOA range around 0^o. With $\mathbf{f}_{\mathrm{MDF/MDF}}$, the array's ULA-like structures increase and contain more closely-spaced microphones concerning the endfire direction to allow further directivity improvement. In contrast, with $\mathbf{f}_{\mathrm{MWNG/MDF}}$, the SCCA replicas sense directional signals around the endfire direction with a small phase shift, which indeed fits a broadband MWNG (or delay-and-sum) array structure.

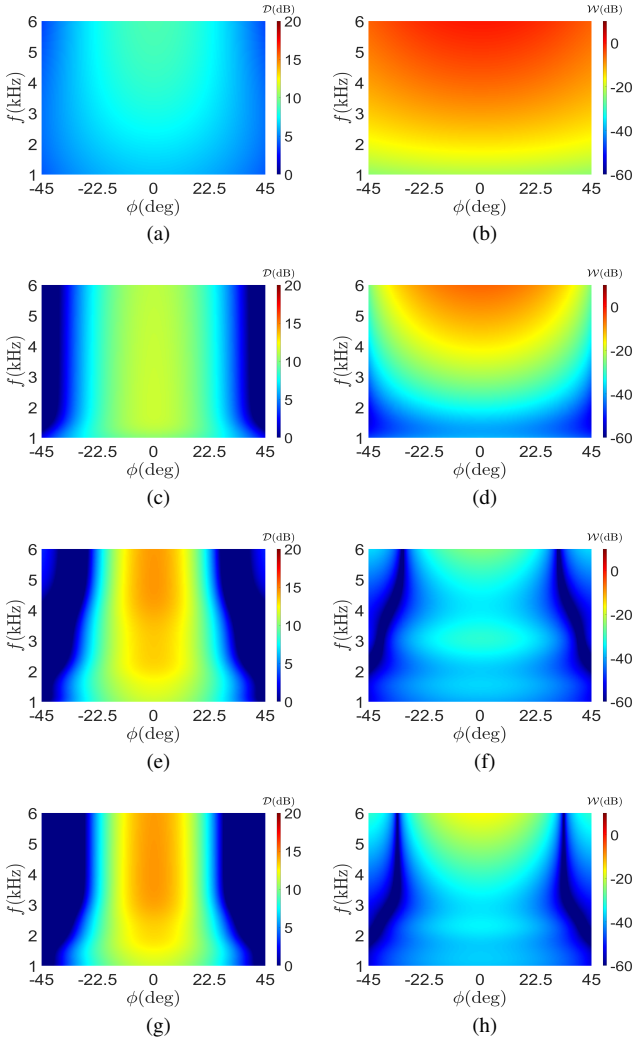

FIGURE 1: DF and WNG of the sub-beamformer $\mathbf{h}_{\mathrm{MDF}}$ corresponding to a single SCCA replica as a function of the azimuth angle ϕ with $M = 20$ and different values of K. (a) DF with $K = 4$, (b) WNG with $K = 4$, (c) DF with $K = 8$, (d) WNG with $K = 8$, (e) DF with $K = 12$, (f) WNG with $K = 12$, (g) DF with $K = 16$, and (h) WNG with $K = 16$.

Next, in Figure 3, we analyze the DF and WNG of $\mathbf{f}_{\mathrm{MDF/MDF}}$ and $\mathbf{f}_{\mathrm{MWNG/MDF}}$ with the setting above, along with two existing well-known beamformers containing 24 microphones: a (complete) UCCA with 6 rings and 4 equally spaced microphones per ring $\mathbf{f}_{\mathrm{UCCA}}$, and a differential uniform linear array (DULA), $\mathbf{f}_{\mathrm{DULA}}$, with a small and constant interelement spacing of 1 cm. Both beamformers are designed as MDF beamformers, optimized considering the minimization of the diffuse noise field, in a similar manner to (25). Due to the practical drawbacks discussed in Section I (mainly, tolerance to small DOA deviations and large physical size), we compare our proposed approach to more traditional and common beamformers. We immediately observe that the WNG in low frequencies is higher for

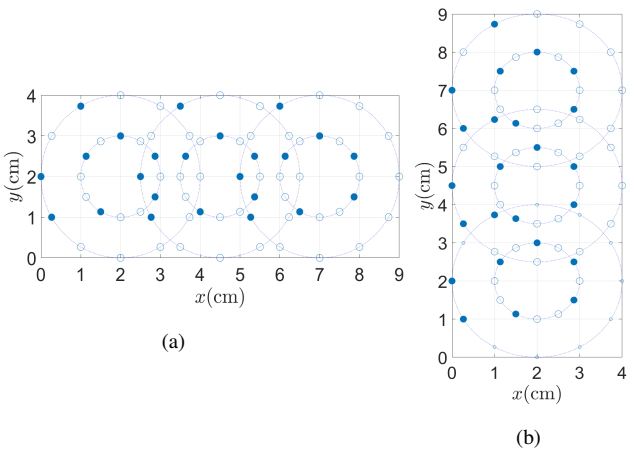

FIGURE 2: Optimal array topologies for the two proposed global beamformers. (a) $\mathbf{f}_{\mathrm{MDF/MDF}}$ and (b) $\mathbf{f}_{\mathrm{MWNG/MDF}}$. The desired speech signal impinges on the array from the positive x-axis direction.

$\mathbf{f}_{\mathrm{MWNG/MDF}}$ than for $\mathbf{f}_{\mathrm{MDF/MDF}}$. In contrast, both exhibit a preferable WNG than $\mathbf{f}_{\mathrm{UCCA}}$ and $\mathbf{f}_{\mathrm{DULA}}$ either in large DOA deviations or in frequencies higher than 2 kHz. Considering the DF measure, $\mathbf{f}_{\mathrm{DULA}}$ is superior if the DOA deviation is small. However, its performance sharply drops when the DOA deviation is significant. In contrast, the linear SCCA beamformers exhibit a better tolerance to significant deviations, with $\mathbf{f}_{\mathrm{MDF/MDF}}$ outperforming $\mathbf{f}_{\mathrm{MWNG/MDF}}$. We note that $\mathbf{f}_{\mathrm{UCCA}}$ is more tolerant to deviations than $\mathbf{f}_{\mathrm{DULA}}$ considering both measures, yet none are as tolerant as $\mathbf{f}_{\mathrm{MDF/MDF}}$ or $\mathbf{f}_{\mathrm{MWNG/MDF}}$.

Figure 4 depicts the DF and WNG of the four discussed beamformers concerning the elevation angle θ. We observe that $\mathbf{f}_{\mathrm{MDF/MDF}}$ and $\mathbf{f}_{\mathrm{MWNG/MDF}}$ are even more tolerant to DOA deviations concerning the elevation angle than to the azimuth angle. We note that this property, although not directly optimized for, is obtained as an artifact of the 2-D array geometry, which is less sensitive to elevation angle deviations, having no microphones to sense acoustic pressure differences in the direction perpendicular to the array plane. In contrast, the performance of $\mathbf{f}_{\mathrm{DULA}}$ with elevation angle deviations is similar to its performance with azimuth angle deviations as the underlying one-dimensional (1-D) array geometry senses variations concerning merely a single angle in the same way. Addressing $\mathbf{f}_{\mathrm{UCCA}}$, we note that its geometry may be seen as an ensemble of small DULAs uniformly spread on a circle circumference, which implies that the 1-D geometry artifact discussed with $\mathbf{f}_{\mathrm{DULA}}$ roughly applies to this beamformer as well.

We end this part by addressing the beampatterns of the two discussed beamformers concerning both the azimuth and elevation angles. The beampatterns are depicted in Figure 5. We observe that both beamformers exhibit a roughly constant

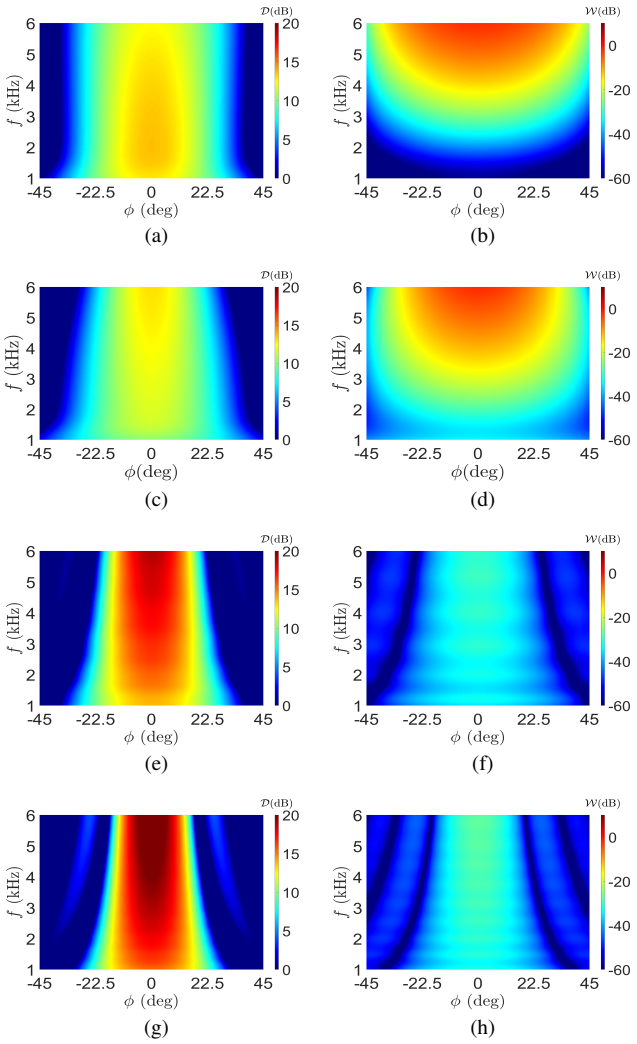

FIGURE 3: DF and WNG as a function of the azimuth angle ϕ with the four discussed beamformers. (a) DF with $\mathbf{f}_{\mathrm{MDF/MDF}}$, (b) WNG with $\mathbf{f}_{\mathrm{MDF/MDF}}$, (c) DF with $\mathbf{f}_{\mathrm{MWNG/MDF}}$, (d) WNG with $\mathbf{f}_{\mathrm{MWNG/MDF}}$, (e) DF with $\mathbf{f}_{\mathrm{UCCA}}$, (f) WNG with $\mathbf{f}_{\mathrm{UCCA}}$, (g) DF with $\mathbf{f}_{\mathrm{DULA}}$, and (h) WNG with $\mathbf{f}_{\mathrm{DULA}}$.

mainlobe beamwidth concerning the azimuth angle, which indicates their tolerance to DOA deviations from another perspective. It is also evident that angles outside the mainlobe (or deviation range) are greater reduced with $\mathbf{f}_{\mathrm{MDF/MDF}}$ than with $\mathbf{f}_{\mathrm{MWNG/MDF}}$ which implies better directivity of the former. Considering the elevation angle, we note that the mainlobe is even wider, with approximately 10 dB attenuation at a DOA deviation of 50^o with $\mathbf{f}_{\mathrm{MDF/MDF}}$ and 60^o with $\mathbf{f}_{\mathrm{MWNG/MDF}}$. In contrast, nulls only appear when the deviation is as extreme as 70^o.

C. Speech Signals in Noisy Reverberant Environments

In this section, we demonstrate the practicality of the proposed approach through simulations of speech signals

TABLE 1: Average PESQ and STOI scores of the time-domain enhanced signals corresponding to the discussed beamformers. We set $T_{60} = 250$ msec in these settings.

	PESQ				STOI			
$(\Delta_\theta, \Delta_\phi) =$	$(0^o, 0^o)$	$(20^o, 0^o)$	$(40^o, 0^o)$	$(40^o, 40^o)$	$(0^o, 0^o)$	$(20^o, 0^o)$	$(40^o, 0^o)$	$(40^o, 40^o)$
$Y_{\rm ref}$	1.87	2.01	1.83	1.94	0.68	0.77	0.73	0.74
$\mathbf{f}_{\rm MDF/MDF}$	2.47	2.46	2.52	**2.44**	0.79	0.84	0.79	0.70
$\mathbf{f}_{\rm MWNG/MDF}$	**2.50**	**2.85**	**2.77**	**2.45**	**0.82**	**0.89**	**0.90**	**0.83**
$\mathbf{f}_{\rm UCCA}$	2.36	2.35	2.27	2.24	0.78	0.85	0.81	0.79
$\mathbf{f}_{\rm DULA}$	2.38	2.08	2.15	1.97	0.78	0.78	0.76	0.69
$\mathbf{f}_{\rm DS}$	2.09	2.31	2.14	2.04	0.74	0.83	0.80	0.72

TABLE 2: Average PESQ and STOI scores of the time-domain enhanced signals corresponding to the discussed beamformers. We set $T_{60} = 500$ msec in these settings.

	PESQ				STOI			
$(\Delta_\theta, \Delta_\phi) =$	$(0^o, 0^o)$	$(20^o, 0^o)$	$(40^o, 0^o)$	$(40^o, 40^o)$	$(0^o, 0^o)$	$(20^o, 0^o)$	$(40^o, 0^o)$	$(40^o, 40^o)$
$Y_{\rm ref}$	1.93	2.04	1.80	1.91	0.54	0.67	0.66	0.67
$\mathbf{f}_{\rm MDF/MDF}$	2.07	2.14	2.16	**2.28**	0.58	**0.75**	0.68	0.63
$\mathbf{f}_{\rm MWNG/MDF}$	2.14	**2.36**	**2.26**	2.20	**0.61**	**0.75**	**0.78**	**0.73**
$\mathbf{f}_{\rm UCCA}$	**2.18**	2.09	2.01	2.04	0.59	0.73	0.68	0.67
$\mathbf{f}_{\rm DULA}$	2.15	1.99	2.00	1.95	**0.61**	0.66	0.64	0.60
$\mathbf{f}_{\rm DS}$	2.03	2.13	2.01	2.00	**0.61**	0.74	0.73	0.66

in noisy and reverberant environments and with DOA deviations of the desired sources. We use a room impulse response (RIR) generator [37] to simulate the reverberant noise-free signals received in the four beamformers discussed in the previous part and with the same configurations, that is, $\mathbf{f}_{\rm MDF/MDF}$, $\mathbf{f}_{\rm MWNG/MDF}$, $\mathbf{f}_{\rm UCCA}$ and $\mathbf{f}_{\rm DULA}$, each consisting of 24 microphones. In addition, we simulate the traditional linear DS beamformer, denoted by $\mathbf{f}_{\rm DS}$, as a reference. We carry out the simulations in four distinct scenarios where the desired speech signal source is relocated to form different deviations in its DOA: $(\Delta_\theta, \Delta_\phi) = (0^o, 0^o)$, $(\Delta_\theta, \Delta_\phi) = (20^o, 0^o)$, $(\Delta_\theta, \Delta_\phi) = (40^o, 0^o)$ and $(\Delta_\theta, \Delta_\phi) = (40^o, 40^o)$, where Δ_θ represents the DOA deviation concerning the elevation angle and Δ_ϕ represents the DOA deviation concerning the azimuth angle. We note that each scenario is characterized by a different reverberation pattern (as the location of the speech signal source varies). Hence, the scenarios cannot be directly compared but instead individually analyzed.

In each of the four scenarios, the array is located on the $z = 1$ m plane and centered around the $(x, y) = (3, 3)$ m coordinate of a $6 \times 6 \times 4$ m room. The RIR is simulated with two different values of T_{60}, 250 msec and 500 msec, where T_{60} is defined by Sabin-Franklin's formula [38]. In addition, two simulated noise fields are present: a white thermal Gaussian noise and a spherically-isotropic diffuse noise, with the latter being 30 dB more powerful than the former; overall, the input SNR is set to $\mathrm{iSNR} = 0$ dB. The desired speech signal, $x(t)$, is a concatenation of 24 speech signals (12 speech signals per gender) with varying dialects that are taken from the TIMIT database [39] and sampled at a sampling rate of $f_{\rm s} = 1/T_{\rm s} = 16$ kHz. The speech signal enhancement is performed in the short-time Fourier transform (STFT) domain using 75% overlapping time frames and a Hamming analysis window of length 256 (16 msec).

We analyze and compare the average PESQ [40] and STOI [41] scores of the time-domain enhanced signals with each beamformer in all the scenarios discussed above. The results are shown in Tables 1 and 2 for $T_{60} = 250$ msec and $T_{60} = 250$ msec, respectively, along with the scores of the noisy speech signal as it is received by a reference microphone located in the $(3, 3, 1)$ m coordinate and denoted by $Y_{\rm ref}$. When $T_{60} = 250$ msec, that is, when the reverberations are mild, $\mathbf{f}_{\rm MWNG/MDF}$ outperforms all other beamformers by a great deal in terms of both scores. This is particularly stressed when $(\Delta_\theta, \Delta_\phi) = (40^o, 0^o)$ or $(\Delta_\theta, \Delta_\phi) = (40^o, 40^o)$, i.e., when the DOA deviations are significant. This results from the beamformer's tolerance to DOA deviations concerning both angles, as elaborated in the former part, and its superior WNG performance even in low frequencies. In contrast, when the reverberations are strong and $T_{60} = 500$ msec, $\mathbf{f}_{\rm MDF/MDF}$ and $\mathbf{f}_{\rm MWNG/MDF}$ are not better than the other beamformers in the zero-

This article has been accepted for publication in IEEE Open Journal of Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2023.3339433

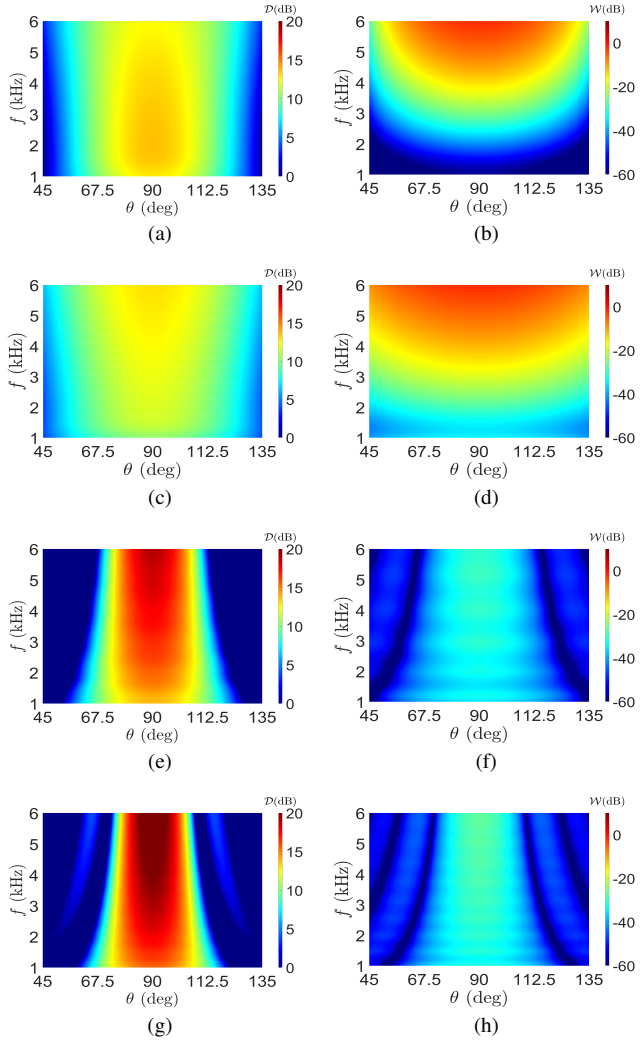

FIGURE 4: DF and WNG as a function of the elevation angle θ with the four discussed beamformers. (a) DF with $\mathbf{f}_{\mathrm{MDF/MDF}}$, (b) WNG with $\mathbf{f}_{\mathrm{MDF/MDF}}$, (c) DF with $\mathbf{f}_{\mathrm{MWNG/MDF}}$, (d) WNG with $\mathbf{f}_{\mathrm{MWNG/MDF}}$, (e) DF with $\mathbf{f}_{\mathrm{UCCA}}$, (f) WNG with $\mathbf{f}_{\mathrm{UCCA}}$, (g) DF with $\mathbf{f}_{\mathrm{DULA}}$, and (h) WNG with $\mathbf{f}_{\mathrm{DULA}}$.

deviation scenario in which $(\Delta_\theta, \Delta_\phi) = (0^o, 0^o)$ as their ability to attenuate the undesirable reverberations is, in this case, inferior. Nevertheless, in all other scenarios, either $\mathbf{f}_{\mathrm{MDF/MDF}}$ or $\mathbf{f}_{\mathrm{MWNG/MDF}}$ are shown to outperform the traditional beamformers in terms of both scores, with the former being superior in terms of the PESQ score in the $(\Delta_\theta, \Delta_\phi) = (40^o, 40^o)$ scenario and the latter otherwise. We deduce that the proposed approach is preferable either in mild reverberant environments or with significant DOA deviations of the desired speech signal.

VI. Conclusions

We have presented an SCCA-based KP beamforming approach that maximizes the broadband array directivity over a wide range of DOA deviations. As opposed to several previous approaches, with the proposed method, the necessity of acquiring the precise direction of the desired speech signal in advance is avoided. First, we optimize the locations of the microphones on an SCCA considering a wide predefined deviation range of the desired signal's DOA and design an MDF sub-beamformer. We empirically evaluate the optimal value of the parameter K, corresponding to the number of sub-beamformer microphones, as a compromise between the DF and WNG measures considering the deviation range. Then, to extend the SCCA sub-beamformer without adding significant computational complexity, we design two ULA sub-beamformers: an MDF sub-beamformer along the x-axis or an MWNG sub-beamformer along the y-axis. The proposed global beamformers are obtained as a KP of either one of the ULA sub-beamformers and the SCCA sub-beamformer to yield $\mathbf{f}_{\mathrm{MDF/MDF}}$ and $\mathbf{f}_{\mathrm{MWNG/MDF}}$, respectively. We analyze the beampatterns and the DF and WNG measures of the proposed global beamformers and compare them to traditional beamformers. We show that the proposed beamformers exhibit a considerably larger tolerance to DOA deviations of the desired speech signal concerning both the azimuth and elevation angles than the traditional beamformers. Finally, we perform speech signal simulations in noisy and reverberant environments and varying DOA deviation scenarios. Considering both the PESQ and STOI scores, it is demonstrated that our proposed approach is preferable, either in mildly reverberant environments or when the desired speech signal deviates from its nominal DOA.

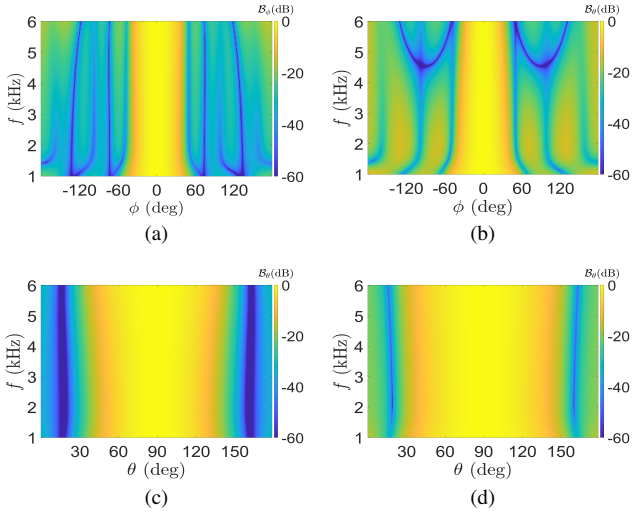

FIGURE 5: Beampatterns as a function of the frequency and either the azimuth angle ϕ or the elevation angle θ with the two proposed beamformers. (a) $\mathbf{f}_{\mathrm{MDF/MDF}}$ as a function of ϕ, (b) $\mathbf{f}_{\mathrm{MWNG/MDF}}$ as a function of ϕ, (c) $\mathbf{f}_{\mathrm{MDF/MDF}}$ as a function of θ, and (d) $\mathbf{f}_{\mathrm{MWNG/MDF}}$ as a function of θ.

REFERENCES

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

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

[3] 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.

[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] S. Gannot and I. Cohen, *Adaptive Beamforming and Postfiltering*, pp. 945–978, Springer Berlin Heidelberg, 2008.

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

[7] G. Itzhak, J. Benesty, and I. Cohen, "Nonlinear kronecker product filtering for multichannel noise reduction," *Speech Communication*, vol. 114, pp. 49–59, 2019.

[8] G. Itzhak, J. Benesty, and I. Cohen, "Quadratic beamforming for magnitude estimation," in *Proc. 29th European Signal Processing Conference, EUSIPCO-2021*, Aug 2021.

[9] V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3255–3266, 2021.

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

[11] 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.

[12] Y. I. Abramovich, G. J. Frazer, and B. A. Johnson, "Iterative adaptive Kronecker MIMO radar beamformer: Description and convergence analysis," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3681–3691, 2010.

[13] L. N. Ribeiro, A. L. F. de Almeida, and J. C. M. Mota, "Tensor beamforming for multilinear translation invariant arrays," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 2966–2970.

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

[15] 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.

[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*, Sep 2022.

[17] G. Itzhak, I. Cohen, and J. Benesty, "Robust differential beamforming with rectangular arrays," in *Proc. 29th European Signal Processing Conference, EUSIPCO-2021*, Aug 2021.

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

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

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

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

[22] Y. Buchris, I. Cohen, and J. Benesty, "Frequency-domain design of asymmetric circular differential microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 4, pp. 760–773, 2018.

[23] S. C. Chan and H. H. Chen, "Theory and design of uniform concentric circular arrays with frequency invariant characteristics," in *2005 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2005, vol. 4, pp. iv/805–iv/808.

[24] S. C. Chan and H. H. Chen, "Uniform concentric circular arrays with frequency-invariant characteristics—theory, design, adaptive beamforming and doa estimation," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 165–177, 2007.

[25] X. Zhao, G. Huang, J. Chen, and J. Benesty, "On the design of 3d steerable beamformers with uniform concentric circular microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2764–2778, 2021.

[26] R. Sharma, I. Cohen, and B. Berdugo, "Controlling elevation and azimuth beamwidths with concentric circular microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1491–1502, 2021.

[27] A. Kleiman, I. Cohen, and B. Berdugo, "Constant-beamwidth beamforming with concentric ring arrays," *Special Issue of Sensors on Sensors in Indoor Positioning Systems*, vol. 21, pp. 7253–7271, 2021.

[28] A. Kleiman, I. Cohen, and B. Berdugo, "Constant-beamwidth beamforming with nonuniform concentric ring arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1952–1962, 2022.

[29] L. C. Parra, "Steerable frequency-invariant beamforming for arbitrary arrays," *The Journal of the Acoustical Society of America*, vol. 119, no. 6, pp. 3839–3847, 2006.

[30] V. Tourbabin, M. Agmon, B. Rafaely, and J. Tabrikian, "Optimal real-weighted beamforming with application to linear and spherical arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 2575–2585, 2012.

[31] O. Rosen, I. Cohen, and D. Malah, "FIR-based symmetrical acoustic beamformer with a constant beamwidth," *Signal Processing*, vol. 130, pp. 365–376, 2017.

[32] T. Long, I. Cohen, B. Berdugo, Y. Yang, and J. Chen, "Window-based constant beamwidth beamformer," *Special Issue of Sensors on Speech, Acoustics, Audio Signal Processing and Applications in Sensors*, vol. 19, pp. 1–20, 2019.

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

[34] 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.

[35] 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.

[36] 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.

[37] E. A. P. Habets, "Room impulse response (RIR) generator," 2008.

[38] A. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, Springer International Publishing, Switzerland, 2019.

[39] "Darpa timit acoustic phonetic continuous speech corpus cdrom," 1993.

[40] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, 2001, vol. 2, pp. 749–752 vol.2.

[41] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, Sep 2011.